Reviewer-1

This study primarily seeks to examine sensitivities in convection-resolving numerical simulations of tropopause-overshooting convection and cross-tropopause transport to the choice of turbulence parameterization. A secondary focus is on sensitivity to the use of single- or double-moment microphysics parameterization. The model used is ICON and one of the turbulence parameterizations tested is the new 2TE scheme. Given the recency of the 2TE scheme and the need to more broadly evaluate it, I find the work to be timely and of interest to the upper troposphere lower stratosphere community. The study is focused and comprised of expected analyses based on similar approaches taken in prior studies. I found the graphics to be mostly well constructed. My comments are largely minor in nature and directed toward analysis design choices or explanation, but I anticipate the amount of work to address them being closer to major revision.

General Comments:

1. There are a few places where the text lacks direction/completion and fails to make a point that I could follow. Examples include Lines 41-47, 234-239, These elements should be revised for clarity.

Reply: Thanks for the suggestion. The lines were restructured for better clarity in the revised version.

2. The analysis would benefit from a more thorough evaluation of overshooting using approaches with differing strengths & weaknesses. Namely, the use of the water vapor - longwave IR brightness temperature difference (BTD) for diagnosing overshooting is a bit underwhelming as it has been shown to poorly isolate convective overshoots (e.g., Bedka et al 2012; http://dx.doi.org/10.1175/JAMC-D-11-0131.1). While I appreciate the need to have an observation-based evaluation tool for the model, some alternative approaches could be explored and included, such as the temperature difference between the longwave brightness temperature and tropopause temperature.

Reply: We thank the reviewer for pointing put this. To address this limitation while retaining an observation-based evaluation framework, we have adopted a hybrid approach that combines mature convective cell detection with the BTD criterion. Specifically, for the revised analysis we use output from the Cb-TRAM (Cumulonimbus Tracking and Monitoring) algorithm applied to MSG/SEVIRI observations and restrict our analysis to Stage-3 convective cells, which represent the mature phase of deep convection where overshooting tops are most likely to occur.

$$\text{Cb-TRAM Stage-3} = 1 \text{ AND } \text{BTD}(6.2\text{–}10.8\ \mu\text{m}) > 0\ \text{K}$$

This approach follows the philosophy of Bedka et al. (2010), where spatial confinement of active convective cores is used to mitigate the known deficiencies of the BTD-only method. By limiting the BTD criterion to objectively detected mature convective cells, we substantially reduce false detections associated with thin cirrus and anvil regions, thereby isolating the most dynamically relevant convective towers penetrating into the upper troposphere–lower stratosphere (UTLS). The reviewer also suggests the use of IR brightness temperature relative to the tropopause temperature. While this is a valuable complementary metric, it requires consistent, high-resolution tropopause estimates collocated in space and time with the satellite observations. Given the scope of the present study and our emphasis on a fully observation-based evaluation, we focus here on the Cb-TRAM-based hybrid method, which has been extensively validated against independent observations and lightning data (Zinner et al., 2008, 2013; Merk and Zinner, 2013).

References

1. Bedka, K. M., Brunner, J., Dworak, R., Feltz, W., Otkin, J., and Greenwald, T.: Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients, *J. Appl. Meteor. Climatol.*, 49, 181–202, https://doi.org/10.1175/2009JAMC2286.1, 2010.
2. Bedka, K. M., Dworak, R., Brunner, J., and Feltz, W.: Validation of satellite-based objective overshooting cloud-top detection methods using CloudSat cloud profiling radar observations, *J. Appl. Meteor. Climatol.*, 51, 1811–1822, https://doi.org/10.1175/JAMC-D-11-0131.1, 2012.
3. Merk, D. and Zinner, T.: Detection of convective initiation using Meteosat SEVIRI: implementation in and verification with the tracking and nowcasting algorithm Cb-TRAM, *Atmos. Meas. Tech.*, 6, 1903–1918, https://doi.org/10.5194/amt-6-1903-2013, 2013.
4. Zinner, T., Mannstein, H., and Tafferner, A.: Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data, *Meteorol. Atmos. Phys.*, 101, 191–210, https://doi.org/10.1007/s00703-008-0290-y, 2008.
5. Zinner, T., Forster, C., de Coning, E., and Betz, H.-D.: Validation of the Meteosat storm detection and nowcasting system Cb-TRAM with lightning network data – Europe and South Africa, *Atmos. Meas. Tech.*, 6, 1567–1583, https://doi.org/10.5194/amt-6-1567-2013, 2013.

3. The diagnosis of cross-tropopause transport could benefit from a few changes. First, it is stated that a cold-point tropopause definition is used, but such a choice is commonly problematic in the extratropics because it often leads to a biased tropopause identification. Figure 12 helped alleviate my concerns somewhat, but I would recommend an alternative choice such as a smoothed lapse-rate tropopause definition be used instead. In any case, diagnosing the tropopause and overshooting in a model simulation can be challenging because the tropopause is poorly defined in a convective core, such that a broader environmental reference tropopause height (e.g., a spatial average or median) may need to be used for reliable assessment. Second, only ice mass above the tropopause is considered for diagnosing cross-tropopause transport. Because the partitioning of water between the vapor and condensed phase can be quite sensitive to the choice of microphysics parameterization and mixing frequency, I would argue that both ice and water vapor distributions should be evaluated in Section 3.3. This could be done, for example, by examining total water instead (i.e., the sum of vapor and condensed mass) and differencing the result from the pre-convective tropopause-relative total water profile (which is likely all vapor).

Reply: We thank the reviewer for this important suggestion. We agree that a cold-point tropopause can be problematic in the extratropics and have therefore replaced the CPT with a vertically smoothed WMO lapse-rate tropopause throughout the analysis. To avoid diagnosing overshooting relative to locally distorted tropopause heights, we define an environmental reference tropopause height as the median lapse-rate tropopause over non-convective columns within the surrounding region (44.75–47.75°N, 5.85–10.5°E). Cross-tropopause transport and overshooting diagnostics are evaluated relative to this environmental reference, while the locally diagnosed lapse-rate tropopause is retained in the cross-sections to illustrate the convectively perturbed UTLS structure. This approach provides a more robust assessment of deep convective transport into the lower stratosphere in the extratropics.

We have also analyzed the tropopause-relative profiles of total water, computed as the sum of vapor and condensed phases, and evaluated their differences relative to the pre-convective environment. The figure is now added as supplementary figure in the revised version.

4. It is not stated what type of CAPE is computed for Figure 10. I am surprised that the differences in vertical velocity amongst the simulations were small compared to the CAPE diagnosed. Thus, I am curious if what is shown is surface-based CAPE rather than most unstable CAPE. Thus, it would benefit the evaluation if both surface-based and most-unstable CAPE were evaluated to best appreciate the differences in the simulations and their sensitivity to turbulence parameterization.

Reply: We thank the reviewer for this suggestion. In Figure 10, the CAPE shown corresponds to surface-based CAPE. To further assess the sensitivity of the simulations to the turbulence parameterization, we additionally evaluated the most-unstable CAPE (MU-CAPE) for both the TKE and 2TE simulations. The results show significant differences between the simulations that are consistent with those seen in the surface-based CAPE. This confirms that the differences in convective instability between the simulations are robust.

Specific Comments:

Line 6: insert ":" after "parameterizations"

Reply: Modified as suggested.

Line 8: Recommend changing "schemes" to "parameterizations" for consistency

Reply: Changed as per the recommendation.

Line 102: it is stated that a sponge layer spans 12.5 -- 20 km. Is this true? If so, I'd be *very* concerned about dynamics being unphysically dampened throughout the UTLS layer reached by the overshooting convection. Most prior studies performing similar simulations include a sponge layer in the top ~5 km of the model, well removed from layers impacted by the convection.

Reply: In the present study, the sponge layer is at 20 km with model top located at 30 km compared to the operational configuration whose sponge layer is 12.5 km and model top 20 km. We have clarified in the revised manuscript.

Figure 4, 6, & 7: it is impossible to know which lines correspond to the single-moment (SM) and double-moment (DM) simulations with the included key. Either the key should be updated or the caption should comprehensively describe the association.

Reply: Modified as suggested. The legend now differentiates clearly between SM and DM.

Line 293: "updrafts strenght" should be "updraft strength"

Reply: Modified as suggested.

Reviewer 2

Review of Evaluating Turbulent and Microphysical Schemes in ICON for Deep Convection over the Alps: A Case Study of Vertical Transport and Model-Observation Comparison by Alladi et al. In this research article, a case study of deep convection over mountainous terrain is examined across 4 high resolution ICON model runs with variations in the utilized turbulence and microphysics schemes. Specifically, the authors utilize model configurations with single- and double-moment microphysics as well as with the default TKE turbulence scheme and the newer 2TE turbulence scheme. Thanks to the model's fine spatial resolution, deep convection is well-represented and does not require parameterization. The model runs are compared with satellite measurements, including retrievals of overshooting cloud tops, ice cloud cover, ice water path, and ice optical depth. The authors find that, in this case study, the choice of microphysics scheme and turbulence scheme have a significant impact on the intensity of convection and frequency of overshooting cloud tops. The article's methods are sound and showcase interesting differences between the microphysics and turbulence schemes with respect to the case study at hand. I think it will be a valuable resource to the communities developing these parameterizations and exploring related processes. However, the paper requires some refinement before it is ready for publication. My comments on these issues below.

General comments:

1. Overall, the language and readability of this paper could use some improvement. I point out some specific examples below, but the paper could use a top-to-bottom re-evaluation in this area, ideally with a focus on sentence flow, grammar, and spelling. In some cases, like L41-47 and L341-343, relevant recent findings are described but not well-connected to the research at hand or the preceding sentence, leading to a stilted flow that doesn't lead the reader down a logical path. Along with the need for the described overall improvements, these specific passages should be revisited.

Reply: We thank the reviewer for this constructive comment. We have modified these lines in the revised manuscript for better clarity.

L341-343: A recent study conducted by Gettelman et al. (2024) over the Southern Ocean highlighted that the DM scheme significantly enhances the representation of supercooled liquid and ice clouds by predicting both the mass and number concentrations of hydrometeors, leading to higher ice water path (IWP) values compared to the SM scheme.

A recent study by Gettelman et al. (2024) over the Southern Ocean demonstrated that simulations employing a double-moment (DM) microphysics scheme produce higher ice water path (IWP) than those using a single-moment (SM) scheme. This behaviour was attributed to the explicit prediction of both hydrometeor mass and number concentrations in the DM framework, which allows for a more realistic representation of supercooled liquid and ice cloud populations and, consequently, a more physically consistent cloud-phase partitioning compared to the SM scheme.

2. The impacts of this paper should be elaborated upon further in the Conclusions section. In its current state, the structure of the Conclusions section feels clipped, mostly summarizing the key results without too much discussion of the broader impacts. As you state, the analysis is based on a single case, and further studies across diverse regions and conditions are needed to contextualize the results. I'm sympathetic to that. But based on these results, do you have any preliminary suggestions as to what sort of experiments this research will motivate? Are there specific areas of weakness in the model that have been exposed that need to be explored in future work? Were any aspects of your findings unexpected? How do these results contribute to the development and evaluation of the broader ICON framework?

Reply: Modified the conclusion section as suggested.

3. You cite Gettelman et al., 2024, (hereafeer G24) a few times in the report, describing their result of increased IWP in double-moment (DM) schemes as a consequence of enhanced ice nucleation, reduced sedimentation of smaller cloud ice particles, and enhanced accretion of ice to snow (these effects are also discussed more explicitly in L271-277). While this is a useful citation and comparison, and I wouldn't necessarily expect the effects to be different, I think a little more work needs to be done to make adequate comparison between the single-column, Southern Ocean study of G24 to the authors' deep convective case in complex terrain over Western Europe. Are you able to examine some of the causal pathways towards higher IWP that are described in G24? I'm not trying to suggest that you perform a whole other set of simulations exploring the sensitivity of various parameterization setups as they do in G24- that seems like an unreasonable increase in scope. But additional evidence linking these conclusions would be useful. Are you able to compare upper-level sedimentation and aggregation between your different model outputs? What about graupel amount, which is relatively similar in the results from G24 but is expected to be more different in a deep convection case (such as the one presented here)? Some of these results can be seen by the reader, qualitatively, by looking at Figure 12, but it's not explicitly discussed in the text. Overall, closer examination of the process differences between the SM and DM runs would be helpful for understanding the causal root of those differences and contextualizing this work in the overall literature.

Reply: We thank the reviewer for this constructive suggestion. To strengthen the process-level comparison with Gettelman et al. (2024; G24), we have expanded the analysis in the revised manuscript. Specifically, we compare vertical profiles of hydrometeor mixing ratios during convective and pre-convective periods relative to the local lapse-rate tropopause (Fig. A1 in the revised manuscript). The results show (i) enhanced upper-level QI and QV in DM during convection, consistent with more efficient upward transport and retention of ice, (ii) scheme-dependent differences in vertical phase partitioning below the tropopause, and (iii) contrasting graupel responses between TKE and 2TE simulations, indicating sensitivity of riming and mixed-phase processes to turbulence–microphysics coupling.

In addition, the profiles of hydrometeors (QC, QI, QS, and QG) averaged over the analysis region are investigated and discussed for both SM and DM simulations in the revised manuscript.

Specific comments:

Figures 4, 6, 7: Due to the formatting of the linestyles and the legend, it is challenging to distinguish between SM and DM model runs. For instance, here's the Figure 7 legend: With close examination, you can see the slightly shorter line length that the -DM runs have compared to the -SM runs (marking them as dashed lines), but that detail is diJicult to discern. I suggest you change the formatting of these legends to make the diJerence more obvious.

Reply: Modified the legends as suggested in the Figures 4, 6 and 7.

Figure 4: Extremely minor note, but your time axis for this diurnal cycle plot uses zeropadded single digits (e.g., "03" instead of "3") for the tick labels. You may want to consider matching the formatting of your other diurnal plots (Figs 6, 7, 8, and 10).

Reply: For uniformity the time axis in Figures 6, 7, 8 and 10 have been changed using the zero padded digits.

Figure 7: The high IWP from observations between 4 and 11 UTC is quite different from the model output here. Is this due to the advection of high clouds mentioned in L203 with respect to Fig. 4? If so, it may be useful to remind the reader of this detail. If not, it should be addressed.

Reply: It is due to the advection of the high clouds mentioned before. This detail is reminded again for readers benefit as suggested.

Figure 9: For context, it would be useful to show the cross-section line (magenta line in Fig.

3) in these plots as well.

Reply: Added as suggested.

Figure 10: Are these time series for each day simulated (July 7, 8, and 9)? Because the plots are not labeled with the day they are representing and all analysis so far has been from July 8, the introduction of this new analysis is jarring. The temporal domain of each subplot needs to be specified (perhaps with labeling the columns of your subplots with the day in question?). Assuming this interpretation is correct, I am confused at the intent behind the introduction of these extra days of data. (b) and (e) are well motivated (assuming they are for July 8) because this is the case we have examined thus far. We have context for this data. But what additional information do we get from seeing the time series of CAPE and w for 7/7 and 7/9? Is the intent to show that the diJerences in model behavior are not limited to the circumstances present in the main case study day? Whatever the case, more rationale is needed to support the reason for the inclusion of the additional time series (or they should simply be cut). As-is, I find the 7/7 and 7/9 plots distracting and confusing more than helpful. For instance, in (a) there is a conspicuous diJerence in CAPE between 2TE-SM and 2TE-DM that is not seen as strongly in plots (b) and (c), but I have no context for what the causes for this discrepancy may be as this day is not examined or displayed anywhere else in the paper.

Reply: We thank the reviewer for this careful observation. All-time series shown in Figure 10 correspond to 8 July only; no additional days are included in this figure. The confusion arose because the date was not explicitly stated in the original caption or panel annotations. To avoid any ambiguity, we have revised the figure caption to clearly state that all panels show the diurnal evolution on 8 July over the same box region for the different ICON simulations.

Figure 11 and 12: Units for potential temperature are not specified. Additionally, the subplot labels (a-d) are very difficult to see with the contour background. Consider moving these labels outside the plot. The same is true for the CPT legend in Figure 12.

Reply: We thank the reviewer for these helpful suggestions. The units of potential temperature have now been explicitly specified in the figure and also in figure captions. In addition, the subplot labels (a–d) have been repositioned outside the plotting area to improve their visibility against the contour background. The CPT legend in Figure 12 has also been relocated to enhance readability. These changes have been incorporated in the revised manuscript.

L42: This sentence ("However, the deep convective…") is awkward and not grammatically correct.

Reply: Changed

L51: NWP is not defined.

Reply: Defined

L105: FAO is not defined.

Reply: Defined now.

L148-149: "CiPS" is not directly stated to be an acronym for the Cirrus Properties from SEVIRI algorithm.

Reply: Stated at the place where first used.

L293: "strength" is misspelled.

Reply: Modified.

L295-296: The line described here (and plotted in Figures 11 and 12) is the cross-section shown by the magenta line in Figures 3 and 5, and this is described as such in the Figure 11 caption. For the sake of readability, it would be useful to also state this explicitly in the text.

Reply: Mentioned as suggested.

L335-337: Awkward sentence structure and incorrect capitalization of "Its", recommend reworking.

Reply: Restructured as suggested.

L336: ABL is not defined (although PBL is on L46)

Reply: Changed to the atmospheric boundary layer.