Interactive comment on "Three-dimensional variations of atmospheric CO_2: aircraft measurements and multi-transport model simulations" by Y. Niwa et al.
Reply to anonymous referee #2
Y. Niwa et al.

The investigators use 4 atmospheric transport models (3 online and 1 offline) driven by 2 different datasets of surface CO_2 flux (Flux 1 and Flux 2). The outputs from different numerical experiments are then compared to aircraft measurements taken at mid to upper troposphere as part of the CONTRAIL project. I find the paper informative and contributes to the further understanding of the behavior of various atmospheric transport models we use to interpret atmospheric CO_2 in terms of carbon sources and sinks. I recommend acceptance with major modifications.

We are grateful for your time to review our paper. We also appreciate for giving us many fruitful comments and suggestions. Our replies to the comments are described below.

Overall, the study presents some interesting model-to-model intercomparison results. The differences amongst these models in simulating radon (Figure 3) and CO_2 (Figure 4) are attributed by the authors to differences in the vertical mixing and cumulus convective parameterization schemes employed in various models used in this study. In this context, I have some questions/comments based on what is stated in the paper . . ..

(1) For Figure 3, the authors state that the "low radon concentration suggests that vertical transport of MJ98-CDTM is slower than those of the other models." Now, based on what is presented in Section 2.2.2 (MH98-CDTM0 and in Section 2.2.4 (NIES), it is my impression that the convective schemes employed in these two models are similar, yet the difference (particularly at 300 hPa) in Figure 3 between the two models is quite noticeable. Why is that? I would also like to suggest switching the 300 and 500 hPa columns of diagrams, putting the 500 hPa column to the left.

As suggested, we switched the 300 and 500 hPa columns. Furthermore, we changed JJA radon figures into JAS ones and newly added 850 hPa column in the leftmost side following the Reviewer#1's comment.

It is our feeling that the scheme difference between MJ98-CDTM and NIES is quite large. Although the cumulus convection schemes used in the two models are Kuo-type, MJ98-CDTM uses the near original Kuo scheme and NIES uses Grell scheme. Moreover, NIES uses precipitation rate for convective updraft.

(2) In the sentence following the one quoted above, the authors state that "the simulated radon concentrations are rather comparable with each other . . ." I strongly suggest deleting the word "rather."

As suggested, we deleted "rather".

(3) In Figure 4, why does NIES look quite different from MJ98-CDTM for JFM? Isn't the convective parameterization scheme used in each of the models very similar?

We think that the difference comes from the shallow convective parameterization of MJ98-CDTM, which is not used in NIES. The different parameterizations for deep cumulus convection may be another cause. However, in JFM, deep cumulus convections are not so active in the northern hemisphere, therefore we consider shallow convection as more probable cause than deep cumulus convection. There are also other possible reasons to contribute to the difference such as boundary layer scheme and difference in wind data which is treated differently in offline and online models. NIES model uses mass flux correction, while MJ98-CDTM does not. In the manuscript, we changed the following sentence.

"For JFM, MJ98-CDTM simulated…and NICAM-TM simulated smaller ones."
[from Page 9, line 27 to Page 10, line 3]
=>"For JAS, both MJ98-CDTM and NIES simulated larger CO_2 vertical differences over northern land, although ACTM and NICAM-TM simulated smaller ones. For JFM, MJ98-CDTM simulated smaller vertical differences over northern lands than the other models did. Probably it is because the shallow convection scheme of Tiedtke (1989) only used in MJ98-CDTM tends to mix concentrations at lower altitudes more strongly. There are also other possible reasons to contribute to the difference such as boundary layer scheme and difference in wind data which is treated differently in offline and online models. NIES model uses mass flux correction, while MJ98-CDTM does not."

(4) Also in Figure 4, unless I missed it in the main text, which flux (Flux 1 or Flux 2) was used to generate the diagrams in the figure? This information should at least be mentioned in the caption.

We used Flux2. We modified both the main text and the caption as follows.

"July–August–September (JAS)"

[Page 9, line 27]

=>"July–August–September (JAS) calculated from Flux2."

[Caption of Figure 4]

"NIES (lowest panels)" => "NIES (lowest panels) using Flux2."

In regards to the comparison between the model output and the observation, I make the following questions/comments/suggestions:

(1) In Section 3.2, the authors calculate average correlation coefficients in their attempt to establish the "reasonableness" of the model reproduction when compared to the observation (e.g., 1st paragraph, 2nd sentence: ". . .. models reasonably reproduced the observed vertical profiles: average correlation coefficients . . .. are 0.62 and 0.71, respectively" for Flux 1 and Flux 2.) How statistically significant are these numbers, and from each other? Every time one performs a statistical calculation comparing one variable with another, it is essential to establish statistical significance level. Here is another example from the paper (and there are others) in Section 3.3 (1st paragraph, 3rd sentence) where the authors state that "seasonal amplitudes simulated from Flux 2 are large and closer to the observed one than those from Flux 1." Is the difference between Flux 1 and Flux 2 statistically significant at, say 95% confidence level?

First of all, in the previous manuscript, we derived the average correlation coefficients by simply averaging. However, we have realized that an average of correlation coefficients in a number of samples does not represent an "average correlation". Therefore, we changed the way to derive the average correlation coefficient. We apologize for our misunderstanding. In the revised manuscript, correlation coefficients are transformed into Fisher's z prior to averaging and the averaged coefficient is derived by back transforming the averaged z. Therefore, the coefficient values are different from those in the previous manuscript.

According to the comment, we checked the significance of the revised correlation coefficient at 95 % confidence level and found that all the average correlations in the manuscript are significant. However, we also found that the difference of correlations between Flux1 and Flux2 are not significant. Accordingly, we changed the sentences as follows.

"averaged correlation coefficients of each vertical profile between the observation and the model mean are 0.63 and 0.71,"
=> [Page 10, line 13]
"average correlation coefficients are 0.83 and 0.85 (significant at 95 % confidence level)"

We removed the following sentence from the first paragraph of Section 3.2.
"In addition, the result suggests that vertical profiles have measurable sensitivity to surface flux. Moreover, because the CONTRAIL measurements were not used in the inversion of Flux2, i.e. independent data, the improvement of the correlation by Flux2 shows some validity of the inversion."

"the averaged correlation coefficients are 0.80 and 0.87"
=>[Page 11, line 2]
"the average correlation coefficients are 0.92 and 0.94"

=>[Page 14, lines 15-16] newly added
"Over all the nine areas, the models reasonably reproduced seasonal variations both with Flux1 and Flux2 (Table 5)."

For seasonal amplitude, we also modified the values. In the previous manuscript, we calculated the averaged seasonal amplitude from 4 modeled mean seasonal variation. In the revised manuscript, we derived the averaged seasonal amplitude by averaging 4 modeled seasonal amplitudes. From those calculated values, we performed t-test to investigate the significance of the difference between Flux1 and Flux2. As a result, we found that the differences are significant only for SSA and AUS. According to the result, we modified the main text.

"Furthermore, Flux2 improves correlations…favourable to simulate CO_2 for this period. However,"
=>[Page 14, lines 18-19]
"However, most of those changes are still not significant at 95 % confidence level, i.e., model–model differences are large compared to the changes by the fluxes. Furthermore,

We removed the following sentence from the second paragraph of Conclusions.
"In terms of the correlation coefficient, root-mean-square difference, and seasonal amplitude, the CO_2 concentration field simulated from Flux2 is closer to the observed one than that from Flux1, indicating some validity of the inversion that produced Flux2."

Also, the simulated seasonal amplitudes and correlation coefficients in Table 5 were changed according to the change of the calculation described above. Furthermore, we added the following sentences in the caption of Table 5.
"The simulated amplitudes are averaged for the four models. Bold font in the Flux2 column represents a value significantly different from Flux1 at 95 % confidence level. The average correlation coefficient is derived by back transforming the averaged Fisher z. All the correlations are significant at 95 % confidence level."

(2) In Figure 5, there are large differences, depending on the geographical location, between the model output and the observed vertical profiles. When the authors state that (Section 3.2, 1st paragraph, 2nd sentence) "the models reasonably reproduced the observed vertical profiles" I think it might be helpful to put this in the context of between-model differences.

Because the correlations are found significant, we did not change the sentence "the models reasonably reproduced…". Instead, we added sentences to discuss about the model differences as follows.
[Page 10, lines 18-24]
=>"Although general transport features are similar in ACTM and NICAM-TM as shown in Fig. 2 and 3, the differences of the vertical profiles between the two models

are comparable to those between other models in some locations (e.g. IND). It suggests that vertical profiles are sensitive to local/regional transport process. The differences at IND may arise from the different wind fields (ACTM uses NCEP2 and NICAM-TM uses JCDAS for the nudging data) or the different Mellor-Yamada type scheme for vertical turbulent mixing."

(3) In the caption for Figure 5, there is a mention of two panels showing time-altitude cross-sections, one for 2006 and another one for 2007. This needs to be explained and discussed in the main text.

As suggested, we added an explanation for the time-altitude cross-sections in the main text as follows.

[Page 10, lines 10-12]

=>"Figure 5 also presents time–altitude cross-sections of the observed daily ΔCO_2 for 2006 and 2007, showing that much more data were obtained for 2007 than for 2006 in most areas."

(4) Section 3.2.1 (3rd paragraph, 4th sentence): I have dealt with many atmospheric models and I find deficiencies in these models troublesome. When the authors make a statement like "we cannot completely attribute the model-observation discrepancy to the model deficiency" how much can one actually attribute the discrepancy to incomplete or defective model representation of the real atmospheric dynamics? The authors need to provide sufficient proof.

The proof is that ΔCO_2 simulated by MJ98-CDTM, which has the weakest vertical mixing as shown in Fig. 4, is more largely different from the observed one in the FT. As the reviewer pointed out, those sentences are not so clear. Therefore, we made those clearer and discussed more carefully as follows.

[Page 11, lines 9-20]

=>"Most simulated vertical gradients from PBL to FT are smaller than the observed ones for JAS, except EAS. One probable cause is a deficiency of the model vertical mixing. Actually, Stephens et al. (2007) reported that the TransCom3 models have overly strong vertical mixing from PBL to FT during boreal summer. In this comparison, however, weakening vertical mixing might not improve the results because ΔCO_2

simulated by MJ98-CDTM, which has the weakest vertical mixing as shown in Fig. 4, is more largely different from the observed one in the FT. Therefore, although the possibility of transport processes other than vertical mixing causing the model–observation discrepancies cannot be ruled out, we consider that flux uncertainty is significant to the simulated PBL-FT gradients. It is because the PBL-FT gradients were changed greatly by selection of the surface flux for JAS. To investigate transport uncertainties further, we should compare the simulated radon results with vertical radon observations (if available) but this is left for the future work."

(5) Section 3.2.3 (1st paragraph, last sentence): Please delete the phrase "more or less" and use a more acceptable word like "generally" if one has to.

Following the suggestion, we changed "more or less" to "generally". [Page 13, line 1]

(6) Section 3.2.3 (2nd paragraph, 3rd sentence): How do you know that the "model-observation mismatches" are not due to a shortcoming in the atmospheric model dynamics?

As the reviewer pointed out, it is difficult to know the causes of the model-observation mismatches. We have realized that those sentences are going too far, therefore we deleted whole the second paragraph of Section 3.2.3 and the sentence of "Another notable … dry atmosphere condition." in Conclusions.

(7) Section 3.2.3 (4th paragraph, 1st sentence): The phrase "marginally failed" needs to be quantified. Or just simply take out the word "marginally."

As suggested, we deleted "marginally".

In order to make the manuscript simpler and more understandable, we modified the figure and table captions and replace "seasonal mean variation" with "monthly mean variation" in the manuscript.

We replaced Patra et al. ACPD (2011) by Patra et al. ACP (2011).

Because the comments by the reviewers were helpful in revising the manuscript, we would like to add the following sentence in the acknowledgement.

=>" We also thank the anonymous reviewers for their valuable comments on this manuscript."