Authors' Response to Referee #2

Dear Robert Nelson,

We'd like to thank you for this review with many very interesting follow-up thoughts. Thank you also for starting to apply the correction to real OCO-3 data which is very much appreciated. Please find below our point-by-point responses and proposed manuscript changes (in blue) to your comments.

Major changes include:

- Improvements in the methodology by using a bilinear interpolation, which improves the results especially for small altitude variations, but changes most of the figures in the manuscript
- We added all regions and more statistics to Table 3
- We added a recommendation how deal with retrievals that retrieve the surface pressure in the NIR
- Improvements in the method description and a more cautious approach in describing the results, e.g. with respect to scattering

In addition, we found an error in Figs. 21 and 22 where we stated that GOSAT had a spatial sampling size on the order of 0.5 x 0.5 km^2, which we corrected to approx. 10 x 10 km^2.

Kind regards and on behalf of all authors,
Michael Weimer

General comments:

The manuscript entitled, "Importance of subpixel Earth surface reflectance and altitude for atmospheric trace gas retrievals from passive satellite instruments" investigates the impact of subpixel albedo and altitude heterogeneity using real albedo and altitude datasets combined with simulated NIR/SWIR observations designed to estimate XCO2. The problem is that typical XCO2 retrieval algorithms assume the altitude of a given footprint is simply the mean altitude of the DEM (e.g., Copernicus GLO-90) when in reality it is weighted by the surface brightness of the scene. They combine Sentinel-2 30 m albedo data with the Copernicus 90 m DEM and the FOCAL GHG retrieval to investigate the magnitude of this effect. They first look at an idealized case, then move on to real-world cases (the Belchatow power plant, Black Forest, Mont Blanc, and Germany). Deviations from the true XCO2 can be several ppm depending on the scene, but it can be mostly corrected by using the albedo-weighted altitude within the retrieval itself. They also developed two flavors of a post-processing correction, which are almost as effective.

Overall, the manuscript is well-written, and I recommend publication after my minor comments are addressed.

Response: We thank Robert Nelson for his detailed review and recommendations. Please find our responses to the specific comments below.

P4 L94: "It should be noted that there are approaches to simultaneously retrieve the surface pressure using the NIR band of GHG satellites"

A few prominent retrieval algorithms retrieve surface pressure. This manuscript could benefit from a brief description on a potential correction method, which would make this research more applicable to those retrieval teams. Because XCO2 is the CO2 column divided by the O2 column (dry air assuming we know the O2 concentration), I think you can simply correct the CO2 column using either the WCO2 or SCO2 dz and correct the O2 column using the NIR dz. You could then convert the dzs to dps using the hydrostatic approximation and combine these corrections to adjust XCO2.

Response: We agree that it is beneficial to add such a recommendation to the manuscript, but there is one major issue we see with this approach: As the O2 and CO2 columns are derived from different bands, they (in general) have different surface altitude / surface pressure values when weighting with the subpixel albedo. For instance, at Belchatow we found that a higher altitude is needed in the NIR, but a lower altitude is needed in the SWIR bands. That means that the O2 column includes a different number of particles than that corresponding to the CO2 column. We think that this will lead to a mismatch and rather recommend a combined correction for XCO2 or even better to use the albedo-weighted average of the surface altitude / surface pressure directly in the retrieval as suggested in the current manuscript. On the other hand, it should be easy to convert the weighted z values in the post-processing equation to pressure differences. We highlight this recommendation in the revised manuscript at the end of Sect. 4.4:

"Some retrieval algorithm retrieve the surface pressure as part of the retrieval.

Equation (30) can be converted to include pressure differences by first converting the $z_{s,weighted,i}$ *values to their differences to the mean altitude* $\Delta z_{s,weighted,i}$ *and subtracting the coefficient values from* $z_{s,mean}$ *in Eq. (30):*

$$\Delta XCO_{2,\,fit} = -\ 7.0\text{x}10^{-6}\ \Delta z_{s,weighted,NIR} +\ 1.0\text{x}10^{-2}\ \Delta z_{s,weighted,SWIR1} + 3.2\text{x}10^{-2}\ \Delta z_{s,weighted,SWIR2}$$

$$+\ 1.2\text{x}10^{-3}\ \sigma_z + 5.1\text{x}10^{-5}\ z_{s,mean}$$

Each Δz in this equation can be converted to a pressure difference using the hydrostatic approximation for pressure:

$$\Delta z = -\ H\ \Delta ln(p)."$$

It has to be kept in mind that when retrieving GHGs we are not really interested in the O2 column in the NIR band, but the dry air column that is representative for CO2, i.e. the dry air column with the highest weight in the SWIR bands. Therefore, assuming that the surface pressure is retrieved only from the NIR band, it could also be possible to retrieve the surface pressure in the NIR and then subtract the difference of the effective altitudes between NIR and SWIR converted to the pressure difference to get the surface pressure for the respective bands which can then be used to derive XCO2. We added this after the new paragraph above:

"When retrieving the surface pressure in the NIR band, only, another possible way to get the dry-air column representative for the CO_2 *column would be to use the retrieved NIR surface pressure and use the difference of the effective surface altitudes in the NIR and SWIR, converted to the pressure difference, to calculate the surface pressure in the SWIR. This new surface pressure can then be used to determine the dry air column which can be used to compute* XCO_2*."*

I went ahead and did this for a real OCO-3 SAM over Belchatow on 18 October 2024 (OCO-3 ACOS v11.0 Lite raw XCO2 data, using ~6 months of HLS S2 data averaged together to get sufficient albedo coverage) and got results very similar to your simulations:

[see figure in the original Referee comment]

OCO-3's footprints are 1.6x2.2 km2, slightly smaller than CO2Ms. You can actually see a ring around the edge of the central mine, because this effect only occurs where there are altitude gradients.

Response: We highly appreciate this first application to real satellite data and it is great to see that similar results can be found for this OCO-3 SAM. We are currently working on the application to real satellite data, which is out of the scope of this current study. Our recommendation to improve the coverage of the albedo data, however, would be to do a multi-annual mean for a specific month rather than averaging six consecutive months in order to conserve the annual cycle of the surface reflectance (due to seasonal vegetation changes etc.). See also our response to your comment to P30 L 576 below.

P35 L655: "the subpixel albedo composition is usually measured with another instrument and might be provided at a later time."

It might be worth further discussion on the nuances of ingesting non-simultaneous high-resolution albedo measurements. In a desert, for example, S2 could make a single overpass a week later and you could probably use that albedo data to make the XCO2 correction. Over Poland, however, you might have to average 6+ months of S2 data to get enough spatial coverage.

Response: We agree and added the following sentences to line 728: "In addition, the spatial coverage of the S2 data is important because it depends on cloud-free and low-aerosol scenes. Therefore, single overpasses close to the XCO_2 measurement time might be sufficient e.g. in desert regions whereas averages over months might be needed over regions with many cloudy days."

P30 L576: "However, when assuming that we can get a good estimate of the surface albedo using climatologies, it might be better to use the albedo-weighted surface altitude directly as input to the retrieval"

I suspect this is mostly true, even if it takes several months of averaging albedo data to get sufficient coverage. It would be interesting to look at how long it takes to get sufficient albedo coverage from S2 for certain locations and how much the albedo can change during this time, but this is beyond the scope of your work.

Response: We started to create such a climatology on a monthly basis, but agree that this is beyond the scope of this study to be presented here. Our experience is consistent with your statement: It takes several months to get a complete coverage in the mid-latitudes. There is a significant change in albedo over time depending on the seasonal vegetation change, which we intend to show in a follow-up publication. For Belchatow, for instance, we see changes of the SWIR2 albedo in the surroundings of the coal mining field between 0.05 (January) and 0.30 (July), which might also be affected by snow cover.

P27 L535: "However, the surface albedo can change over time, e.g. due to snow cover"

Would this be an issue for real CO2M observations? Typically, snow-covered soundings are of poor quality and are screened out. When creating an albedo climatology, you'd also want to remove snow/ice pixels.

Response: There is no dedicated filtering for snow/ice in FOCAL-CO2M and it might be difficult to estimate in advance of the retrieval whether a subpixel region is covered by snow. It is, however, indeed likely that the snow-covered spatial samples have a poor quality, e.g. due to their low albedo in the SWIR, and are filtered during the retrieval post-processing. Because there is no dedicated filtering, we decided to keep the snow pixels in the S2 data, but we remove data with high aerosol load (Fmask > 192) or with cloud (Fmask = 2), adjacent to clouds (Fmask = 4) and cloud shadow

(Fmask = 8) contamination. In addition, we removed cases with extremely high albedo (larger than 2.0) or lower than zero in any band. We added this information to the manuscript after line 233 in the description of S2:

"The HLS data are provided with a Fmask for quality assessment. In each downloaded S2 tile, the data contaminated with high aerosol load (Fmask > 192), clouds (Fmask = 2), adjacent to clouds (Fmask = 4) or cloud shadows (Fmask = 8) are removed. In addition, each tile is filtered for high surface reflectance (α > 2.0) or reflectance lower than zero in any band. We deliberately keep the snow data, although they might have poor quality and would be filtered out, because it might be difficult to estimate in advance if a spatial sample is partly covered by snow and because there is no explicit filter for snow in the FOCAL-CO2M retrieval algorithm."

Specific comments:

P9 L239 "Note that the time and viewing geometry are different between S2 and CO2M which is not relevant for the simulated data here, but could lead to errors in later applications, e.g. due to different shadows of the mountains at the time of measurement.":

Other (rare) cases might be where it rains (and darkens the surface) over half the pixel during the elapsed time. Or there's snow present during the first observation but melts before the second!

Response: Thank you for these additional cases, which we added to this line in the manuscript: "[…], e.g. due to […] time of measurement, if rain darkens part of the surface in the time between the two measurements, or if snow falls or melts in the meantime."

P11 L285: "we define the region of interest by using a S2 tile provided by the HLS dataset, average the S2 timeseries of cloud-free surface reflectance measurements using the data of the year 2019"

Does this mean you only applied the "Cloud" mask in Fmask for the HLS S2 data? Was there a reason you didn't also apply the other masks, like "Cloud shadow"? Also see my previous point about removing snow/ice scenes.

And was there a reason you used all of 2019? Looking at the HLS S2 data over Belchatow, there are some strange high albedo values on the edges of the pits during more recent years.

Response: We used 2019 because it is the first full year with two satellites for S2 (better coverage), which is then closest to the Copernicus DEM data which are "acquired during the TanDEM-X mission between 2011 and 2015" (Copernicus Programme, 2026). We added this information to line 209 of the manuscript for Copernicus DEM:

"[…] better than 4 m and is derived from satellite measurements in the time range between 2011 and 2015"

and to line 285 for the S2 data:

"The year 2019 is used because it is the first full year of measurements of S2 with two satellites (better coverage) closest to the acquisition time in the CopDEM data."

Thank you also for pointing out that we neglected to describe the Fmask values we used for S2. As described above as response to P27 L535, we also added this information to the manuscript.

References:

Copernicus Programme: Copernicus Digital Elevation Model, 90 m resolution (GLO-90) [dataset], https://doi.org/10.5270/ESA-c5d3d65, accessed on 26-Feb-2026, 2026.

Fig. 5: it's interesting that for small area fractions at 8 km, using the albedo weighted surface altitude for NIR+SWIR1 gives slightly worse errors than not using any albedo weighting.

Response: Yes, this is true and the reason is probably that when using the average in the SWIR1 band, where the altitude-dependence is reversed compared to using the average only in SWIR2 and where consequently the altitude 8 km has the smallest error, leading to a smaller total error at 8 km. It is not exactly the case but the total error in the first row can be seen as a linear combination of the errors of the three following rows. We added the following sentence to line 361 of the manuscript:

"In addition, it can be seen that the error at 8 km is larger when applying the average only in the SWIR2 band compared to the average in all bands, which shows that the error in the first row is a linear combination of the error in all bands."

In addition, there was a small error in the description in the next line where we replaced "last" by "first row".

Fig. 13: when using the albedo-weighted average surface altitude directly in FOCAL, you still get differences up to +/- 2.5 ppm, which means you would still try to filter out much of this data, correct?

Response: Yes, this is true and we are aware that this region pushes the retrieval to its limits and should be seen rather as a theoretical case than a case actually occurring in retrievals with real data. To highlight this, we replaced "Nevertheless" in line 441 of the manuscript by:

"Although these data probably have to be filtered"

In addition, we added the following sentence to the end of the Mont Blanc section:

"On the other hand, it is clear that this estimated bias up to ±2.5 ppm is larger than the accuracy requirement for CO2M so that they will likely be filtered out in retrievals with real satellite data."

Fig. 15: for the full Germany case, by my eye it looks like higher altitudes generally have more forests, which are darker in all bands. As you mention in the Black Forest case, this is especially concerning because it could result in XCO2 errors manifesting as CO2 fluxes.

Response: It is true that the mountains in Germany are mainly covered by forests which is typical for mountain ranges with moderate peak altitudes (~1500 m) in the mid-latitudes. In addition, there is agriculture and small towns in the valleys, which usually have higher surface reflectance in both the NIR and SWIR compared to the forests at higher altitudes. We decided to add this to line 418 in the manuscript:

"Agricultural activity and small towns lead to generally higher albedo values in the valleys of the Black Forest compared to the forest regions at higher altitudes."

Sec. 4.3: "This fraction is close to zero until σ_z = 20 m":

Can you give a rough estimate of what fraction of Earth's surface, for CO2M-sized pixels, has σ_z <= 20 m? This would give context for the magnitude of the problem and inform existing science teams on how much of their data they would have to reprocess if they implemented this new altitude parameterization.

Response: We evaluated 11 days (corresponding to one repeat cycle) of simulated orbits of one CO2M satellite with the result that globally about 80 % of all samples over land have surface roughness less than 20 m. This means that about 20 % of the land data would have to be reprocessed. We added this information to this sentence in L. 527:

"[This fraction is close to zero until σ_z = 20 m] (corresponding to about 20 % of global CO2M spatial samples over land) […]"

Note, however, that the 0.5 ppm requirement used here for accuracy includes many components, including that of "heterogeneous scenes" (Buchwitz et al., 2020), which is on the order of 0.2 ppm, so that an even higher fraction of land data would be affected. Therefore, we added this sentence to the conclusions (line 650):

"We note that although 0.5 ppm was used in this study, the actual error budget for heterogeneous scenes and surface pressure is only part of the total accuracy and on the order of 0.2 ppm (Buchwitz et al., 2020)."

Buchwitz, M., S. Noël, M. Reuter, H. Bovensmann, A. Richter, J. Landgraf, P. Veefkind, and H. Bösch (2020): Study on Consolidating Requirements and Error Budget for CO2 Monitoring Mission (CO2M-REB): Error Budgets and Performance for anthropogenic CO2 Monitoring Mission (CO2M), Tech. rep., https://www.iup.uni-bremen.de/carbon_ghg/CO2M-REB_TNs/CO2M-REB_TN-3000_v2p2.pdf

Sec. 4.5: "Overall, the application as direct input to the retrieval removes the topographic structures from the retrieval results, but seems to increase the variability over homogeneous scenes, where the average altitude was already a good assumption."

Can you expand on the reasoning behind this? It's hard to tell that there's increased XCO2 variability in northern Germany in Fig. 15(f)

Repsonse: Thank you for this comment. We realized that this larger variability is due to the fact that we used a nearest neighbor interpolation for estimating the meteorological profiles from the spectrum database, which limits the analysis to an accuracy in altitude of 4 m, or approximately 0.16 ppm in XCO2. We replaced this by a bilinear interpolation in albedo and altitude with respect to its nearest neighbors in the database which reduces the variability with albedo-weighted altitude for these homogeneous cases. This changes all related figures in the manuscript and the supplementary material (Figs. 4 – 6 , 9, 11, 13, 15 – 22, A1, S2, S3, S5 – S8) and the post-processing equation. We additionally added the following description of the interpolation to line 303 of the EGUsphere version of the manuscript:

"[…] and interpolate the meteorological profiles to their corresponding average albedo and average or albedo-weighted altitudes using a bilinear interpolation in the albedo-altitude space of the spectra database."

Furthermore, we removed the part of the sentence cited in the reviewer's comment after "but seems to increase […]".

Sec. 5: "One main assumption was that the altitude dependence of the dry-air column-averaged mole fraction in the lowest parts of the atmosphere, where the profile is cut or extended, is small."

And, to a lesser degree, that extending/cutting the meteorological profiles is a valid assumption.

Response: We added exactly this wording to this sentence.

Sec. 5: "especially relevant for greenhouse gases."

And especially for CO2, of course.

Response: We included “CO2 and other” after the “for” in this sentence.

Technical comments:

P1 L5: typical defined as “column-averaged dry-air mole fraction of carbon dioxide”
Response: Corrected.

P2 L36: “to Earth’s surface”
Response: Corrected.

P11 L285: “an S2 tile”
Response: Corrected.

P11 L286: “for the 30-m pixels”
Response: Corrected.

Fig. 4 caption: “Because some of the panels have values that are close to zero”
Response: Corrected.

P15 L344: “The non-monotonous values within one fraction are a result”
Response: Corrected.

Figs. 6 & 16: might be worth putting the standard deviation of dXCO2 for each panel
Response: Thank you for this suggestion which we applied and which immediately shows the improvements using the albedo-weighted average of the surface altitude. We added this to both figure captions:
“The σ values correspond to the standard deviations of the XCO_2 values (in ppm).”

Furthermore, we added this text to line 366 related to the description of Fig. 6:
“This is also reflected in the standard deviations for both methods: With the unweighted average, the standard deviation is 0.9 ppm whereas it is 0.01 ppm when applying the albedo-weighted average of the surface altitude.”

Related to Fig. 16, we added this for the Belchatow example (line 492):
“[…] , which also reduces the standard deviation from 0.12 ppm to 0.05 ppm.”

And this for the Black Forest example (line 497):
“The standard deviations are reduced from 0.41 ppm} to 0.05 ppm.”

And for the Mont Blanc example (line 501):
“[…], resulting in a reduction of the standard deviation from 1.78 ppm to 0.39 ppm.”

Finally, we added this to the description of the example in Germany (line 514):
“[…] with a reduction of the standard deviation from 0.52 ppm to 0.09 ppm.”

P17 L364: “compared”
Response: Corrected.

Fig. 6 caption: “and the shaded ranges”
Response: Corrected.

P17 L374: “structure of Earth’s”
Response: Corrected.

P17 L375: “real numbers of subpixels”
Response: Corrected.

P18 L385: “one of the largest point sources”?
Response: We corrected this to “strongest”.

P20 L425: “The surface altitude differences are generally larger in the SWIR bands than in the NIR band.” Repeated information?
Response: Yes, we removed this sentence.

P22 L440: “or other differences”
Response: Corrected.

Fig. 16: coincidence that (b) and (c) have the same N? Or typo?
Response: The Sentinel-2 tiles have approximately the same size. (a) was a zoom into one S2 tile, therefore fewer data. (b) and (c) were generated for a whole S2 tile, and (d) are many S2 tiles. Therefore, the number of CO2M spatial samples in (b) and (c) are the same. We do not account for CO2M spatial samples at the edge between two S2 tiles in this study which is also part of the explanation why the numbers of spatial samples in (b) and (c) are the same.

Fig 19(b): z should be “zs,mean” for consistency
Response: Corrected.

P31 L594: clarify what “complete dataset” means
Response: We added “the dataset including spatial samples in Germany, the Alps and the Pyrenees as in the previous examples”.