We would like to thank referee #4 for the effort in evaluating, commenting, and thus improving our manuscript and for the time spent to perform this work.

This paper calibrates the long term NDACC record using XCO2 a model derived from TCCON, Mauna Loa, and the CarbonTracker model. The error estimates from these comparisons are on the order of 0.4% (1 ppm), which is large for CO2 error, but small compared to variations in other remotely sensed species. The authors are aware of the sensitivity of the NDACC and have accounted for it through the application of the NDACC averaging kernel. Their model is validated with the TCCON system and then used in the calibration of the NDACC sensors back much longer than the TCCON record, using input from Mauna Loa.

We report and discuss the averaging kernel (Fig. 4). However, for our purpose we need a simple NDACC/FTIR product that well captures XCO2 variations on timescales larger than one month. In this context the sensitivity of our NDACC/FTIR XCO2 retrieval is not a problem, since it is clearly sufficient for our purpose (see, for instance, Figs. 6, 7 and B1).

General Comments

The differences between retrieving with two very different priors (NDACC (TCap) and NDACC(WACCM), red and black, respectively, showing a blunted seasonal cycle for the flat prior in figure 1) plus the analysis in section 5 shows that the NDACC product has almost no sensitivity to the surface and is driven by the mid-Troposphere. Comparing directly to TCCON will rely heavily on the a priori information for the surface to 5 km. For this reason, direct comparison to TCCON in Figures 6 and 7 are not of much interest for validation and can be confusing to the purpose of the paper. For example, figure 1 makes the sensitivity point shown in Figure 7. The primary validation is the indirect validation of NDACC through comparison of NDACC to the model, and model comparison to TCCON shown in Figures 5 and 8. Also the analysis in section 5 should be referred to in regards to figure 1, which is otherwise very puzzling.

Yes, the NDACC spectra have less sensitivity in the lower troposphere if compared to the TCCON spectra. This is clearly shown in Fig. 4. We can stress this even more in the manuscript.
However, with our NDACC XCO2 product we do not aim on capturing day-to-day CO2 variability. We want to capture variabilities on timescales longer than a month. On these timescales the model is valid and we can use the difference between the NDACC XCO2 product and the XCO2 model as proxy for the NDACC data stability.

1) Figure 6 is the most valid comparison between NDACC and TCCON since it is for a NDACC retrieval that applies the TCCON strategy (application of a strongly varying apriori that follows the typical annual cycle and trend). It shows that the TCCON and the NDACC XCO2 retrievals reveal very similar annual cycles and long-term trends. This means that NDACC has a similar capability than TCCON in capturing the XCO2 variations that take place on the timescales above one month (which are the timescales that are of interest in our study). Interesting in this context is Fig. B1. It documents that the NDACC retrievals reveal the same deviations from the TCCON a priori model as TCCON retrieval. Assuming that the TCCON a priori model is a first guess of the XCO2 variations NDACC can reveal deficits in this first guess in a similar manner as TCCON. Fig. 6 and Fig. B1 provide the overview of the consistency between our NDACC XCO2 and the TCCON XCO2 product.

2) Figure 7 shows the comparison for a rather simple NDACC retrieval (fixed a priori, not varying TCCON a priori). We show that even with this simplification, our NDACC product can follow the long-term XCO2 trend and the seasonal variation. However, here the seasonal variation is dampened. This dampening is consistent for all the sites and can be easily considered. In conclusion: our simplified NDACC retrieval can also capture the XCO2 variations on the timescales that we need for using it as stability proxy. This is an essential outcome of our study.

I would additionally stress why CO_2 is chosen for validation in the paper. The reader would also be interested in what the other species produced by NDACC are, and what their variability is.

CO_2 is chosen because atmospheric CO_2 is rather stable. The weak variability of XCO_2 can be modelled within better than 1% and it can thus serve as good stability reference. O_2 or N_2 would be even better suited, however, they are not accessible (O_2) or not sufficiently well accessible (N_2) by the NDACC spectra.

Specific Comments

Abstract, line 11. "As XCO2 model" change to "An XCO2 model".

See answer to referee #3: will be changed to: Our XCO_2 model is a simple regression model fitted to CarbonTracker results and the Mauna Loa CO_2 in-situ records.

Section 2.3, Eq 1, line 18. This equation should specify the units, e.g. 'where Ps is the surface pressure' change to 'where Ps is the surface pressure in Pascals'. The first term should have Avogadro's number multiplied and the molecular mass should be in kg/mole. H2Ocol in the second term should be specified as molecules/m2.

Will be changed to: "where P_s is the surface pressure in Pascals, m_{dryair} the molecular mass of the dry air (~$28.93 \cdot 10^{-3}/N_A$ kg/molecule), m_{H2O} the molecular mass of the water vapour (~$18 \cdot 10^{-3}/N_A$ kg/molecule), H_2O_{col} the water vapour total column amount (in molecules/m^2), and N_A is Avogadro's constant (~$6.022 \cdot 10^{23}$ molecules/mol). H_2O_{col} is a result….."

Section 4.1 line 18 First, the rationing for our NDACC product is made by DPC (Eq. 1) and for the TCCON product by O2 (Eq. 4). I think should be "ratioing" but could be worded better.

Wording will be improved.

Conclusions This process was done in order to calibrate NDACC spectra. If I understand, the bias is fine (as it could be explained by spectroscopy and/or O2 ratioing); the stability of the difference is the important metric? This should be made more clear.

ok

Figure 4. For a column measurement, the y axis is better in hPa as 50% of the column is between 0 and 5 km, whereas on this figure 0-5 km is only about 12% of the plot. What does a column averaging kernel value of 4.5 mean? This should be explained, as the ideal column AK would have a value of 1 (as defined in Connor et al., 2008 Eq. 8). A value of 4.5 says that differences between the true value and the prior at 30 km would by multiplied by 4.5. Is this trying to make up for the fact that the AK is 0.25-0.5 between 0 and 5 km?

We'll add the hPa on the right y axis. As suggested, we'll add an explanation of the column averaging kernel values.

In Figure 5-8, middle plot, are the axes per mil or ppm or percent?

All middle plot axes of Fig. 5-8 are in percent (unfortunately missing in Fig. 6 but will be added)

Figure 7, 8 should refer back to figure 5 rather than Figure 6.

ok