We would like to thank the reviewer for taking the time to evaluate our manuscript and suggest changes. These comments have lead to a substantial improvement of this work. We detail the changes we have made below.

*R2:Technical Details.*
*1. P. 20365, line 15: need just one "of" before Holtslag* *Done.*
*2. P. 20368, line 14: should be "similarity"* *Done.*
*3. P. 20369, line 5: should be "data were"* *Done.*
*4. P. 20374, line 21: remove "for example"; change "compared" to "comparison"* *Done.*
*5. P. 20376, line 10: should be "data were"* *Done.*

*R2: 1. Much of the paper provides weak ascertains (e.g. "Some of the offset of the models towards higher CO values may be due to the idealised CO tracer having a longer lifetime than CO in the atmosphere" in Section 4.4 and "The timing of the transport events is probably of lesser importance for average tracer mixing ratios" in Section 5) that need to be strengthened with supporting evidence. On the other hand, the abstract does a fairly nice job of stating what the findings of the study are. The same conviction expressed in the Abstract should be used in the body of the manuscript.*
*We have firmed up many of the statements made throughout the paper, also by supporting them with more details from paper 1.*

*R2: 2. In the Model Description text, the models should be described in a consistent fashion. I find that it is much more useful to include a brief description or terminology of a method than just providing a reference. In Table 1, the convective parameterization used in each model should be included; less important is the reference which can be included in the text.*
*We have changed table 1 so that the convective parameterisation is given and the reference for the models is given now in the text. We have included terminology and brief descriptions of the various schemes used for convection or boundary layer transport/mixing in the model descriptions.*

*R2: 3. This paper should be able to stand on its own, without the need to have Part 1. Because*
*of this, there is a need to incorporate more information into Part 2. For example, a plot showing the regions of study would help greatly. A summary of the results of Part 1 would provide necessary background information. Specifically, section 4.1 says what Part 1 addresses (compare modeled and observed cloud top height distributions to investigate the capabilities of convective parameterizations), but does not say what the conclusions are. Certainly these results will relate to the tracer transport results!*
*We have added more comparisons with paper 1 throughout the manuscript, to improve the analysis. The specific statement mentioned by the reviewer was an introductory remark, and is followed by more exact statements about the model meteorology in this section. We now also present the main conclusions of paper 1 in the introduction section.*

*R2 4. In the Results section, there are many places where explanations of differences between models are not given. Gaining insight as to the causes of these differences would guide the community as to ways to improve parameterizations in the large-scale models. Specific issues are included in the next section.*
*We have improved this in many places throughout the manuscript, both generally, and in reply to the specific comments below.*

*R2: Title: As Reviewer 1 suggests, the title should be changed because the composition of the troposphere is not addressed. Idealized tracer transport is the topic. Why not use a title that describes only what is discussed in the paper, e.g. Intercomparison of tracer transport in large-scale chemistry transport models for tropical deep convection regions. (In my opinion, the regional-scale models are behaving much like a global model but with the addition of lateral boundaries.)*

The title has now been changed to "Representation of tropical deep convection in atmospheric models – Part 2: Tracer transport"

*R 2. Abstract: While I applaud the presentation of the Abstract, I am concerned about over-stating some results. For example, the importance of the boundary layer scheme was based on only one comparison with another PBL scheme. However, there are many PBL schemes that were not tested. How do the authors know that ": : :tracer profiles are found to be strongly influenced by : : : boundary layer mixing parameterisations: : :" when it could be that the Louis (1979) scheme is the anomaly? In other words, I am not sure the paper provides enough evidence to give such a broad conclusion.*
*Think carefully about each conclusion stated and whether there is strong evidence to support that conclusion. If the authors think that it is not the role of this paper to conduct further analysis on a topic, then the conclusion must be stated as "this study indicates a process is important and should be further investigated in more detail".*
*We have taken this statement into account in the revised version of our manuscript, and provided more evidence to support some of our assertions, in particular by including more information from paper 1 in the analysis.*

*R2:26. Lastly, there has been similar work done at the mid-latitudes (and tropics). While these previous studies do not conduct the same type of analysis as this paper, or use the same perspective, it might be smart to cite them as important previous studies that have shaped our current perspective. Ones that I suggest are: Barret et al., ACP, 2010; Barth et al., ACP, 2007; Jacob and Prather, Tellus B, 1990; Lawrence and Rasch, JAS, 2005; Mahowald et al., JGR, 1995; Tost et al., ACP, 2010.*

*Thank you for pointing this out. These citations have been included, along with brief descriptions.*

*R2:. Section 2.2 KASIMA CTM. Besides including a reference to the model, an explanation of how the lower boundary (at z = 4 km) is handled would be useful.*
A reference for the model and description of how the tracers are set at the lower boundary has been added.

*R2: Section 2.9, CATT BRAMS: Nothing is said about the advection scheme. Again, it is better to describe the type of scheme (e.g. semi-Lagrangian, flux-form, monotonic, positive definite) than to just give a reference.*
We have now improved the description of CATT-BRAMS, as suggested.

*5. Section 2.10, WRF NWP: There are a couple of options for the type of advection in WRF. The first is not necessarily conserving and is not positive definite, while the second is positive definite. Thus, it is possible that WRF results have numerical issues with advecting tracers. Please elaborate as to which advection scheme is used.*
The order of the advection scheme was 5 in the horizontal and 3 in the vertical (Wicker and Skamarock, 2002). The advection scheme is conservative, but not positive definite nor monotonic.

Positive definite and monotonic options do not work because large Courant numbers near the poles invalidate the limiters and Fourier filtering for the polar filter is not monotonic or positive definite.

We have updated the text to include this information.

*R2: For the regional models, how are the lateral boundaries handled (both for meteorology variables and tracers)? CATT-BRAMS was the only regional model studied here. A description of how the boundaries are dealt with has been included.*

*R2: Section 3.1, T6h description: Is the tracer continuously replenished in the z < 500 m region, so that T6h = 1 at every time step for z < 500 m?*
*Yes, that is correct. We have added text to clarify this in the manuscript.*

*R2: Section 4.1, first paragraph. First, the paragraph is poorly constructed. The background information should come before the Figure is introduced. Include in the background information a summary as to what was found in Part 1. Define the regions via a plot and state the definition of the tropics early (current version of text should have it on line 21 of p. 20369).*
*The paragraph has been re-organised as the reviewer suggests. A plot of the regions has been added, and the regions are defined. Instead of including a summary of the findings of part 1 at the beginning of the section, we have included more direct discussion of the results of part1 in the analysis of the experiments in part 2, to explain particular model differences.*

*R2: Section 4.1, second paragraph. I found this analysis of the mean convective outflow height to be interesting. There has been work done (Mullendore et al. JGR 2009) that has determined outflow heights of a storm in South America using radar reflectivity. It would be interesting to compare that one case study to your South America results. Obviously it cannot give validation to model results, but can indicate what might be expected for that region. Plus, it is a good step for bringing global-scale climate studies closer to mesoscale meteorology work.*
*We have included this reference, with discussion.*

*R2: P. 20373, lines 14-26. There is much reporting of how one model compares to another, with little explanation of why that difference occurred. Remember, many readers will not know the details of these models and cannot interpret the differences meaningfully. Please explain why differences occur.*
*We have added further comments on the causes of the differences (both here and throughout the paper) where these can be determined. As there are several factors which contribute to the tracer profile produced by a certain model, it is however difficult to be exact, unless comparing two versions of the same model which differ in only one respect (for example pTOMCAT and pTOMCAT_tropical).*

*R2: P. 20373, lines 28-29. The CATT-BRAMS model has the T6h mixing ratios almost completely decaying. Where does it go? In CATT-BRAMS, the convective transport essentially completely ceases between the convective peaks. T6h has a lifetime of 6 hours, therefore it simply decays. The other models have a more constant convective transport source into the upper troposphere. We have added text to clarify this.*

*R2: P. 20374, lines 23-25. One of the conclusions that I drew from this study is that the PBL parameterization is also important for convective transport of tracers. First of all, the mesoscale meteorology community already knows this for cloud-scale studies.*

*It is generally because each PBL scheme predicts different PBL heights and different stability of the boundary layer which feeds to the convection parameterization. (a) However, I am not familiar with the Louis (1979) PBL scheme. Could the text explain how the scheme works and hopefully include what it tends to do to PBL heights. (b) In Figure 3, the TOMCAT-Louis curves appear to not have any effect from convection. Could the authors examine whether convective transport was occurring or not? I did not see these results in Part 1 of the paper – I wonder how it compares with cloud top distributions, convective intensities, etc.*

We have added text both to the model description section and to section 4.1 to clarify this. The Louis scheme is a local scheme with no eddy transport. The height of the boundary layer is low using the Louis scheme.
The Louis scheme does also not account for entrainment at the top of PBL (the limitation by using by Louis scheme, see Stockwell and Chipperfield (1999).
The convective transport parameterisation is the same in both model versions.

*R2: P. 20375, lines 5-9 describe differences between simulations with and without nudging. How good is the analysis that drives the nudging for the Maritime Continent? In other words, is there a paucity of data in that region that can then affect model results? These lines are another place where there is no explanation as to why the results are different. The text only says that the curves are different.*

The reasons for the differences between the two models are hard to pinpoint with the data available – unfortunately we have no archived meteorological fields for the simulation with UMUKCA-UCAM(R1). There is not much we can do here except describe the differences.

*R2: Paragraph spanning pages 20375 to 20376. Do the differences between models worry the authors? What do these differences say about the community's capability in predicting locations of convective transport? Answering this question would be a worthy conclusion of the paper.*
*The differences in the location of the enhanced tracer mixing ratios are very small between the models, and we don't find these worrying. The plots show the main difference is in the magnitude of the variation of the tracer mixing ratios. We have added some text to emphasise this. Later in the paper we now include TES CO maps, where we can comment on the location of the highest tracer mixing ratios.*

*R2: 15 Would MOPITT data or other satellite data be helpful for interpreting Figures 4 and 5? One might expect that biomass burning would affect the MOPITT CO at 200 hPa,* but at 90 hPa it may have a minor signal and CO enhancements may be found. Just a thought.
We now have TES map plots of CO, see below. These are much better than the TES CO line plots in the previous version of this manuscript.

*R2:16. P. 20376, lines 25-27 seem to be too detailed. What is the reader suppose to learn?*
*We wanted to point out that the tropical mean is strongly influenced by what goes on over the maritime continent, and that in the case of UMUKCA-CAM (R1), the situation is different, as the tropical mean cycle looks more like that over west Africa. We agree that we didn't express this well, the text in the paper has been modified.*

*R2: 17. P. 20376, line 28 to P. 20377 line 6 compare the T20 seasonal cycle to the precipitation rates in Part 1. In my opinion, this discussion should come sooner in the paper.*
*Second, why should T20 (20 day lifetime) be correlated with precipitation rates? I would think that T6h would be better correlated.*
*To me, the model results are showing that PBL tracers are lofted to the 150-200 hPa region (Figure 1) and then lifted by large-scale air flow to <100 hPa which explains why T20 is much more abundant above the 100 hPa level. Is this the conceptual picture the authors are trying to convey?*

*We located this discussion where it currently is in the paper, because it relates to the regional seasonal differences in transport, whereas the early part of the paper is concerned more with the shape of the average vertical profiles. It is correct that T20 seasonal cycle should not necessarily be related to precipitation – if the main transport process affecting tracer concentrations in that area is not convective transport. As stated by the reviewer, the tracers are lofted to 200-150hPa, and further ascent is due to large scale transport processes. We have added text to this section accordingly. T20 is a better tracer of this large scale transport, the T6h mixing ratios become extremely small for some models above 100hPa.*

R2: P. 20377, line 16. Again, why use T20 to locate rapid transport by convection. It seems that T6 would be a better tracer.
*This section has been removed, as recommended by reviewer 1.*

# SECTION 4.4

R2:19. In Figure 8, only some of the models are plotted. How do the regional-scale models compare with the measurements?
*CATT-BRAMS was the only regional scale model, and is a NWP model, therefore lacking the possibility to include chemistry. Actually only round 2 models ran CO, and among those WRF, CATT-BRAMS and UM-UCAM_highres could not run CO either. The remaining model results are plotted.*

*R2: 20. P. 20379, lines 25-28. Why do models over-estimate measurements?*

*Additional analysis now included in the paper indicates that there is simply too much vertical transport, and most likely this is not due to too strong convection in the models. This is now discussed in the section comparing measurements and models, as well as in the conclusions.*

*R2:21. P. 20380, line 20. I thought the step-and-stare mode of TES gave better vertical resolution compared to the Global Survey mode. Was not the step-and-stare mode used?*

*To the best of our knowledge, the step-and-stare observations do not have any more vertical information, and the retrievals are done in the same way for both modes, with the only difference being the distance between observations along the track. The step-and-stare mode gives better horizontal resolution but only in a very restricted area, which doesn't make it very suitable for global comparisons. We have changed the plot of the TES data to show maps of CO now, as suggested by reviewer 1, this gives us a more useful comparison.*

*R2: 22. P. 20381, lines 6-17. Why do the models underestimate TES data?*
*This plot has been replaced, and the new plots showing maps of CO at 350 and 195 hPa enable a far better discussion and analysis. Basically, the models mostly overestimate the background CO values due to a too rapid non-convective transport, and underestimate the convective peaks, due to low surface values (peaks are averaged out due to the model grid), or too little convective transport.*

*R2: 23. Discussion section: The results of the TOMCAT-Louis simulation are brought up again as an important result, however there is no explanation of what causes the result. Please include discussion of the characteristics of the Louis scheme that are causing differences in model results.*
*Done.*

*R2: Discussion section: The implication of using the same ECMWF data to drive the models yet getting different results in convective outflow height is that the ECMWF data is not controlling the outcome. Instead, the internal model dynamics and physics are critical to the tracer transport profiles*.
That was indeed the point we wanted to make. We have changed the text slightly to emphasise this point.

*R2:25. P. 20383, lines 5-15. The authors state that the timing of convection is not important for average mixing ratios in the UTLS. I would agree with that from the perspective of what is entering the stratosphere. However, timing is very important for the PBL stability (as is said in the next sentence) and for emission strength of PBL tracers which then affect the magnitude of the tracer lofted to the UT.*
*This is a good point. The sentence in question has been revised to state that for the composition of air entering the stratosphere, the timing (multi day time scales) of the events is not so important, but in terms of diurnal cycle, timing could play a role if it lead to systematic differences in the stability of the boundary layer between the models, during convection.*