Response to interactive comment on "Assessing stratospheric transport in the CMAM30 simulations using ACE-FTS measurements" by Felicia Kolonjari et al. from Referee #1

RC = Reviewer Comment
AR = Author Response

RC: The authors compare free-running and nudged simulations using the CMAM middle-atmosphere model to ACE-FTS measurements of long-lived tracers. The analysis is generally well grounded and based on established analysis techniques, such as tracer-tracer correlation plots. Generally my impression is that the paper takes in a lot of information, making this a fairly dense read. For the future, I recommend to the lead author to break up such works into into smaller, separately publishable pieces. I don't think it would be adequate to recommend this course of action for the present paper as this is only a matter of presentation. The captions of some figures could be more detailed; for examples see below. In the comparison of models versus satellite measurements, a more thorough discussion of the effect of measurement uncertainty on tracer-tracer plots would be desirable. For example, the density plots of N_2O versus CFC-11 and CFC-12 (figures 13a and c), in the case of the satellite, are probably affected by measurement noise giving the JPDFs a fuzzy appearance. Such noise is absent in the model, making for a skewed comparison. Possibly averaging kernels of the ACE-FTS measurements could be used to define random noise to be added to the model data, making them more comparable to the measurements. If this is not practical, at least some text to this effect would be good to have.

AR: We thank Referee #1 for their thoughtful comments on our manuscript. We recognize the manuscript is long but feel that to have a complete discussion, the various components of the paper that have been included here are necessary. We will consider this feedback on future manuscripts. As per your suggestions below, some figure captions have been edited. The ACE-FTS retrieval does not routinely produce averaging kernels so these are not available for the analysis. Also, a discussion of the effects of measurement uncertainty has been added to the JPDF discussion (detailed below).

RC: Similarly, the discussion of differences in stratosphere-troposphere intrusions / extrusions mentions that resolution might factor into this comparison. At least for the detection of such structures in the data, this can be accounted for by removing small scales from the satellite data using a low-pass filter. Then the two datasets are nominally at the same resolution. This would however not address that the simulation of cut-off systems is fundamentally sensitive to numerical diffusivity in the model, causing reduced incidences of such systems.

AR: Philosophically, we have tried to approach the measurement-model comparisons in this study as directly as possible, knowing that the model is always representing a smoother version of reality than ACE-FTS does. We decided to not filter the measurements in any way so that the differences due to small scale features could be identified. If we were to apply a low-pass filter to the observations, we would not be able to answer the questions we posed. In addition to this, we would not be able to apply a three-dimensional low-pass filter on the ACE-FTS profiles (which would be

the only appropriate way to do such a comparison since the model fields are effectively smoothed across all three dimensions). The point about the finite resolution of the model is well taken and should be mentioned, but to start to modify the observations seems like a slippery slope.

P18L13-14 has been changed to reflect this sentiment:

Original sentence: "Understanding how CMAM30HR simulates this exchange assists in the interpretation of mixing effectiveness in the model and the impact of its vertical resolution on the comparisons."

New sentence: "Understanding how CMAM30HR simulates this exchange assists in the interpretation of mixing effectiveness in the model and the impact of its finite resolution on the comparisons."

RC: Regarding the differences in age-of-air between the free-running and nudged version of the model, my impression is that this is partly caused by mass non-conservation in the nudging fields, whereby artificial divergence caused by relaxation towards reanalyses causes noise in the vertical motion fields. The effect of this might be increased numerical diffusion and a reduced age. Since CMAM is based in a spectral dynamical core, one could consider, in separate experiments, to only nudge divergence or only vorticity, to try to control this behaviour.

AR: This is an interesting idea that we will pass along to the CMAM development team.

RC: On the whole, the above amounts to a recommendation to publish after a minor revision.

Minor comments:

RC: P3L3: This sentence reads a little awkwardly – modelling and observations are independent activities. How about "The BDC is well characterized in models but remains poorly constraint in obs" or so?

AR: We agree that this sentence should be rephrased.

Original sentence: "Despite significant progress in modelling, the BDC has been poorly constrained by observations (Butchart, 2014)."

New sentence: "The BDC is well characterized in models but remains poorly constrained by observations (Butchart, 2014)."

RC: P7L10: It's certainly possible to rescale the fields to construct approximations for the other tracers. But this requires further assumptions.

AR: We chose to use a parallel set of halocarbon species with adjusted boundary conditions because we felt the assumptions that would be required to construct approximations would complicate the interpretations of the measurement-model comparisons. Given the time-varying relative contributions of individual halocarbons to a particular model species in the troposphere, as air parcels enter the stratosphere and air parcels of different ages are mixed together, to untangle the contribution would require assumptions about both the mean age and the full age spectrum. It was felt that such an approach would introduce significant uncertainties.

Original sentence: "Because of the time-varying contribution of the individual halocarbons to the tropospheric concentration of the model species, it is not possible to re-scale the concentration of the model species to recover a concentration that could be compared with observations."

New sentence: "Because of the time-varying contribution of the individual halocarbons to the tropospheric concentration of the model species, the numerous assumptions that would be required to rescale the model species concentration to recover a concentration that could be compared with observations would introduce significant uncertainties."

RC: P7L20: Worth mentioning / discussing Meinshausen et al., Geosci. Model Dev., 2017 here. They have constructed boundary conditions for CMIP6 simulations that follow very similar ideas.

AR: The following sentence has been added after the sentence on P7L20:

"The application of hemispherically-defined lower boundary conditions based on observations is consistent with the proposed approach for the upcoming sixth phase of the Coupled Climate Model Intercomparison (CMIP6) project (e.g. Meinshausen et al., 2017)."

RC: P8L6: It remains a little unclear to me how you can have systematic differences between the LBCs used to constrain the simulations and the long-term observations, when the obs were used to construct the LBCs.

AR: The halocarbon LBCs were hemispheric averages obtained from the HATS network. While the data from the individual stations contributed to the hemispheric average, there are latitudinal dependencies in the surface measurement data. These comparisons indicate that differences exist but they are very small. The N_2O comparisons exhibit larger differences because the LBC was a global average.

RC: P8L27ff: Perhaps not drag but noise in the vertical motion. The w fields in the nudged and free-running model might show some differences.

AR: We agree that differences in the vertical motion between the free running and nudged simulations could certainly be contributing to the differences in the residual circulation. A figure has been provided below to demonstrate the differences in w^* in the tropical region. Our aim with including a comparison to the free running simulation was only to provide an important caveat to the results for the CMAM30 nudged simulation. While the use of a nudged simulation allows for a

time- and space-matched comparison to the ACE observations, we wanted to make the reader aware that the residual circulation in CMAM30, along with age of air and the distribution of long-lived tracers, is different than that which we find in the freely running version of the model. An analysis of the cause of these differences would be a completely separate study and is well outside of the subjects addressed here. We have modified sections of the paper where we speculate on possible causes of the differences to make it clear that the reasons for the differences are unclear at the present moment and are outside of the scope of the paper.
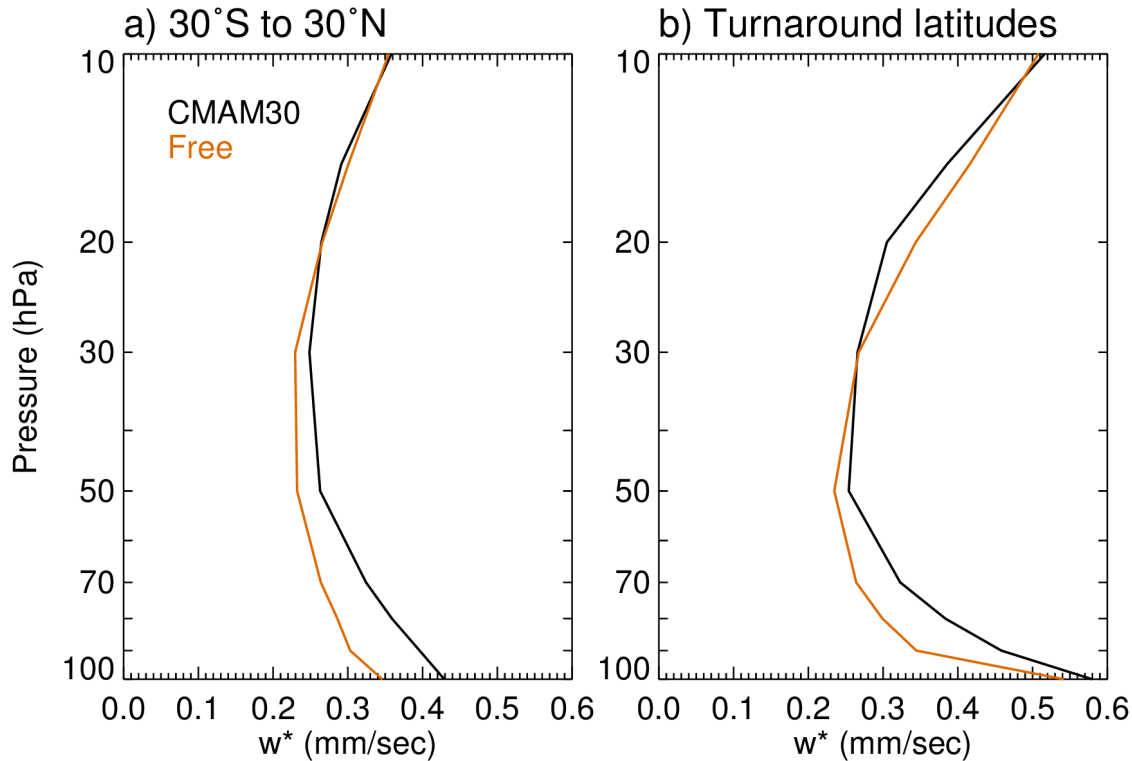
a) 30˚S to 30˚N   b) Turnaround latitudes

AR: Annual-mean has been added to this line as well as the Fig.6 caption.

Original sentence: "The zonally-averaged distribution of N_2O is presented in Fig. 6.".
New sentence: "The zonally-averaged annual-mean distribution of N_2O is presented in Fig. 6.".

Original line in Fig 6 caption: "Zonally averaged latitude-altitude distributions of N_2O."
New line in Fig 6 caption: "Zonally-averaged annual-mean latitude-altitude distributions of N_2O."

AR: We agree with the reviewer that the capability of a relatively low-resolution model such as CMAM (at T47) to correctly model the dynamical evolution of cut-off systems that produce intrusions would be a consideration for a freely running model. In this analysis, however, the dynamical fields are nudged to reanalysis fields derived from a high resolution atmospheric model (T255 for ERA-Interim). Since the synoptic scale and larger (to T21) in CMAM30 are nudged to the ERA-Interim data, there should be, though admittedly it has not been shown, a good representation of the formation and lifetimes of cut-off systems. As for homogenizing the resolution of the observations and model, we note that ACE provides high resolution (up to ~3 km) vertical profiles but only about 15 profiles a day, so these profiles are widely spaced horizontally. While it would be possible to perform vertical smoothing of the observations, it is not clear whether the smoothed observations would be more comparable with the model, as the model fields are the result of a finite horizontal and vertical resolution.

Original sentence: "Therefore, it is unlikely that there is a physical mechanism or deficiencies in the model leading to the differences observed in Fig. 11 and the differences are primarily due to the model resolution."

New sentence: "Therefore, the differences are primarily due to the finite horizontal and vertical resolution of the model, which leads to differences in the representation of stratosphere-troposphere exchange events."

RC: P34: More detail in the caption please. Which species, which network, which measurement principle, why are there these systematic differences when the measurements had been used in constructing the LBC for the model?

AR: The boundary conditions for N_2O were not derived in a special manner for the CMAM30 run. The global averages of N_2O were based upon an older IPCC Assessment Report - the A1b scenario for the 4th Assessment Report and also used for CCMVal-2. The N_2O time series uses observations only up to 2000, then it becomes a projection so there are differences to be expected there. The CFC-11 and CFC-12 measurements used in the comparisons in Fig. 1 are an updated version compared to the data that was used for the boundary conditions in the CMAM30 simulations. The minimal differences observed are due to updated values in the observations. Further detail has been added to the caption of Figure 1.

Original Figure 1 caption: "Comparison of CMAM30HR run to surface measurements, relative differences calculated as the site subtracted from the CMAM30HR simulation, divided by the average of the two, as described in the text. The differences and the uncertainties included are the mean and standard deviation of relative differences over the time series."

New Figure 1 caption: "Comparison of CMAM30HR simulations of CFC-11 (blue x), CFC-12 (green diamond), and N_2O (black circle) to the HATS surface flask network of measurements at various locations around the world. Locations of measurement sites are indicated by latitude. Relative differences are calculated as the difference between the concentration at the surface site and the lowest model layer of the nearest neighbor gridbox to the site in the CMAM30HR output,

divided by the measured concentration. The relative differences were calculated based on the monthly averaged observations and simulations. Shown here are the mean of the differences between May 2004 and June 2010 and the error bars indicate one standard deviation of the mean of the relative differences over the time period."

RC: P46: Here's where I think measurement uncertainties make this a skewed comparison. The model output would ideally be folded with the averaging kernels and a-priori assumptions used in the retrievals of the ACE-FTS measurements before comparison with those measurements.

AR: We agree that it would be ideal to incorporate averaging kernels and a-priori assumptions in measurement-model comparisons. However, we are unable to do this because there are no averaging kernels available for the ACE-FTS dataset because we do not use optimal estimation in the retrieval process, and so we have no averaging kernels.

To address the measurement uncertainty concerns, the following text has been added to the end of section 3.1.2:

"Hegglin and Shepherd (2007) have shown the impact of ACE-FTS measurement uncertainties in joint PDFs by comparing the full model output, subsampled model output, and ACE-FTS measurements. They found that there was larger variability in the ACE-FTS joint PDFs compared to those of the subsampled CMAM output."