Response to Anonymous Referee #1

We thank Referee #1 for their thoughtful comments and questions.

Revisions to the text are necessary to address comments from both referees and the Editor's comments and we have prepared a revised manuscript.

Below we will respond to the comments and questions of Referee #1.

Comment #1:

The two things that are missing are:

1) predicted errors propagated from the interferent errors, These errors are calculated empirically and should also be compared to their predicted impact, at least for temperature and water. These errors can be calculated using Jacobians, following Connor et al. (2016) Eq. 6, or Worden et al. (2004) (second term of Eq. 18).

Response:

This was done in Appendix D where we show the smoothing, noise, interference, and retrieval errors for retrievals with real spectra (Fig. D8). The interference error (which includes contribution from the water vapour profile scaling) is the smallest contribution, but it does not include the effect of temperature, which is not retrieved. GFIT does not compute the derivative of the spectrum wrt temperature. We would expect the interference error to increase after the implementation of a temperature retrieval.

We did not include estimates of the forward model error, which would require estimating a covariance matrix and computing the Jacobian for each of the non-retrieved forward model parameters (e.g. each spectroscopic parameter, temperature, surface pressure, etc.), that would correspond to $\boldsymbol{S_b}$ and $\boldsymbol{K_b}$ in Connor et al. 2016. The effort of deriving $\boldsymbol{K_b}$ is equivalent to that of implementing a temperature retrieval in GFIT2.

Figure D8 was updated to present results from the 8 days with Lamont data, with the mean+-standard deviation of each error components instead of just showing results from 2012-01-14. And the following was added at the end of Section 3.2.2 to bring more attention to this analysis:

"

Additional analysis of the vertical sensitivity of the retrieval is presented in Appendix D, as well as a decomposition of the retrieval error into the interference, measurement noise, and smoothing errors as shown in Fig. D8. The interference error is the smallest (<0.5%) contribution but does not include the effect of temperature errors. The measurement noise error decreases from ~1% at the surface to ~0.2% at pressures less than 0.6 atm (> 5 km), and the smoothing error dominates and decreases roughly from ~3% at the surface to 1% at the top of the atmosphere.

"

Comment #2:

2) The performance of the profile retrieval with real spectra is only reported as a column result (e.g. Fig 13). The standard deviation and bias versus AirCore should be calculated at all pressures and summarized in the paper.

Response:

All the CO_2 profiles and difference profiles shown in Section 3 are for the average of retrieved profiles from spectra meeting the coincidence criterion with AirCore. The right hand panels of these figures show the differences with AirCore. Fig. 8 and 9 show results of retrievals from real spectra. However these are indeed missing the standard deviation, both from the set of profiles averaged, or as the square root of the diagonal elements of the retrieval covariance matrix. The latter is shown in Fig. D8, which was updated to show statistics over the 8 days with AirCore profiles over Lamont, instead of just on 2012-01-14. Although it is reported there as a fraction of the a priori uncertainty and may not be straightforward to interpret in parallel with the profile figures, especially since for real spectra the prior uncertainty is described by Equations 1-4 from section 2.2. A second panel was added to Fig. D8 with the a priori uncertainty profile itself. These errors were summarized as shown in the response to Comment #1.

The retrieval error ranges from ~40-100% of the a priori uncertainty, with that fraction roughly increasing with altitude, and the a priori uncertainty decreases with altitudes from ~5% (~20 ppm) at the surface to ~1% (~4 ppm) at 5 km altitude and above.

The differences between the retrieved CO2 profiles and AIrcore profiles as shown in right panels of Fig. 8 and 9 are larger than the retrieval errors (obtained with the square root of the diagonal elements of the retrieval covariance matrix).

Comment #3:

Line 17. To help the reader understand the scope of the paper, I would change the second sentence of the abstract to, "With these improvements, CO2 profiles were obtained from sequential retrievals in five spectral windows with different vertical sensitivities using synthetic and real spectra."

Response:

The text was updated as suggested.

Comment #4:

Add to the paper: It was unclear to me (or maybe I missed it) what is retrieved. Is surface pressure? Water profile? Water scaling? Temperature profile? Temperature scaling? This could be listed at the start of methods.

Response:

Table 2 presents all the retrieved parameters. In GFIT the state vector is only made of retrieved parameters, the caption of Table 2 was updated to:

“Components of the state vector in GFIT2 profile retrievals. These are all the retrieved parameters.”

Comment #5:

Line 82. I would add, "... and would also allow TCCON to be used for validation of TIR satellite products, e.g. from AIRS and GOSAT/2, and vertically resolved NIR GOSAT and OCO-2 experimental products."

Response:

The text was updated as suggested.

In addition to the suggested text, a reference to AIRS was added on Line 41 where the other satellites are first mentioned.

Comment #6:

Line 117. I would add a sentence saying, "OCO-2/3 and GOSAT/2 use the Weak1 and Strong bands as well as a band centered at 0.765 um not used by TCCON."

Response:

Some TCCON stations with Si detectors measure the O_2 A-Band, although it is not part of public data releases and is indeed not used for deriving Xgas products. The OCO-2 window centered near 1.61 μm corresponds to the TCCON1 window rather than the Weak1 window.

The text was updated as "OCO-2/3 and GOSAT/2 use two windows comparable to the TCCON1 and Strong windows to retrieve CO_2, and use the O_2 A-band (centered near 13158 cm^{-1})."

Comment #7:

Line 147. The sentence, "We see no advantage to fitting noncontiguous windows in parallel, rather than in series, and then averaging the results." is incorrect. Mathematically these are very different. The way to fit sequentially is called a sequential update and discussed in Rodgers, 2000 or Dudhia et al., 2002 and involves setting the constraint for the next step to the error covariance from the previous step.

Response:

We recognize the wording used may be confusing. By "sequential retrieval" we mean that retrieved parameters are obtained from each spectral window separately, as opposed to a joint-band retrieval. This is different from the "sequential update" method described in Rodgers, 2000.

As described in Rodgers, 2000, the "sequential update" method does not present advantages in the n-form formulation, which is the one used in GFIT2. This should be clear from the description of the GFIT2 implementation presented in Appendix B. What Appendix B does not describe is the detail of the implementation of the "inversion step" behind equation B2.

GFIT2 uses the algorithm described in Rodgers 2000 "Which formulation for the linear algebra?: The n-form" (section 5.8.1.1) but for the Levenberg-Marquardt method and using the HFTI algorithm (Lawson and Hanson, 1974) for obtaining matrix pseudoinverses by solving problems of the type $Bx = I$, with I the identity. The HFTI algorithm is also used to solve for the state update (equation B2 in the preprint).

Computational efficiency was not a concern for this study. By "advantage" we mean an improvement in the precision and accuracy of retrieved quantities, or a better vertical resolution. If the problem is linear enough, like is the case for CO_2, doing retrievals in each window separately and combining the results in

post-processing should be equivalent to a joint-band retrieval but with the added possibility of diagnosing window-specific issues.

The text was updated to:

"In TCCON post-processing the total columns retrieved from different retrieval windows (CO_2 from the TCCON1 and TCCON2 windows, for example) are averaged after removing window-dependent multiplicative biases, using retrieval errors as weights."

Comment #8:

Line 189. It would help the reader to have the value for sigma listed for 100, 500, 200, and 1 hPa.

Response:

The a priori uncertainty is defined on an altitude grid and will thus vary slightly at given pressure levels. The following Figure R1 of sigma with pressure was added in section 2.2, using the pressure grid of the a priori used for each day with Lamont spectra.

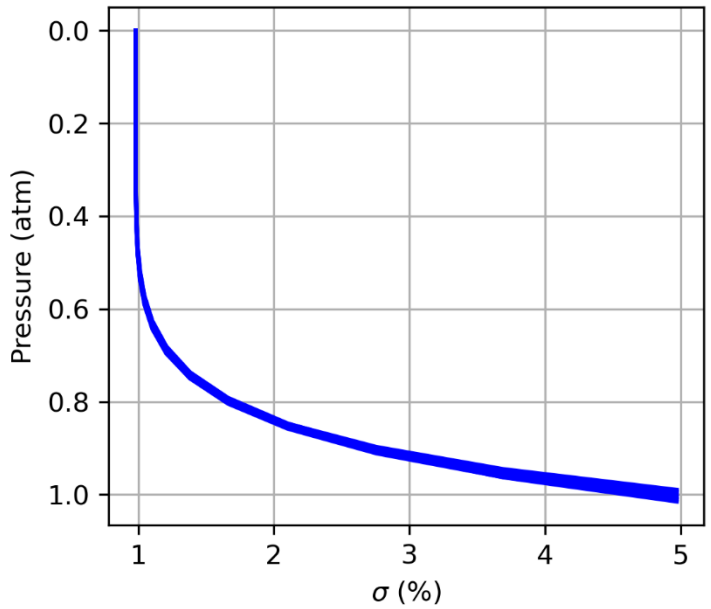

Figure R1: A priori uncertainty profiles for each of the 8 dates presented in Table 3. These are defined by Eq. (1). Since σ is defined on an altitude grid, it varies slightly with pressure.

Comment #9:

Line 194. A 2 km length scale seems very narrow, particularly above 4 km. Just a comment.

Response:

The ensemble of ObsPack aircraft profiles over Lamont used to determine the length scale are typically between 0.5-5 km. The full width at half maximum of the rows of the correlation matrix built from these profiles was ~2 km, and we use the same value for the rest of the profile, which might indeed be

inadequate above 5 km. Using a larger length scale above 5 km would increase the CO_2 profile deviations from the truth in profiles presented in section 3.2.1

Comment #10:

Line 228. Add a sentence, "The DOFS are shown in figures 3-6, and 8-9."

Response:

The text was updated as suggested.

Comment #11:

Figures 3-5. These errors should be compared to the predicted errors.

Response:

This paragraph was added before section 3.1.1:

"The total retrieval random error for the retrievals presented in this section is ~4.5% (~18 ppm), the contribution of random noise is ~0.8% (~3 ppm), see Appendix D for definitions of total and measurement noise errors. When the deviations from the truth are larger than the a priori uncertainty (~20 ppm), it means the perturbation applied has a severe effect on the retrieval. Of course this can be mitigated by using a stronger a priori constraint or a measurement covariance matrix that reflects expected systematic errors, and not just random noise, but always at the cost of reduced sensitivity to CO_2 too. The goal here is to estimate the relative effect of different kinds of expected systematic errors on retrieved profile shapes. Stronger constraints can only reduce the amplitude of the deviations from the truth, but the same structures would remain. When the perturbation to a parameter other than CO_2 results in deviations from the truth much larger than those presented in Section 3.1.1, it means that errors in that parameter will dominate the variability in the retrieved CO_2 profiles regardless of the retrieval constraints."

Comment #12:
Figures 3-5. Question that might be out of the scope of this paper: do these errors affect TCCON XCO2? In other words can the same tests from Figures 3-5 be done for the standard TCCON retrieval?

Response:

Although the results were not shown in this study, each of the sensitivity tests were also performed with scaling retrievals. An upcoming paper by Laughner et al. following the GGG2020 public release will include details of the effect of different perturbations on new TCCON Xgas products. Similar sensitivity tests have previously been presented for TCCON Xgas products in Wunch et al. 2011 and in the GGG2014 documentation (Wunch et al. 2015).

The effect of the sensitivity tests of Section 3.1 is shown in Figure R2. It shows the effect of each perturbation on the retrieved XCO_2 in the following way: $\Delta Bias$ is the average (for the 8 dates with AirCore profiles over Lamont) of the absolute change (relative to the unperturbed case) in the absolute difference between the retrieved and AirCore XCO_2. The labels of the horizontal axis of Fig. R2 indicate the perturbation applied. Thus, when a "Profile retrieval" bar is larger than a "Scaling retrieval" bar, it means

the XCO_2 retrieved from the profile retrieval is more sensitive to the perturbation considered than the scaling retrieval. The perturbations are labelled A to L.

It shows that XCO_2 from profile retrievals is as expected less sensitive to the CO_2 a priori profile (perturbations B and C), but is also more sensitive to temperature errors (perturbations G-H-I) except in the Weak windows. All windows are relatively insensitive to the H_2O perturbation considered (perturbations D-E-F). Perturbation L shows the effect of using the Voigt line shape to fit a synthetic spectrum generated with the qSDV+LM line shape. There is no change in the Weak2 window as it does not have qSDV+LM line parameters implemented, and little change in the Weak1 window as qSDV+LM is only implemented for the interfering CH_4 lines in that window. Perturbation L illustrates how using the Voigt profile was the most important source of mismatch between retrieved and true XCO_2 in the TCCON windows.

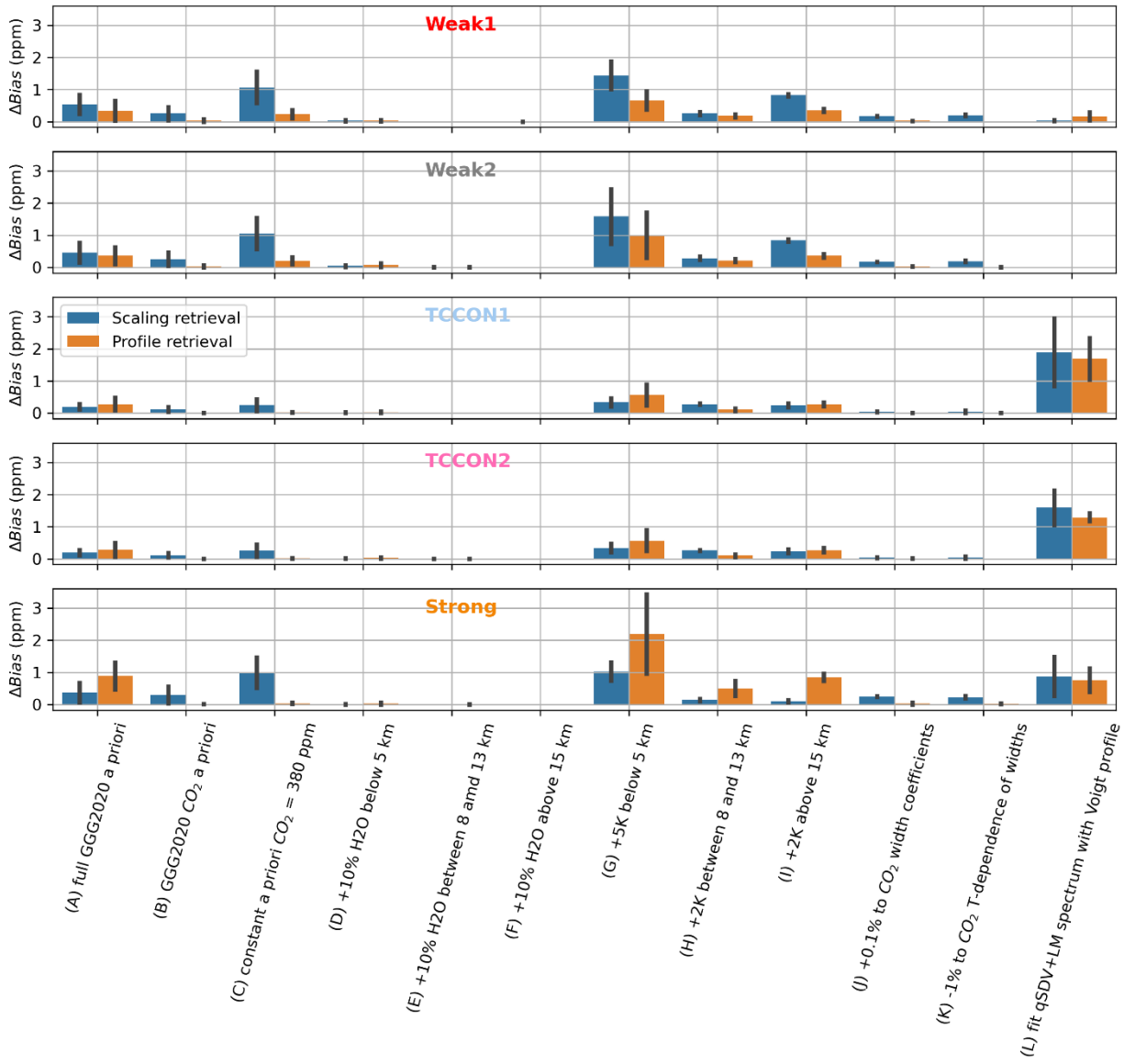

Figure R2: Absolute change in the absolute difference between the retrieved and true XCO_2 relative to the reference case (fitting a synthetic spectrum using the same a priori used to generate it), each bar is the average of that quantity from the 8 days with AirCore profiles considered in the study, and the error bars are the corresponding standard deviations. The labels of the horizontal axis indicate the kind of perturbation applied before fitting the synthetic spectrum.

Comment #13:
Figure 8. I do not understand how the result can be up to 50 ppm off when the a priori is set to 6 ppm

Response:
By "when the a priori is set to 6ppm" I assume "a priori uncertainty" was meant. The a priori uncertainty is ~5% (20 ppm) at the surface for the retrieval results of section 3.2. The response to **Comment #8** should also address this comment.

Comment #14:
Table 5. My note: I have not found information content to be useful because it is not aware of many of the systematic errors in the system.

Response:
Like the averaging kernel, the DOFS give information on the retrieval sensitivity to CO_2 in the absence of systematic biases, this is useful information only on the potential of the profile retrieval.

Comment #15:
Figure 13. My note: The variability of the results for xco2 for the different bands was surprising.

Response:

The variability between different lines is caused by window specific biases. And the error bars of ~4-5 ppm are that large, even for the scaling retrievals, because the retrieval errors were not scaled like is done in TCCON post-processing. This scaling of the retrieved error is done to better reflect the variability of XCO_2 in consecutive spectra, since the retrieval error is an overestimate of the random error due to the presence of systematic residuals. The scaling is illustrated with a set of GGG2014 TCCON data from Eureka in 2018.

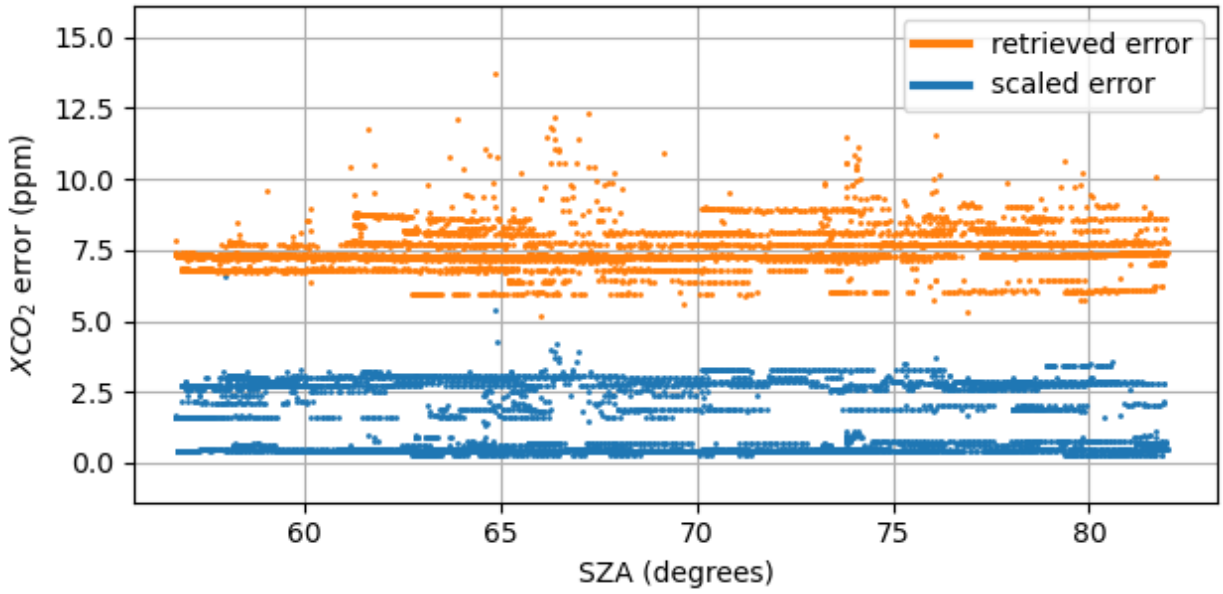

Figure R3: XCO_2 error derived from the retrieval error (orange), and after a scaling is applied to better reflect the XCO_2 variability from consecutive spectra (blue), example from Eureka 2018 GGG2014 TCCON data.

L. Lawson and R. J. Hanson, Solving Least Squares Problems, Prentice-Hall, Inc., 1974, Chapter 14.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V. and Wennberg, P. O.: The Total Carbon Column Observing Network, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 369(1943), 2087–2112, doi:10.1098/rsta.2010.0240, 2011b.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G. and Wennberg, P. O.: Documentation for the 2014 TCCON Data Release (Version GGG2014.R0). CaltechDATA, doi:https://doi.org/10.14291/tccon.ggg2014.documentation.r0/1221662, 2015.