<u>**Response to reviewers**</u>

Thank you to both reviewers for your considered and constructive comments. The main changes to the manuscript resulting from the open discussion are the addition of a short appendix and additional detail provided in the manuscript's conclusions to contextualise the results, given key uncertainties. The appendix includes an assessment of the sensitivity of the results to key methodological choices and an assessment of the feasibility of performing the tracking operations on large datasets. Below, we respond point-by-point to all review comments.

<u>**Response to reviewer #1**</u>
Review comments left justified
Author responses indented

This study investigates the relationship between convective updrafts simulated by the global storm-resolving model ICON and their corresponding anvil properties. The authors first introduce a couple of new methodologies to improve the three-dimensional tracking of convective cores and their linkage to associated anvils. Specifically, they propose a combined tracking framework in which different components of deep convective systems are initially identified and tracked separately, and subsequently merged to analyze the storm as a coupled dynamical system. These developments build upon the existing tracking library tobac.

Using the new tracking methodology, the authors identify and analyze deep convective cores and their associated anvils and updrafts over the Amazon region. The relationships between updraft properties and anvil characteristics are quantified. The results demonstrate that both the updraft area and updraft intensity influence anvil area, with updraft area emerging as the more dominant control. A key finding is that systems with the largest and most intense updrafts are associated with anvils that are up to four times larger in area compared to the smaller and less intensive convective system of the tracked distribution. This result suggests that changes in anvil extent are not linearly proportional to convective core properties, but instead reflect nonlinear and potentially amplifying dynamical processes that can substantially enhance anvil growth.

The paper reads very well and is of clear relevance to both the global km-scale modeling community and the storm-tracking community. The analysis is thorough, carefully executed, and clearly presented. I have a few minor comments and suggestions for clarification (also a few points that would be worth taking up in the discussion), but overall I believe the paper is suitable for publication once these points are addressed. No additional data analysis is required.

General minor comments:

Importance of vertical resolution: Based on the presented analysis and development of the enhanced 3D tracking capabilities, do you have a sense of how important the vertical resolution is? Many km-scale models with similar grid spacing may have only ~ 50 vertical levels or so, depending on the vertical coordinates used. I am thinking in particular of Fig. 7e) which shows that you are not missing too many deep convective systems that have strong updrafts but are spatially disconnected from the anvil cloud. I wonder if that could potentially change with the vertical resolution.

> The number of vertical levels will primarily impact the subsequent analyses, rather than the performance of the tracking algorithm. As you refer to, the method relies on the updrafts connecting, at any time, to the tracked condensate field; the implicit assumption is that the two will co-occur at some point. This should not be overly dependent on the vertical resolution, as the mass condensed by the updraft will initially occur within the grid box and will be transported upwards or laterally whilst obeying conservation of mass. As long as the thresholds used to define the updrafts and clouds respect the distribution of values reported by the given dataset, the physics encoded into the km-scale models should ensure that the updrafts always co-occur with the cloud base.

Tropics vs. mid-latitudes: Similarly, it would be helpful to discuss potential caveats if the same method was applied outside the tropics. Do you expect the tracking enhancements to work equally well or were they so tuned for a specific cloud regime that they might only be useful for the tropics?

> This manuscript presented two tracking enhancements (1) the eroded-contiguity tracking, and (2) the multivariate approach to linking updrafts to anvil clouds.
>
> Development (2) has no limitations to the dataset or domain with which it is presented. The basic premise is that two objects can be tracked, in any manner, and these tracks can later be combined such that either object is permitted to precede or succeed the other.
>
> Development (1) has caveats that make it more and less suitable as an approach depending on the characteristics of the field in question, but is not subject to dataset- or domain-specific tuning. This approach offers the most usefulness where object shapes are complex or the magnitude of the field can vary heterogeneously within an object, both cases that challenge centroid-based or standard overlap tracking approaches. It relies on a time resolution that is small enough to ensure the objects in question cannot substantially displace themselves from one time to the next. If these characteristics are descriptive of a dataset, the method should offer meaningful enhancements, regardless of the exact field being tracked (liquid or ice).
>
> An additional consideration includes the threshold used to define cloud boundaries. Where this is more relaxed, the eroded-contiguity tracking will be more likely to connect more neighbouring and merging clouds. Where it is stricter, the condensate field associated with a given initial trajectory will be more likely to be uniquely tracked.

> More generally, the method presented here can be understood as an attempt to uniquely associate cloud with the updraft(s) that most closely fed it. Thus, we expect the method to be similarly successful in comparable endeavours, but **not applicable** to, say, frontal structures in the mid-latitudes or stratocumulus decks. We have added the following acknowledgement to the main text:
>
> *L204: "This method is likely generalisable to shallow convection and broader geographic regions, but may not be helpful in cases where link between the source of hydrometeors and the cloud deck are more complex or inseparable, such as frontal cloud in the mid latitudes or stratocumulus decks."*

Microphysics: The presented analysis is based on cloud ice which heavily depends on the liquid vs. ice partitioning in the applied microphysics scheme. In addition, the (somewhat arbitrary) partitioning between different hydrometeor types as well as the particle fall speed that determines the removal of ice are factors that affect anvil growth. Can you just add a brief discussion which of the studied aspects might look different in simulations with different microphysical parametrizations?

> We have expanded our previous acknowledgement of this possibility to the following paragraph.
>
> *L394: "Additionally, the model representation of these processes may result in an over or underestimation of the realistic response of anvil evolution to convective transport. For example, Jeevanjee and Zhou (2022) show total anvil amount decreases with increasing model resolution, suggesting that anvil production here may be overestimated compared to simulations in which turbulence is more highly resolved. In contrast, adopting a two-moment microphysical scheme instead of the single-moment scheme employed in the NextGEMs simulation assessed may imply that anvil production is underestimated, as studies have shown an increase in total ice under two-moment microphysics (Igel et al., 2015), including in the ICON model (Sela et al., 2025). Additionally, differences in the lifecycle of cloud ice from Lagrangian, rather than domain-mean, perspectives in response to different overall model configurations and aerosol loadings are highly variable, even disagreeing on the sign of the change (e.g., Saleeby et al., 2025). Overall, these uncertainties (i) emphasise the continued need to constrain the sensitivity of large-scale cloud properties to small-scale processes and (ii) indicate that the exact scalings reported between convection and anvil properties are likely inexact. However, the universality of these biases within the model suggests that the relative scaling between narrow and wide cores is qualitatively robust, and this claim is further supported by sensitivity tests performed to methodological choices and different measures of convective intensity."*

Feasibility/performance of 3D tracking: Can you provide some information about the performance of the 3D tracking (in terms of computational costs and time)? It would be great to know if the same analysis would be feasible to do on a global scale or if the analysis of storm-resolving models always still needs to be confined to a smaller region.

> Thank you for this suggestion; this is very helpful information to have included in the manuscript. The CPU time required to perform the 3D tracking per GiB is approximately 80 seconds, and scales linearly with dataset size (see Figure S3). Testing the method on a single day in the same NextGEMs simulation dataset but over the tropical belt, defined as from 15 S to 15 N, we found the tracking to take roughly 22 CPU hours to complete. Although we posit that this is a conservative estimate and that the time taken can be reduced by making use of more parallel computing nodes and fewer intermediate checkpoints than were used to perform this test.

Open science: Lastly, I wonder if there is a plan to publish the code enhancements as they would be hugely beneficial for the larger cloud tracking and high-resolution climate modelling communities.

> Yes, the algorithm used, including the new code enhancements, is available for public access in the GitHub repository published at https://github.com/mathilde-ritman/global3d_track. In addition to this, the eroded-contiguity method will be available to use within upcoming versions of *tobac*. The developments needed to link the results of one tracking method with another, used to link the two tracking operations and the convective cores and anvils, will also be available in upcoming versions of *tobac*. This detail has been added to the 'open source' section of the paper.

Detailed comments:

Fig. 1. I suggest plotting the liquid and ice water path separately, because one could gain a little insight into the phase partitioning of deep convection. Since the entire analysis is based on cloud ice, it would be great to see if there is a significant area in the tropics that has quite high LWP values but no high IWP values.

> Presenting the global distribution of frozen and liquid condensate separately provides useful regional context to the clouds studied in this manuscript, and Figure 1 has been updated accordingly. The updated figure shows that extending the analyses presented to liquid clouds would capture a broader region, particularly over the Amazon forest and the Pacific.

Table 1: Why did the authors choose the same thresholds for the centroid detection threshold and the boundary threshold for each variable? Given that cloud ice shows huge variations in magnitude within a single convective cloud, would it not make more sense to have a higher threshold for the centroid detection and a lower threshold to define the boundary of the cloud?

> The threshold choice is an important consideration in cloud tracking, particularly in cases where the fields in question are highly heterogeneous (such as ice

> clouds), and we are happy to provide additional detail on the logic behind the final algorithm used.
>
> As mentioned, *tobac* can use different thresholds to define the object centroid and the boundary. When defining object centroids, tobac calculates the centre point within each region that exceeds the given threshold value. When the centroid detection is provided with multiple thresholds, this operation is performed repeatedly, meaning that objects that contain regions of higher ice water concentrations will have centre points that are closer (in space) to where the condensate is most concentrated. This is particularly useful when fields vary greatly in magnitude between and within objects, and are very large.
>
> In our algorithm, we use two thresholds to define the ice cloud centroids, the larger being one order of magnitude greater than the smaller, the latter of which was also used when defining the boundary of the cloud. So, in answer to the direct question, we do provide a higher threshold for the centroid detection.
>
> During method development, we tested providing additional higher ice thresholds, but found that while providing more and higher thresholds did improve the robustness of the tracks, the results were still prone to issues, particularly once convection ceased and anvils began decaying. The eroded-contiguity step was specifically developed to combat this, and since it makes no consideration of the magnitudes of the fields, only their existence, differences associated with the number of thresholds provided to *tobac* centroid detection were no longer impactful.

Fig. 3: Since the authors decided on the ice water content threshold based on its distribution but chose a more arbitrary value for vertical velocities, I would suggest showing the ice water content distribution in this figure instead. This would contain more information that supports the choices made for the presented analysis. If you still want to keep this figure too, maybe attach it as supplementary material.

> We agree that this is a logical suggestion, and have included the ice water content distribution in Fig. 3, along with a comment (see below) on the logic used to choose the threshold. We would also suggest that the preceding discussion on why the vertical velocity choice was semi-arbitrary is nicely supported by the vertical velocity distribution, and so we have modified Fig. 3 to keep two of the three vertical velocity distribution panels previously presented.
>
> *L138: "Based on the peak mass mixing ratio (below which we assume dry pixels), we used thresholds of 5x10−6 and 5x10−5 kg kg−1."*

L. 9: I suggest replacing "convective intensity" with "updraft intensity" in the abstract, since this describes more accurately which variables were analyzed as the controlling factors for anvil clouds.

> Thank you, this suggestion offers a neat improvement to the specificity of the language used in the manuscript. This change has been made where relevant (for example, see L9). In response to the second review, we additionally performed some sensitivity tests to check how greatly the result presented in Figure 14 varied with the choice of convective intensity metric. We found that while the magnitude varied, the qualitative result comparing narrow and wide cores did not. When referencing this result, we use ‘convective intensity’ instead (for example, see L296).

L. 52 - 55: Why is the “scatteredness” or lack of convective organization a result of the overestimated vertical motions and precipitation intensity? The causal relationship between those two sentences is unclear to me.

> Indeed, we did not provide a clear link between these two statements. We had stated that “as a result [of overestimated vertical motions and precipitation intensity], the tropics tends to be characterised by more smaller and fewer larger clouds, and by less convective organisation, compared to reality”. The first implicit argument was that due to greater precipitation intensity and vertical motion, precipitation *efficiency* is increased, meaning clouds are less likely to maintain development into larger systems. The second implicit argument was that, because they tend to be smaller, dynamic interactions between clouds are reduced and mesoscale organisations are less likely to develop; this claim was supported by the study referenced (Becker et al. 2021). However, while both relationships may be plausible, neither were justified in detail and additional processes may influence the biases mentioned. Hence, in the interest of clarity and brevity, we rephrase the statement to the following:
>
> *L54: “… and the tropics tends to be characterised by more smaller and fewer larger clouds, and by less convective organisation, compared to reality.”*

L. 81: Since the enhancements of the existing tracking library focuses on the linking method of features with complex geometries and combining tracks identified on two different atmospheric fields, I would suggest to be specific and clarify what “advancing the capability” refers to here.

> The passage now reads as follows.
>
> *L81: “(i) advance the capability of existing cloud tracking algorithms to better handle the complex anvil morphologies and constituent structures (i.e., updrafts) associated with deep convective clouds in three-dimensional space,”*

L. 134: would replace “data-driven” by “data-informed” to not confuse with actual ML techniques

> Agreed, done.

L. 194: What does “coincident in time and space” mean here exactly? It does not seem straightforward how to combine lifecycle tracks of two components such as updrafts vs. ice clouds. Is the same overlapping technique used that was used for these variables separately? Is the total lifetime of the DCC truncated by the co-occurrence of updrafts and cloud ice or is it possible to have updrafts first which are then followed by a connecting anvil cloud later on (as would be common because I believe there is a time lag between cloud ice formation and the initial updrafts that drive liquid condensation). A few more details here would be helpful.

> Thank you for this request, this step in the algorithm was challenging to set out clearly and we welcome the invitation to ensure that it is clear. The paragraph was rewritten and now reads as follows.
>
> *L194: “At this stage, two separate tracking operations have been described, centroid-based and contiguity-based. These have been performed on both the vertical velocity and cloud ice fields. Next, the results of the two operations must be combined. This was achieved by identifying tracked objects from each method that co-occurred in space at any, but at least one, time. For example, if a cloud at some time, t, was tracked by both methods, then the trajectories of each method, which contain cloud detections from preceding or subsequent times, are combined. That means that the total lifetime of the cloud tracked is derived from the union of the two input tracks. By combining the results of both tracking operations, we additionally ensure that objects that may have been “shredded" vertically by the erosion are reconstituted.*
>
> *“Finally, the same method was used to determine the overall deep convective cloud system. To do this, the updraft tracks were compared to the ice cloud tracks, and, as described above, if one co-occurred in space with the other at any time, t, their trajectories were combined and designated to a single overall system. The result is that updrafts can precede the development of the anvil, and the anvil can be tracked beyond the cessation of convection.”*

L. 211: Refer to the nice visualization of this concept in Fig. 2. (panel 5)?

> Thank you, this suggestion has been implemented.
>
> *L222: “(this concept is visualised in panel 5 of Fig. 2)”*

L. 217: This line is a bit confusing. Does this mean that DCCs that have been determined to be “isolated” could actually be part of a larger complex? If so, I do not understand the criteria made to distinguish between “isolated” and “complex”. The text suggests to me that disconnected updrafts can oc-occur in one system connected horizontally by cloud ice, but would that not exclude that systems with isolated updraft and cloud ice are still part of the larger organized cloud scene?

> Thank you for raising this, we have clarified the text. The sentence you are referring to was intended as a general acknowledgement that DCCs deemed as

> isolated may be associated with large-scale organisation, for example, a squall line. Since we made no attempt to review the large-scale meteorology, we refer to the two cases as ‘isolated’ and ‘complex’, rather than ‘isolated’ and ‘organised’. The method and terminology for the isolated case are unchanged, but the term ‘isolated’ is relative to ‘complexity’ rather than to ‘organisation’.
>
> *L227: “The term “complex” is used rather than “organised” to acknowledge the possibility that isolated DCCs may have occurred within scenes characterised by large-scale organisation.”*

L. 391-392: Can you clarify whether any of the results of your study on the relationship between anvil thickness and convective cores, suggests a revision of previous assessments wrt anvil feedbacks? Or in other words, since radiative effects are the main motivation behind this investigation, can you conclude what the implications of increases/decreases in deep convective core intensity/area would be for ice cloud feedback assuming the relationship stays the same as in this study? Additionally, maybe you could also mention how NextGEMS specifically can be utilized to test whether the relationship will hold in a warmer climate or not.

> This study was strongly motivated by studies providing climate model evidence for reductions in the amount of thick high clouds in a warmer climate and, as a result, changed cloud radiative effects. Clarifying the link between our results and this motivation is a very reasonable request. We have revised the final two paragraphs of the manuscript to provide a more detailed comment on these points.
>
> *L409: “Thus, the greater dependence of anvil expansion on convective area demonstrated in this study provides an important basis for future process-level investigations into how convection controls the characteristics of anvil clouds most relevant for radiation and climate. This result implies that a reduction in the typical size of convective cores or in the intensity of convection in a future climate would see a reduction in high cloud area, and that changes in the frequency of larger convective cores may cause a greater reduction. We also found that both convective area and velocity were associated with greater anvil thickness, which then implies that a decrease in either convective property would reduce both anvil extent and anvil thickness. These results suggest that changes in tropical convection are a plausible mechanism for anvil cloud thinning, meaning that if the frequency of larger, stronger convective clouds reduces (e.g., Bolot et al., 2025), the net radiative effect of tropical deep convection may shift to more positive values and result in a more positive feedback on the rate of warming than previously supposed (Sokol et al., 2024; Raghuraman et al., 2024; Deutloff et al., 2025).*
>
> *“Convection-permitting climate models are a new and promising tool in challenging our understanding of how cloud processes interact with global climate. Experiments using convection-permitting models to simulate warmer climates offer an important test of whether the relationships assessed in this*

> *manuscript hold under a warmer global atmosphere, or whether other changes, such as changes in the microphysical evolution or diurnal cycle of convection, dominate changes in the radiative effect of tropical deep convection. In such efforts, Lagrangian methods that capture cloud development and lifecycle in the dimensionality in which it is simulated will help maximise the insights possible. Ultimately, we argue that studies and methods linking convection to anvil clouds are crucial to ensure that the processes that control anvil structure and spreading are understood and that projected changes can be physically justified."*

Response to reviewer #2
Review comments left justified
Author responses indented

General comments:

This manuscript investigates the relationship between deep convection and anvil cloud evolution using a novel 3D tracking methodology within the ICON global climate model. By implementing a distance-transform-based "erosion" technique, the authors successfully isolate convective cores from their surrounding debris—a necessary step for quantifying how core intensity and organization dictate anvil properties. The study finds that greater convective intensity or convective size both lead to larger anvils, albeit more reliably for convective size. It also finds a significant 4-fold increase in anvil extent associated with convective intensity when convective cores are large, providing valuable quantitative constraints for cloud-radiative feedback studies in high-resolution models. I generally support the study and recommend publication, as it addresses a critical source of uncertainty in climate projections. However, I have some comments regarding the justification of specific thresholds and several technical points that I encourage the authors to consider.

Specific comments:

1. Threshold Justification: The authors describe the 0.75 m s^-1 vertical velocity threshold and the 50% erosion fraction as "semi-arbitrary" or "experimentally determined." While I understand the practical necessity of these choices, the manuscript would benefit from a clearer demonstration of how sensitive the final results (e.g., the 4-fold increase in anvil area with intensity in large convective cores) are to these parameters.

> This is a very valid suggestion, and we agree that this adds greatly to the interpretability of the results presented for the community. Your comment 4, on the choice of the convective intensity metric, also relates to this comment, and we respond to both here.
>
> To address these three points (fractional erosion, vertical velocity threshold, and convective intensity metric), we performed some sensitivity tests. The outcome

of these can be summarised by stating that the key qualitative results of the manuscript are unchanged (in particular, convective intensity still mattered substantially more in wide cores), but that quantitative differences do arise that raise interesting considerations for future work.

The computational cost of repeating the exact tracking and analyses presented for multiple erosion and velocity thresholds is considerable. To get around this, we instead conducted two types of sensitivity tests. The first involved repeating the tracking operation with different erosion thresholds for a 24-hour period in the study domain and assessing the resulting sensitivity of the bulk statistics. The second involved reproducing the final results presented in Figure 14 after re-deriving key quantities from the original tracked dataset presented in the manuscript. All outcomes are detailed in the (new) appendix and are described below.

First, the erosion fraction. We tested three erosion thresholds on the first simulation day over the study domain, 0.25, 0.5, and 0.75. Statistics were then derived from the tracking metadata dataframes, which record the lifetime, extent, and thresholds met for each DCC and convective core tracked. The results are detailed in Figure S1 and are referenced in the main text at L173. They showed little change in the relationships between measures of convection and bulk anvil properties. Our interpretation of this is that while the erosion algorithm choices can change the partitioning of fields to particular labels and the duration for which fields are successfully tracked, they do not incur substantial changes in the fields included in the final dataset, and quantitative relationships between variables are largely preserved.

Second, the vertical velocity threshold. We redefine the convective cores by applying three vertical velocity thresholds (0.75, 1, and 2 m/s) to the column-maximum velocities derived from the convective cores in the original dataset (which were created using a threshold of 0.75 m/s). This approach avoids repeating the tracking and subsequent analyses from scratch for each vertical velocity threshold, thereby avoiding high computational costs. The results are presented in Figure S2. They show that while the threshold choice does not qualitatively change the results, quantitative changes do occur. Mainly, higher vertical velocity thresholds increased the magnitude of the response of the anvil area to convective intensity in wide cores. The implications of this are discussed in the appendix and are referenced in the main text at L294.

Third, the convective intensity metric. We perform the above velocity threshold analyses five times, once for the following five different convective intensity metrics: maximum updraft, mean updraft, maximum convective mass flux (CMF), and maximum convective core depth. The results are presented in Figure S2. Again, the qualitative results of the manuscript are unchanged, but quantitative differences arise. There was little change between the two updraft metrics, and a dampening of the “4-fold increase” by approximately half when using maximum CMF and approximately three-quarters when using the maximum depth. More details are provided in the appendix and is referenced in the main text at L297.

2. Model Bias Context: Since the study acknowledges that ICON overestimates convective intensity and prefers smaller cloud structures, it would be helpful to include a brief discussion on whether these model-specific biases might artificially amplify the scaling relationships reported between core intensity and anvil size.

> We have expanded our previous acknowledgement of this possibility to the following paragraph.
>
> *L396: "Additionally, the model representation of these processes may result in an over or underestimation of the realistic response of anvil evolution to convective transport. For example Jeevanjee and Zhou (2022), show total anvil amount decreases with increasing model resolution, suggesting that anvil production here may be overestimated compared to simulations in which turbulence is more highly resolved. In contrast, adopting a two-moment microphysical scheme instead of the single-moment scheme employed in the NextGEMs simulation assessed may imply that anvil productivity is underestimated, as studies have shown an intensification of total ice under two-moment microphysics (Igel et al., 2015), including in the ICON model (Sela et al., 2025). Additionally, differences in the lifecycle of cloud ice from Lagrangian, rather than domain-mean, perspectives in response to different overall model configurations and aerosol loadings are highly variable, even disagreeing on the sign of the change (e.g., Saleeby et al., 2025). Overall, these uncertainties (i) emphasise the continued need to constrain the sensitivity of large-scale cloud properties to small-scale processes and (ii) indicate that the exact scalings reported between convection and anvil properties are likely inexact. However, the universality of these biases within the model suggests that the \textit{relative} scaling between narrow and wide cores is qualitatively robust, and this claim is further supported by sensitivity tests performed to methodological choices and different measures of convective intensity."*

3. 3D Vertical Coherence: The erosion algorithm treats vertical levels independently. I would appreciate a note on whether this leads to vertical "shredding" of tracked objects and how the tracking algorithm ensures that a single convective system remains vertically integrated.

> This is indeed an important detail. The algorithm is not prone to shredding because (i) the tracking performed on the eroded masks links vertically contiguous pixels and (ii) when pixels lose vertical contiguity that is reinstated when the results of the two tracking operations (*tobac* linking and eroded-contiguity) are shared. We have included the following clarification to the revised text:

> *L199: By combining the results of both tracking operations, we additionally ensure that objects that may have been ``shredded" vertically by the erosion are reconstituted.*

4. Convective Intensity Definition: The authors used lifecycle-averaged intensity (wmax) to categorize objects into intensity quartiles. I wonder how final results would change if wmax were chosen at peak convective activity instead (peak wmax, or alternatively peak convective mass flux, which should occur in the first half of the lifecycle). I think the authors could possibly mention that the results (eg, 4-fold increase in anvil area) are likely sensitive to the definition of convective intensity.

> We agree that this is an interesting and important point. Please refer to our response to comment 1, in which we address both comments.

Technical corrections:

Abstract: Correct the drafting error "associated assoc with".

> Thank you, now corrected.

Table 1: In the vertical velocity row, replace "no applied" with "None" or "Not applied".

> Thank you for picking up on this, now corrected to “none”.

Line 131: Use standard time formatting (14:00–18:00) rather than 1400-1800 for consistency with the figures.

> Agreed, thank you, now done.

Table 1: Thresholds for specific ice water are written both as 1e-6 kg kg^{-1} (Line 139) and in decimal form like 0.000005 in Table 1. Standardizing these to a single format (e.g., 5 10^{-6} would improve readability

> Thank you, now done.

Figure 2: The caption should more explicitly link the sub-steps 3a and 3b in the diagram to the corresponding text in Section 2.2.4.

> A helpful suggestion, now implemented.

Line 324: Correct drafting error "the link between intense convection and".

> Thank you, done.

Line 352: Correct drafting error "excitation by gravity waves".

> Thank you, done.

Line 389: Please reformulate "and robust, and Lagrangian methods" to make the sentence clearer.

Thank you, done.

Line 401: Correct the name: "Daniel Klocke".

Thank you very much, done.