General comments:

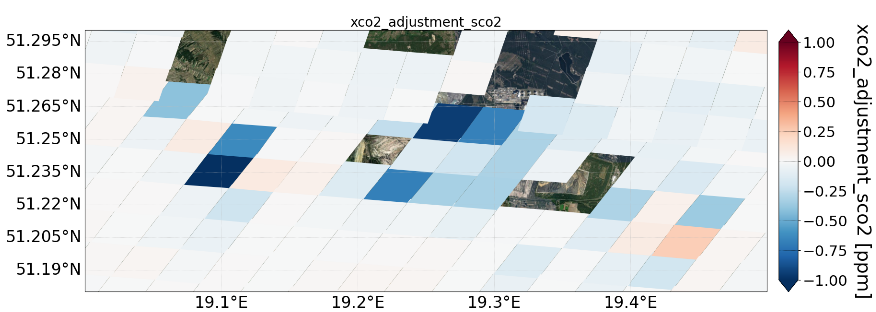
The manuscript entitled, "*Importance of subpixel Earth surface reflectance and altitude for atmospheric trace gas retrievals from passive satellite instruments*" investigates the impact of subpixel albedo and altitude heterogeneity using real albedo and altitude datasets combined with simulated NIR/SWIR 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.

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.

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:



OCO-3's footprints are 1.6x2.2 km^2, 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.

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.

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.

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.

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!

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.

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.

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?

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.

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.

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)

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.

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

And especially for CO_2, of course.

Technical comments:

P1 L5: typical defined as "column-averaged dry-air mole fraction of carbon dioxide"
P2 L36: "to Earth's surface"
P11 L285: "an S2 tile"
P11 L286: "for the 30-m pixels"
Fig. 4 caption: "Because some of the panels have values that are close to zero"
P15 L344: "The non-monotonous values within one fraction are a result"
Figs. 6 & 16: might be worth putting the standard deviation of dXCO2 for each panel
P17 L364: "compared"
Fig. 6 caption: "and the shaded ranges"
P17 L374: "structure of Earth's"
P17 L375: "real numbers of subpixels"
P18 L385: "one of the largest point sources"?
P20 L425: "The surface altitude differences are generally larger in the SWIR bands than in the NIR band." Repeated information?
P22 L440: "or other differences"
Fig. 16: coincidence that (b) and (c) have the same N? Or typo?
Fig 19(b): z should be "$z_{s,mean}$" for consistency
P31 L594: clarify what "complete dataset" means