Response to reviewer 1

We thank the reviewer for raising several important questions and guiding the restructuring of our paper. Comments by the reviewer are in italics.

at the moment the Methods section is too long and the Results section still reads like methods/technical details.

We have trimmed the existing Methods section and moved method-related paragraphs that were in inappropriately in the Results section to the Methods section.

First few paragraphs in Page 1716 do not sound like results.

The reviewer is correct. We remedy this by renaming Sect. 3 “Results and discussion”. The first paragraphs of Section 3 are the TH results and the related discussion. We will create a sub-section named “3.1. Retrieved tangent heights”

page 1716, line #26-27 : Need to mention what is the actual source of this data? May be even discuss of their salient features supporting your idea of choosing G. C. Toon’s data over others.

This is not data. This is an empirical model used as *a priori* for TCCON retrievals. It was chosen for its simplicity. It is a parameterization which requires inputs of year, day of the year, altitude and latitude. It includes a latitude-dependent seasonal cycle, a 0.5%/year increase, an age-of-air correction which is stronger in the stratosphere than the troposphere, and a decrease in concentration above 90 km. The seasonal cycle mimics the true seasonal cycle in terms of amplitude and latitude of amplitude extrema with a minimum amplitude at the south pole, and local maxima in amplitude at 72°N and 31°S of 4% and 0.4% of opposite phase. We now provide the following information, including a reference which addresses which data was used to develop the simple model:

…a priori model (Keppel-Aleks et al., 2012) for the Total Carbon Column Observing Network (TCCON, Wunch et al., 2011) retrievals.

page 1717, line #18 : Bit of an odd formulation, first say reliable and then implausible!

The reviewer is paraphrasing incorrectly. The original manuscript says “generally reliable … on rare occasions … implausible”. We don’t find this odd. Observed spectra in successive sunrise occultations at the same tangent height can differ due to random noise and systematic issues. One suspected issue has been identified where the total transmittance becomes too low at low tangent heights.

page 1719, line #1-4 : cannot understand the link here.

The reviewer’s comment is vague so we are struggling to understand it. The paragraph (p1719, L1-4) describes the HIPPO CO2.X data product, which is the official HIPPO data product and is used for validation in our paper. HIPPO consists of a suite of instruments. The CO2.X data product is a composite with measurements primarily from the Quantum Cascade Laser Spectrometer (QCLS). When the QCLS did not measure, CO_2 measurements from a second instrument where used. This paragraph is moved to the Method section into a sub-section 2.6 relating to the validation data. The post-processing data filters sub-section was also moved to the method (now Sect. 2.5).

page 1719, line #5 : basically 5 out of 12 month, if you count by number of days each of the 5 campaigns cover about 30 days each

Our statement is also correct but perhaps misleading because nine full months were not sampled. Even though the sampling in some of the months may be limited, we only use latitude bins that were sampled during that (fraction of a) month. The HIPPO data in months when only a small number of days were sampled does not appear biased relative to adjacent months. We now write:

The HIPPO campaigns produced ~150 days of flight data, sampling 9 of 12 calendar months with good latitudinal and altitudinal coverage.

page 1719, para 2 and para 3 : these are not results. more inappropriate when the Methods section is long

The reviewer is correct. These paragraphs will be moved to Section 2.6.

page 1719, line #18ff : Why? Is there much difference between the 3 instruments? It could be actually interesting to discuss how different are the 3 data are - e.g., are the differences smaller/greater than uncertainties of your retrievals?

The reviewer raises a question already of interest to us. We were paying some attention to this issue and found qualitatively that the CARIBIC seemed to be slightly high, but did not address it in the original manuscript because most of the validation data comes from CONTRAIL and HIPPO, which agree quite well, much better than ACE-FTS with either in-situ data product. To be more quantitative, we sampled the in-situ zonal monthly mean profile data only for latitudes, altitudes, and months in which ACE-FTS retrieved CO_2 (clear and "cloudy" data sets). We find that there are no spatiotemporal coincidences between HIPPO and CARIBIC. For HIPPO versus CONTRAIL, we find HIPPO is lower on average by 0.37±0.04 ppm (1 standard error, N=138). For CARIBIC versus CONTRAIL, we find CARIBIC is higher by 0.62±0.12 ppm (1 standard error, N=183). The larger standard error of the mean difference with CONTRAIL is also driven by more variability in the CARIBIC data than the other two in-situ data sets.

We have added the following information at P1719,L18:

"The differences between these three instruments at ACE-FTS CO_2 geolocations is <1.0 ppm. CARIBIC appears to have slightly more variability than the other in-situ data sets. HIPPO is preferred because of its latitudinal and vertical coverage."

page 1720 (whenever uncertainties are discussed) : isn't 1% bias is significant for CO2 for improving our understanding of the carbon cycle? It is all the more worrysome because the biases are latitude and height dependent. Please provide justifications what uncertainties are accetable for the usage of your CO2 data product.

Note the overall bias versus in-situ data is <1% (=-1.7±0.4 ppm). As this is only the first version of the global ACE CO_2 profile retrieval, we expect biases will decrease in the next version. A 1% bias in the CO_2 data which is latitude and height dependent makes improving our understanding of the carbon cycle difficult but certainly not impossible. It limits the applications for which the data can be used. Acceptable biases, like the actual biases, are also dependent on height and latitude. Acceptable bias also depends on the application. For example, if one is interested in the interhemispheric gradient at 8 km, one must consider the relative biases in both hemispheres at this altitude. If one is interested in small spatial scale variability, the acceptable bias is at most <4.3 ppm as discussed in Sect. 4 (Conclusions). For some applications, a height or latitude-dependent bias does not matter at all. One such example is the long-term trend. If a bias exists at a certain latitude or altitude, so long as this bias is not time-dependent, it will cancel out when looking at the temporal trend. In the stratosphere, acceptable biases are smaller because the variability is less than in the troposphere. Also, in the southern hemisphere, where the seasonal cycle is small and, as a result, latitudinal gradients are small year-round, the acceptable latitudinal bias is small. In general, the acceptable biases are currently slightly smaller than our actual biases, meaning that we cannot effectively capture some aspects of the carbon cycle. Sect. 4 discusses the maximum natural variation in different dimensions (latitude, altitude, time) which sets an upper limit on the 'acceptable bias':

1) biases between 7 and 12 km should be <8 ppm
2) relative biases (i.e. differential bias) between latitudes at 5.5 km should be <9.2 ppm

The validation indicates that the actual ACE-FTS bias (using all validated heights) between northern mid-latitudes and southern high-latitudes is 5.4 ppm, whereas the actual gradient (based on aircraft data) is 6 ppm at 10 km in April. This allows us to quantify the interhemispheric gradient at a time of year when biospheric uptake is nearly minimal and thus to get a sense of the impact of interhemispheric differences in anthropogenic emissions on the CO_2 concentrations near the tropopause. This is what is shown in Figure 14 (left panel) of the original manuscript. The right panel shows the reduction in the interhemispheric gradient during boreal summer.

page 1721, item#4 in Section 4: I think, it is more important if you could tell the readers how important is the choice of a priori was - not interesting to state that a prioris did not come of from a CTM. I see no scientific merit.

The main point was to say that the retrieval is not sensitive to the *a priori*. We muddled this message. We now write:

"4. no use of a priori information for regularization."