Authors' Response to Referee #3

Dear Referee,

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

This manuscript investigates an important and practically relevant error source in passive satellite retrievals of greenhouse-gas column abundances. Biases in XCO_2 caused by correlated sub-pixel variability in surface reflectance and surface altitude. The authors argue that the conventional use of an unweighted mean surface altitude within a satellite ground pixel can misrepresent the effective light path when brighter sub-pixel areas are preferentially located at higher or lower elevations. To address this issue, they derive an albedo-weighted effective surface altitude and test the concept using Copernicus Sentinel-2/HLS surface reflectance, Copernicus DEM data, and the FOCAL-CO2M retrieval framework. The manuscript applies the method to idealized two-part pixels and to several real-world examples. The study also proposes an empirical post-processing correction and discusses implications for surface-roughness filtering and spatial-sample size.

I recommend minor revision before the manuscript can be considered for publication.

Response: We thank the reviewer for their detailed review and respond to each point in our responses below.

Major comments

1. The manuscript suggests that the albedo-weighted surface altitude should be used in trace-gas retrievals and that it can reduce biases below future mission requirements. This is a potentially important conclusion, but the evidence presented is not yet sufficient to support it in an operational sense.

The study uses FOCAL-CO2M to generate simulated spectra and then retrieves XCO_2 with FOCAL-CO2M using either unweighted or albedo-weighted altitude input. The authors also state that no radiance noise is added, although a realistic uncertainty based on signal-to-noise ratio is used in the inversion.

I recommend that the authors revise the abstract, conclusions, and relevant parts of Section 4 to make clear that the demonstrated improvement applies to self-consistent FOCAL-CO2M simulations under controlled assumptions. Statements such as "below the requirements for future satellite missions" should be qualified, particularly because the Mont Blanc example still shows substantial residual errors even after applying the albedo-weighted altitude. A more appropriate conclusion would be that the method substantially reduces one specific component of terrain-related XCO_2 bias and may be valuable for future retrieval development, but that its operational benefit remains to be demonstrated under realistic measurement and forward-model uncertainties.

Response: We agree that the conclusion to use this in operational products is yet to be demonstrated and made the following adaptions to the manuscript:

Abstract: We removed the statement mentioned by the reviewer. In addition, we added "in these self-consistent experiments" to the last sentence.

Conclusions: Line 697: We included "that there are strong indications" to this sentence. We include "The results using the self-consistent setup suggest that accounting […]" in front of the sentence starting on line 700 and we add the conclusion as suggested by the reviewer to the end of the conclusions: "The method substantially reduces one specific component of terrain-related XCO_2 bias and may be valuable for future retrieval development, but its operational benefit remains to be demonstrated under realistic measurement and forward-model uncertainties which is the subject for a future study."

2. In the theoretical derivation in Section 2, the authors derive an exact effective altitude expression under a simplified Lambert-Beer framework and then apply a first-order Taylor expansion around the mean surface altitude. This leads to the albedo-weighted surface altitude. The derivation is elegant and physically intuitive, but the validity of the first-order approximation is not sufficiently quantified.

The manuscript states that the expansion assumes the deviations of individual sub-pixel altitudes from the mean altitude are small enough. However, later applications include cases with large sub-pixel altitude variations, including mountainous terrain and surface-roughness bins extending to several hundred meters. In such cases, second-order and higher-order terms may not be negligible. The Mont Blanc case appears to be an example where the albedo-weighted approach still leaves residual XCO_2 errors of order ppm, suggesting that the first-order approximation or other simplified assumptions break down for strongly heterogeneous high-mountain scenes.

I suggest the authors add a dedicated analysis comparing the exact effective altitude from their non-linear expression with the first-order albedo-weighted approximation. This should be done across a representative range of altitude variability, albedo contrast, background altitude, solar/viewing geometry, and optical depth. Such an analysis would allow the authors to define a validity domain, for example in terms of surface roughness, maximum sub-pixel altitude range, albedo contrast, or expected XCO_2 residual.

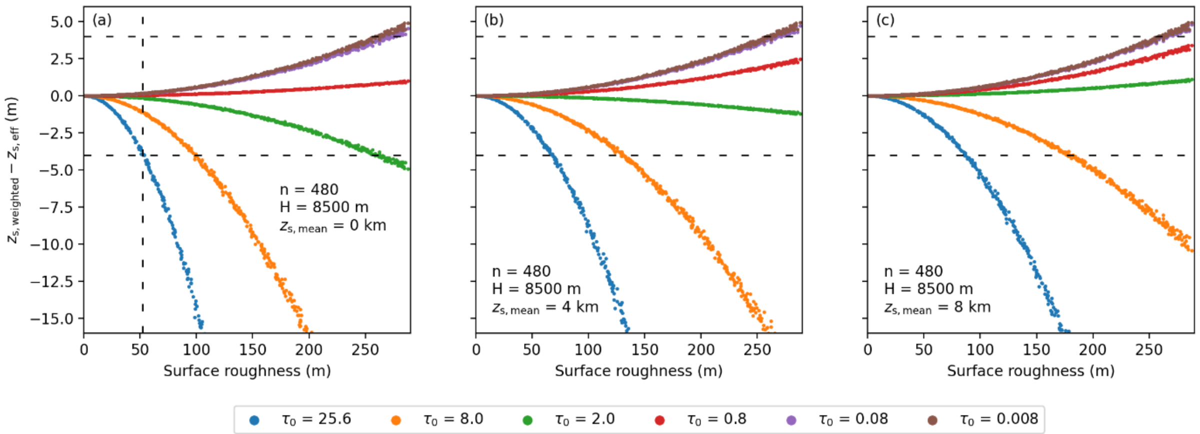
Response: [Note to the editor: This is the same response as for Major Point 2 by Referee #1] We'd like to thank the reviewer for this comment and their ideas about which kind of analysis is needed for this. It is clear that there is a limitation to the Taylor expansion of first order. In the following, 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 paper discusses scattering using a small number of prescribed scattering scenarios. This is useful, but the range appears limited. In the appendix, the authors test no scattering and two scattering cases with different optical thickness and scattering-layer pressure. The resulting correlations suggest that the empirical correction remains broadly useful, but one of the scattering cases has much lower explanatory power than the others.

I suggest the authors avoid statements implying that the proposed correction is valid over a wide range of scattering conditions unless a broader sensitivity analysis is performed. A more complete test would vary aerosol optical depth, scattering-layer height, Ångström exponent, surface albedo level, solar zenith angle, and viewing geometry. At minimum, the authors should explicitly state that the scattering sensitivity is limited to the few prescribed scenarios tested in this study.

Response: We thank the reviewer for this comment and agree that tests of three scattering configurations might not be enough for this conclusion. It is not feasible, though, to do the complete analysis as suggested by the reviewer, as it requires computation of the database of spectra (computation time 10 hours) and of the 3 million CO2M spatial samples (~1 day incl. some parallelization) for every single scattering configuration. To test the range in all parameters mentioned by the reviewer would require 10^6 different setups. We therefore follow the reviewer's last suggestion and remove the first concluding sentence in line 715 and replace the paragraph in this section by:

"To summarize this section, Equation (30) is also applicable to cases without scattering in the atmosphere, but not if the scattering optical thickness is increased to $\tau_{sca} = 0.037$. A more comprehensive study varying all scattering parameters would be needed to confirm that the equation is also applicable for a wide range of scattering regimes."

4. The manuscript presents several real-world examples with useful maps and qualitative discussion. A summary table would greatly improve readability and allow readers to compare the performance of the unweighted, albedo-weighted, and post-processed approaches.

Adding a table for each test region is recommended, which can include the number of valid spatial samples, mean bias, median bias, standard deviation, RMSE, etc. This table should include at least Bełchatów, the Black Forest, Mont Blanc, and Germany.

Such a table would also clarify where the method works very well and where it still leaves large residual errors.

Response: Following the suggestion by the reviewer, we extended Table 3 in the manuscript to show the values for all regions and the whole dataset as a summary. We note that the predicted ΔXCO_2 needed for the root-mean-square error is the constant value of 400 ppm, which means that the standard deviation and the RMSE are identical in this case. Therefore, we show only the number of spatial samples (all assumed to be valid, cf. also our explanation to point 2 of the reviewer), the mean and median bias, and the standard deviation for all regions and methods.

We replaced Table 3 in the manuscript by this new table and added the following paragraph describing the results after line 597:

"In addition, Table 3 shows statistics for all regions analyzed in this study. The median and mean biases are reduced for all regions, except for the Mont Blanc, where the mean bias increases from -0.08 ppm (unweighted) to 0.29 ppm (albedo-weighted) for the reasons discussed in the previous sections. The standard deviations are reduced for the albedo-weighted and post-processing method compared to using the unweighted average. The standard deviations using the post-processing are slightly larger than using an albedo-weighted average directly in the retrieval."

Minor comments

1. The abstract should explicitly state that the study is based on self-consistent simulations, not on real CO2M observations or independent satellite retrieval validation.
   Response: We already mention in line 11, that we "create a variety of self-consistent experiments to test this theory". In order to avoid any confusion, we added "in these self-consistent experiments" to the last sentence in the abstract.

2. Line 75: The phrase "this examples shows" should be corrected to "this example shows".
   Response: Corrected.

3. Line 81: The phrase "the first of three satellite" should be corrected to "the first of three satellites".
   Response: Corrected.

4. In Eq. (13), the summation index appears to be written as (i=a). Please check whether this should be (i=1).
   Response: Corrected to i = 1.