Authors’ Response to Referee Comment #1

Dear Lu Zhang,

Thank you very much for this detailed review of our manuscript. Please find below our point-by-point responses and proposed manuscript changes (in blue) to your comments, which helped a lot to improve the manuscript.

Major changes include the following:

- 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 on how to 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

Review report on “ Importance of subpixel Earth surface reflectance and altitude for atmospheric trace gas retrievals from passive satellite instruments”

The manuscript is generally well structured. The theoretical motivation is clear, and the self-consistent FOCAL-CO2M simulations provide useful evidence that the proposed correction can substantially reduce XCO2 biases in heterogeneous terrain. The paper has clear potential for publication. However, several aspects require further clarification or additional analysis before publication. In particular, the validity range of the first-order approximation, the representative of Sentinel-2 bands for CO2M bands

Response: We thank you for all these comments and will get back to each point separately in our responses below.

Major comments

1. The physical meaning of “albedo-weighted altitude” should be stated more carefully

In a real satellite measurement the contribution of each subpixel to the detector signal is not determined by surface reflectance alone. It also depends on:

the instrument point spread function or spatial response function; viewing and illumination geometry; surface BRDF ....

I recommend that the authors make this limitation clearer in the abstract, introduction, and conclusion. The current wording may give the impression that albedo weighting alone is generally sufficient, whereas the derivation assumes a non-scattering atmosphere, Lambert surface, homogeneous spatial response, and small altitude variability.

Response: We agree with this comment and made the following changes to the manuscript:

Abstract: Line 9: "[…] by using a linearized theoretical analysis at one wavelength with many assumptions including a non-scattering atmosphere and neglecting the influence of different illumination of the detector pixel (the point spread function) and effects of the bidirectional reflectance distribution function.

Introduction: We introduce the word "albedo" at its first occurrence in line 67 as an abbreviation of the surface reflectance in the respective wavelength band. In addition, we add the following sentence after line 76:
"Here, as in the remaining part of this study, we neglect effects due to the point spread function or the bidirectional reflectance distribution function."

In the conclusions, we already clearly state that: "We also assumed that each subpixel contributes equally to the signal on the detector, while, in reality, the illumination depends on the point spread function and might be different in the center and at the edges of the detector pixel." (line 684) and based on the reviewer's comment we added "and the bidirectional reflectance distribution function" to the point spread function.

In addition, we added the main assumptions of the theoretical derivation as mentioned by the reviewer to the conclusions in line 679:
"Apart from the approximations done in the theoretical derivation (e.g. non-scattering isothermal atmosphere, Lambertian surface and small subpixel altitude variability), […]"

2 The validity range of the first-order Taylor approximation needs clearer quantification

The theoretical derivation starts from a nonlinear expression for the effective altitude and then applies a first-order Taylor expansion. This is reasonable, but its applicability depends on the magnitude of subpixel altitude variability and albedo contrast.

The Mont Blanc case clearly shows that the proposed method substantially reduces the bias, but residual errors can still reach about ±2.5 ppm, which is well above the CO2M systematic error requirement. This suggests that the first-order approximation may not be adequate in highly complex mountainous terrain.

- The authors should quantify the validity range more explicitly.
- For what values of subpixel altitude standard deviation σ_z is the method reliable?
- At what altitude second-order terms become important?
- How does the approximation error depend on albedo contrast?
- Does the threshold depend on background altitude, SZA, VZA, or scattering conditions?

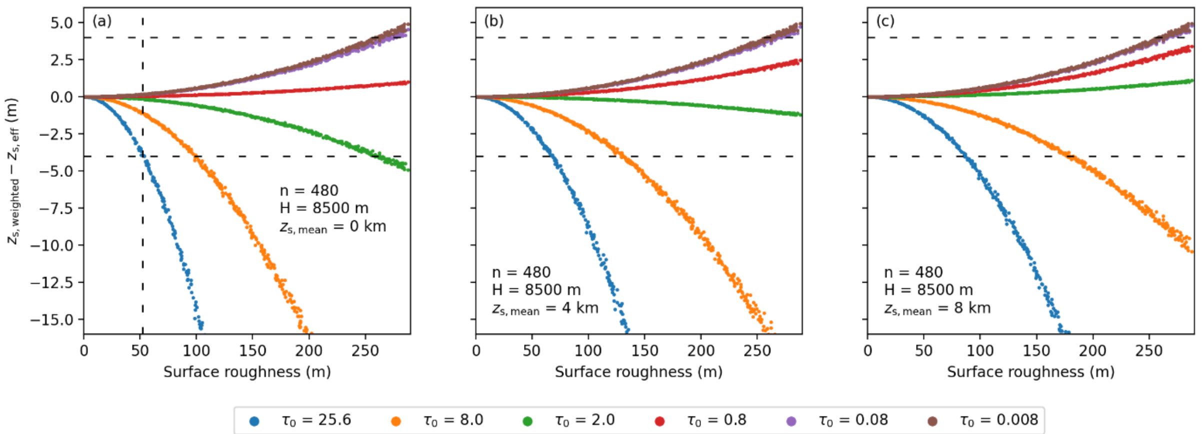
Response: [Note to the editor: This is the same response as for Major Point 2 by Referee #3] We'd like to thank the reviewer for this comment and their ideas about which kind of analysis is needed for this. It is clear that there is a limitation to the Taylor expansion of first order. In the following, we will explain why an analysis as requested by the reviewer is impossible in the real application. In summary, there are two major points we would like to make: (1) We do not know how the retrieval weights single wavelengths, especially not in advance of the retrieval itself, and (2) our results show that, although the approximation of first order is not precise at these altitudes and for these surface roughness regimes, it still leads to smaller errors compared to the unweighted average. Therefore, the results in this study suggest that it is still better to use the albedo-weighted average of the surface altitude in these challenging regions, where the approximation leads to remnant biases.

What we can do, however, is an empirical analysis after the retrieval to investigate where the large XCO2 biases actually come from. The dependence on the surface roughness (standard deviation of the subpixel altitudes within the satellite spatial sample) is already shown in the manuscript where it can be seen that there are limitations related to this parameter. We find that the albedo-weighted average of the surface altitude can be used up to a surface roughness of about 300 m. We undertook further investigations on the limitations related to the albedo and found that the largest XCO2 biases are connected to spatial samples with an average SWIR1 albedo being smaller than 0.07, which additionally depends on the atmospheric scattering. Such small albedo values will lead to a smaller sensitivity to the CO2 concentration because most of the signal arriving at the detector does not reach the surface anymore and actually comes from atmospheric scattering. This makes these types of conditions more challenging for the retrieval of XCO2. This is the case even when the subpixel altitudes are constant for the spatial sample. In summary, this demonstrates that the large XCO2 bias most probably results from the challenging conditions for the retrieval (e.g. low surface reflectance of a surface covered by snow) rather than due to the first-order approximation itself.

This comment also triggered us to improve the method for cases where the altitude variability is small. In the published preprint, we used a nearest neighbor interpolation to determine the surface altitude and related meteorological profiles, which essentially limits the accuracy of the method to the vertical spacing of the database, which is set to 4 m, i.e. an approximate ΔXCO2 of 0.16 ppm. We now apply a bilinear interpolation (in the altitude-albedo space) for the mean and effective altitudes. This does not change the main conclusions, but the variability of ΔXCO2 is now smaller. This changes all related figures in the manuscript and 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 the average or albedo-weighted altitudes using a bilinear interpolation in the albedo-altitude space of the database of spectra."

Regarding the reviewers' actual request, we respond as follows. In the initial draft of the manuscript, we had included the following figure:



It shows more or less exactly what is required by the reviewer's comment at first glance. The x-axis shows the surface roughness (standard deviation of the subpixel altitudes), and the y-axis is the difference between the weighted average surface altitude (see Eq. 23) and that from Eq. 11. These equations depend on the mean surface altitude which is different in each panel (i.e. 0 km in panel (a) , *4 km in panel (b) and 8 km in panel (c)) and on the atmosphere optical thickness at altitude 0 m, τ_0, from Eq. 11. To generate the subpixel variation we used a uniform random distribution of z values ranging from ±1, ±2, …., up to ±500 m, and we then calculate the surface roughness and add the respective point to the plot above. The number of subpixels is set to 480 and the subpixel albedo values are taken from a uniform random distribution of values between 0 and 1.*

The plots show that the error of using the Taylor expansion of first order strongly depends on the value of τ_0, and to a lesser degree on the mean altitude. For a constant τ_0, the error increases in one direction, either positive or negative, and in the case of one wavelength, as assumed in the theoretical derivation in Sect. 2, this error can be directly derived from this figure.

As the τ_0 depends on the cross section, it depends on the wavelength, pressure, temperature and H_2O concentration. The spectra measured by CO2M in the future and by current satellite instruments, such as OCO-2 and OCO-3, are not only at one wavelength but rather at hundreds to thousands of wavelengths, which are used to derive information on XCO2. At each of these wavelengths, the τ_0 will be different and the values shown in the figure cover the whole range of τ_0 values estimated from the simulated spectra used in this study. This would make the effective surface altitude dependent on each single wavelength, where we do not have surface reflectance data. In addition, it is not clear which value of τ_0 is representative for the whole spectral window. To address this issue, we would have to know which wavelength has the highest weight in the retrieval, but in advance of applying the retrieval. In addition, the τ_0 value depends on the CO2 concentration in the atmosphere itself, which is the main quantity to be retrieved by the retrieval and is not known in advance of the retrieval. As a result, we cannot assess the range of validity of the first-order Taylor expansion using this figure. At each wavelength, the error in z due to the first-order approximation will be different in the order of magnitude and even in the sign. One could consider the strictest limit, denoted by the dashed vertical line at surface roughness of 55 m related to the Copernicus DEM accuracy of 4 m, but our results demonstrate that the methodology is applicable to a far larger range of surface roughness values (see Figs. 17 and 18). This is not a satisfying conclusion, of course. If the reviewer has further thoughts on how this τ_0, which is representative for the whole wavelength band, can be estimated, we would be happy to test this. We would like to highlight that due to this specific issue, we increased the complexity of the experiments in Sect. 4.1 step by step, to test how far this one-wavelength equation is applicable in the multi-wavelength retrieval algorithm. Consequently, we added the following sentence to its introduction in Sect. 4 (line 310 in the preprint):

"The complexity of these experiment is increased incrementally to test whether the one-wavelength model of Sect. 2 is applicable to the actual multi-wavelength retrieval algorithm."

Although biases on the order of 2.5 ppm remain for the Mont Blanc region, it can be seen that the differences remain a factor of 4 smaller compared to the unweighted average. This implies that using this new method will also improve the results under such conditions. We consider that this high-altitude mountain region pushes the retrieval and the methodology to its limits (which was the intention) and that these data will most probably be filtered out during the retrieval process using real data in the end, as a result of other terrain- and surface type-related retrieval errors.

As this discussion is not yet included in the manuscript, we replaced the sentence in line 195 by the following paragraph:
"The question remains under which conditions this first-order Taylor approximation is valid. However, an evaluation using Eq. (11) and (23) will not help in the real application because the derivation is based on a one-wavelength retrieval while real retrievals use many wavelengths of different wavelength bands to retrieve the greenhouse gas concentration. We analyzed the parameter τ_0 for all wavelengths used in the retrieval for CO2M, see Sect. 3.3, and found that it covers four orders of

magnitude ranging from values between 0.008 and 25.6. This leads to errors due to this first-order approximation also differing by several orders of magnitude (not shown). As it cannot be evaluated in advance of the retrieval which will be the dominant error cause, this derivation is a demonstration that the albedo-weighted average of the surface altitude is a direct result of the assumptions made in this section, but its value for applications in actual retrievals is yet to be shown. This is why we start with examples with well-defined conditions increasing the complexity of the underlying data and finish with applications using real subpixel albedo and altitude data in the first part of the results section (Sect. 4)."

In addition, we replaced the sentences in lines 438 – 440 by the following sentence to emphasize the problem with the Mont Blanc region:
"In addition, it was estimated that larger surface roughness will lead to larger errors when applying the first-order approximation using Eq. (23). Therefore, we expect high mountain regions, such as around the Mont Blanc, the highest mountain of the European Alps, to show remaining biases even after applying the albedo-weighted average of the surface altitude. This region is chosen to assess how large this effect is when pushing the methodology and the retrieval towards the edge of their applicability."

3 The use of Sentinel-2 bands as proxies for CO2M bands is ok?

For SWIR2, the mismatch is even more significant: CO2M covers approximately 1990–2095 nm, while Sentinel-2 B12 covers roughly 2100–2280 nm. Snow, ice, vegetation, dry soil, rock, and mining surfaces may have substantial spectral differences across this interval.

Clearly state in the conclusion that the current results rely on the approximation of CO2M band reflectance using Sentinel-2 bands.

This directly affects the albedo-weighted surface altitude, especially in scenarios involving snow-covered mountains, vegetation boundaries, bare rock, and mining areas.

Response: We are aware of this mismatch, which was already described in the manuscript. The wavelength ranges are compared explicitly in Table 1 in the manuscript. In addition, we would like to highlight that in both Eq. 11 and its approximation (Eq. 23), the albedo only appears as a weight for sums. Therefore, the absolute value of the albedo is not as important as long as its variability can be scaled to the respective wavelength band. As the albedo of the S2 channels B11 and B12 (which are much farther apart than the difference to the respective CO2M wavelength band) are well correlated, this seems to be a good assumption for CO2M. However, this cannot be tested without actual surface reflectance data in the fit windows used for CO2M. We agree that this was not mentioned in the conclusions and added these sentences after the sentence in line 682, which we hope clarify and sufficiently highlight this issue:

"Furthermore, the Sentinel-2 spectral bands do not exactly match the wavelength bands measured by CO2M. However, the available bands and their data sets cannot be changed. Therefore, this adds to the uncertainty in using Copernicus Sentinel-2 data as a proxy for the real surface reflectance at the CO2M wavelength bands."

Minor/technique comments:

1. The manuscript uses "albedo" and "surface reflectance" . Since Sentinel-2 provides surface reflectance and FOCAL retrieves Lambertian-equivalent albedo polynomial coefficients, it would be useful to clarify how these quantities are related in this study.

Reponse: In short: They are the same. In the introduction, we point out that we use "albedo" in this study as an abbreviation for the surface reflectance of a specific wavelength band. Generally, albedo refers to the reflectivity of a surface, usually as a broad-band quantity in the visible spectrum. Surface reflectance is the wavelength-dependent reflectivity, so the same quantity as albedo, but not only in the visible, which is why we abbreviated it here to albedo. The FOCAL algorithm retrieves the reflectance parameters for each fit window, which we call albedo parameters for the same reasons. We replaced the corresponding sentence of line 269 by:

"FOCAL-CO2M estimates the surface reflectance for each wavelength band (NIR, SWIR1, SWIR2) separately with a polynomial to account for non-linear effects, which are abbreviated as albedo polynomial coefficients in this study."

2. In Eq. (13): the starting point for the summation should be 1 rather than i?
   Response: Thanks, we corrected it.

3. In Eq. (16):It would be better for the subscripts to be consistent with the surrounding context.
   Response: We think that this comment is about an inconsistent use of subscripts "I" and "j". As "I" has always been used within sums, but the derivative in Eq. (16) is done for a specific subpixel altitude, we decided to use another subscript "j" here. We agree that this has not been introduced at this point, which is why we replaced the sentence in line 174 by:

   "so that the derivative of $z_{s,eff}$ with respect to one subpixel $z_{s,j}$ can be written as:"

4. At some points, ΔXCO2 appears to denote a correction term; else where it denotes the deviation from the true input value of 400 ppm. Please check it
   Response: In the whole study, ΔXCO2 always refers to the deviation from the true input of 400 ppm. The correction that is suggested in the results emulates the deviations from the true input of 400 ppm, which makes it a ΔXCO2, too. We've already added the subscript "fit" where it is a fit to the actual ΔXCO2 of the retrieval, therefore we think that this should be clear enough.