Dear reviewer,

Thank you for your insightful comments and constructive suggestions. We understand your concerns regarding the limited validation presented in the manuscript. The primary aim of this paper is to introduce new physics-based mitigation techniques for OCO retrievals, which differ fundamentally from existing statistics-based bias correction methods.

Advancement of this methodology and an expanded validation effort are planned for future work. In response to your feedback, we have included two additional case studies that apply the same bypass parameters and provide a comparison with the baseline method in this response document on page 5. Given the manuscript length, we will include only the first case of these two samples in the appendix.

Your suggestion to streamline the abstract by focusing on the two primary mitigation strategies is well taken. We will shorten the abstract as suggested while retaining the description of the EaR^{3}T-OCO setup within the manuscript. A detailed explanation of both the methodology and experimental setup is essential for clearly conveying the approach and demonstrating its application.

We notice your concerns about the applicability of our method in the presence of shadows. Since the OCO retrieval algorithm pre-screens most cloud-shadow footprints over land, we focus our mitigation efforts on clear-sky and non-shadow footprints. These footprints constitute most nadir observations and are therefore more critical for accurate retrievals than cloud-shadow footprints. A detailed response addressing this concern is provided in the following statements.

Below, we provide detailed responses to each point raised. The comments and suggestions are highlighted in red, responses are in blue, and references to the original manuscript content appear in black.

General comments:

The paper presents a step forward to mitigate biases of retrieved CO2 concentrations from satellite observations (OCO-2/3) due to scattering of radiation by clouds in clear regions. 3D scattering effects are approximated by a linear fit of the radiances against selected wavelengths of a given spectral window.

Thank you for this summary, which accurately captures our work.

Based on another ongoing work, the authors claim that 3D scattering effects for a specific scene can be described by the slope and the intercept of the fit. These two parameters can be used to adjust the cloud contaminated observations to corresponding clear sky spectra.

To illustrate, we reference Fig. 2 from Schmidt et al. (2019) (references at the end of this document), which shows three distinct types of footprints with the 3D cloud effect. The characteristic correlation pattern forms three branches: clear-sky pixels with cloud in the vicinity (green), footprints in cloud shadow (blue) and cloudy pixels (red). For OCO-2/3 land nadir footprints, our analysis focuses on the clear-sky pixels (green), where we apply a linear fit to derive the slope and intercept parameters.

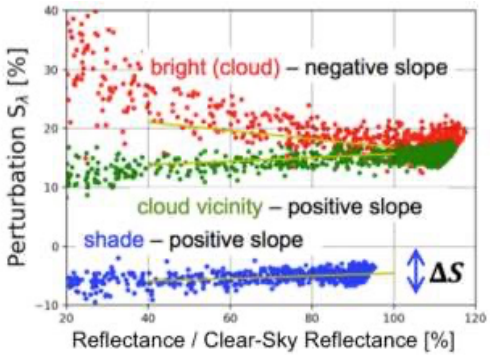


Figure R1. (Fig 2 from Schmidt et al., 2019) Reflectance enhancement (spectral perturbation) as a function of the reflectance for three distinct types of footprints.

In a following step an exponential fit of the slope and the intercept against the effective distance to clouds is performed so that the adjustment of the spectra can be done without 3D radiative transfer (3D RT) simulations for each pixel. In order to avoid 3D RT calculations completely the so-called bypass method is also presented, which uses only observational data to obtain the slope and the intercept as a function of distance from the cloud and "other scene parameters" (the second is not shown in the paper).

In our study, we utilize observational data from MODIS, including cloud and aerosol properties, to better capture the true radiance distribution. However, deriving the slope and intercept requires direct comparisons between 1D and 3D simulations. Without these simulations, it would not be possible to accurately determine the relationship between radiance and cloud distance. Based on the case shown in this study, we also simulate the radiance for different solar zenith angles, surface albedo, and aerosol optical depth to evaluate the impact of these scene parameters on 3D cloud effects.

The basic method using the 3DRT simulations requires cloud input data (cloud optical thickness, effective radius and cloud top height from MODIS observations), surface albedo (also from MODIS) and several assumptions (which are not fully described). It is rather difficult to obtain the realistic setup and therefore not straightforward to obtain the slope and intercept parameters.

The realistic setup for a radiance simulation is indeed always difficult. However, the short time difference between OCO-2 and MODIS-Aqua makes it more straightforward to build up the simulation environment. Since OCO-2 and MODIS Aqua are on the

same orbit with only a 6-minute time difference, the MODIS observations provide a highly realistic environmental setup for OCO-2, ensuring a high degree of consistency between the two datasets. Some future missions, such as CO2M, also aim to have an imager directly alongside the spectrometer to obtain more simultaneous cloud information. Therefore, we think our approach is useful.

While we obtain cloud optical thickness, effective radius, cloud top height, surface albedo, and aerosol properties from MODIS, some additional assumptions are necessary to complete the 3D radiative transfer (3D-RT) simulations. Specifically, MODIS does not provide cloud geometric thickness or the mixing layer height for aerosols, which requires us to make reasonable assumptions based on typical values and the literature. To make the assumptions clear, we will add the following sentences in the simulation setup: "This study makes the assumption of fixed cloud geometric thickness (1 km for cloud top height smaller than 4 km and cloud base at 3 km for cloud top height greater than 4 km), which could lead to some uncertainties in the RT simulations. Further investigation of the impact of cloud properties on the 3D cloud effect is needed in the future."

The bypass method, which derives the parameters only from the observed radiances would be more practicable and it does not rely on other retrieval algorithms and assumptions on input to RT simulations, therefore this is in my opinion a promising approach.

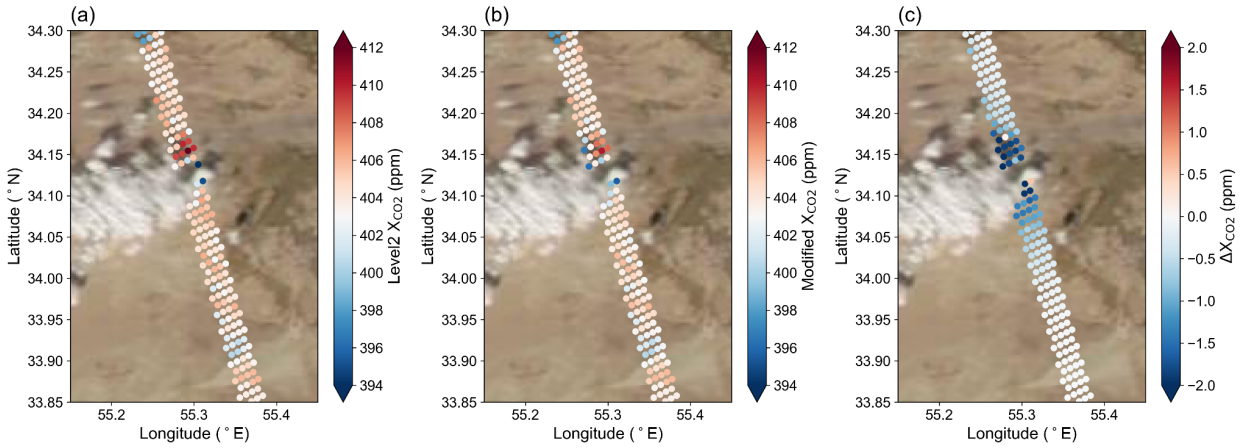
However, we want to clarify that the bypass method is a parameterized *version* of the baseline method, which derives from the result of the (more complicated) RT simulations. That is to say, the baseline method will be more accurate, but also more time-consuming compared to the bypass method. While we aim to eventually apply the bypass method *independently* to OCO-2 footprints without the need for additional RT simulations, further refinement is required. At this stage, we need to perform more RT simulations using the baseline method to build a robust, generalized bypass model that can reliably capture a wide range of scenarios.

The methods are demonstrated for a particular scene, showing the retrieved CO2 concentration of the original OCO retrieval versus the adapted retrieval. For this particular case I see that the adapted retrieval gives smaller CO2 concentrations (up to 6ppm) in the cloud shadow (for which the correction is not designed for as mentioned in the text). For in-scattering, there is almost no difference as far as I can see in Fig.9. The corrected results are similar for both methods, the baseline and the bypass approaches. The truth is not known, and the authors therefore can not prove that they achieve an improvement. A way to validate the approach would be to use synthetic observations with known input concentration.

As mentioned in lines 381-384, we refer to Massie et al. (2023), who found that relatively few cloud shadow retrievals exist in the OCO-2 Lite files due to the pre-retrieval cloud screening algorithms. Footprints affected by cloud shadows present additional

complexities beyond the typical 3D cloud bias, such as significant surface albedo variations and reduced CO_2 absorption, which complicate the analysis further. Our current study therefore focuses exclusively on mitigating 3D cloud biases in areas with cloud-induced enhanced illumination, rather than addressing the complexities of cloud shadow regions. We agree that demonstrating the method's applicability to a broader range of conditions, such as different solar zenith angles, surface types, cloud types, and viewing geometries (nadir and glint), is crucial for generalization. This will be an essential direction for future research to ensure that the method can be reliably extended to diverse scenarios.

To clarify the results in Figure 9, we have included Figure 9c, which highlights the differences after applying the mitigation strategies. We change the scale of the color bar to make the difference clearer as below:



(edited Fig. 9) Figure R2. Replicated from Fig. 9 in the manuscript, with an updated color bar scale in panel (c) adjusted to -2 to 2 ppm to better highlight ΔX_{CO2}.

We also want to state again that the bypass method is the parameterization version of the baseline method. As a result, the corrected results are similar for this case. Regarding the "unknown truth" issue, we acknowledge the limitations of using real observations without a reference. This challenge is common in OCO-2 3D cloud effect research. To address this, we followed established approaches from previous studies (Massie et al., 2021, 2022; Mauceri et al., 2023) by comparing footprints near and far from clouds, using those farther from clouds as a proxy for unbiased X_{CO2} values. The suggestion to validate using synthetic observations is certainly well taken, but not the emphasis of this particular study, which focused on real-world data instead of synthetic data. Besides, such a study has previously been conducted by Emde et al. (2022). Of course, when working with satellite data, there is no ground truth data. However, the anomalies in Fig. R2a suggest that the X_{CO2} in the vicinity of clouds is too high. This is also suggested by a companion publication (Schmidt et al., 2019, Figure 3, included below), which shows that positive slopes of the kind shown in Figure R1 (green) in the

vicinity of clouds lead to positive X_{CO2} bias. Furthermore, we show that our method removes at least part of this positive bias (negative values near clouds in Figure R2c.

It is, technically, possible to directly compare OCO-2 retrievals with data from the ground stations at TCCON. However, this would require the use of target mode, which has its own limitations (here, we focused on nadir-only observations). In the future, we plan to look at target observations as well.

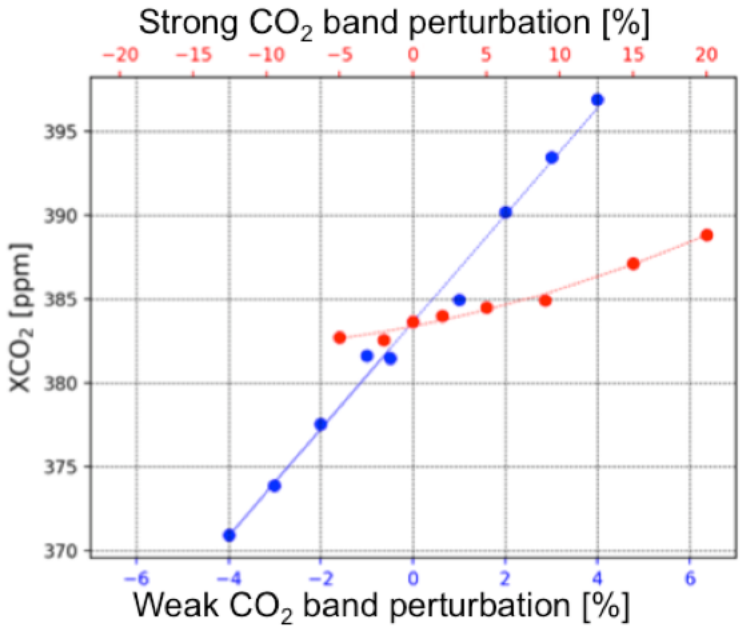


Figure R3. (Adapted from Fig. 3 from Schmidt et al., 2019) Dependence of the X_{CO2} retrieval when applying perturbations to the strong or weak CO_2 band spectra. Positive perturbations in either band introduce a positive retrieval bias in X_{CO2}, while negative perturbations result in an underestimation of X_{CO2}.

Certainly it is important to correct cloud effects in trace gas retrievals, therefore I think that the topic of the paper is appropriate for AMT. However, more validation is needed and I suggest to revise the paper in this respect. Further, several clarifications regarding the setup of the methods are required, for example it is not explained how cloud shadows are excluded.

We agree that additional validation is necessary before our method can be applied operationally. As we explained previously, the primary objective of this paper is to introduce the methodology we developed and to demonstrate its potential effectiveness. We plan to conduct more extensive validation studies in the future, incorporating a broader range of cases, including both land and ocean observations. Please also see our response to the previous point: Aside from TCOON (not applicable here), satellite retrievals of X_{CO2} (and many other geophysical products) only rarely come with ground truth. But looking at Figure R2, it is obvious that the near-cloud bias

is at least significantly reduced by our method. As mentioned above, we plan to extend our study to target mode observations in the future, at which point we can use TCOON for a direct validation of our method.

To address your comment on validation to some degree, we have included two additional cases from the same month and general geographic area as the case in Fig. 2. These cases apply the bypass mitigation method based on parameters outlined in Table R1 (Table 2 in the manuscript) and are compared to the baseline method as validation examples (shown below). 1D-RT and 3D-RT simulations were conducted for these two cases to derive the correct slope and intercept parameters. Thus, we can evaluate the bypass mitigation based on Table R1, with the comparisons illustrated in Figs. R4 and R7.

The results from case 1 indicate that the bypass method yields a mitigation trend similar to that of the baseline method, although with lower magnitudes (i.e., not as accurate as the baseline method). This could be due to differences in surface altitude and albedo, solar geometry, AOD, and other environmental factors. This case demonstrates promising results, yet adjustments to the bypass parameters with more scene variables are necessary for effective operational use. For now, we therefore recommend using the baseline method.

In contrast, case 2 shows less favorable performance of the bypass method compared to the baseline method. In case 2, the baseline method reveals a weaker correlation between ΔX_{CO2} and effective cloud distance, likely due to confounding factors, such as surface albedo effects. This indicates that the bypass method may be less effective in mixed or complex cloud conditions. Given the length constraints of the current manuscript, we have decided to add only case 1 as an example in the appendix.

Regarding your concern about footprints under cloud shadows, we do not exclude footprints that fall under cloud shadows. Instead, we apply the radiance adjustment to all footprints that pass the quality test. In other words, if the retrieval algorithm's pre-screening does not exclude a footprint—even if it is located in a cloud shadow—our radiance adjustment will still be applied, despite the method being primarily designed for bright areas.

Table R1. The same table as Table 2 in the manuscript. Amplitude and e-folding distances for ***s*** and ***i*** fittings of the simulation with a homogeneous aerosol layer in the O_2-A, WCO_2, and SCO_2 bands for 1.0 km geometric cloud thickness of low clouds.

	Slope			Intercept		
	s_{O_2-A}	s_{WCO_2}	s_{SCO_2}	i_{O_2-A}	i_{WCO_2}	i_{SCO_2}
a_s or a_i	0.457 ± 0.094	0.123 ± 0.037	0.250 ± 0.041	0.755 ± 0.327	0.648 ± 0.227	0.847 ± 0.406

d_s or d_i (km)	3.82 ± 0.44	5.04 ± 0.89	4.58 ± 0.78	2.69 ± 0.32	2.91 ± 0.31	2.35 ± 0.33

Additional case 1:

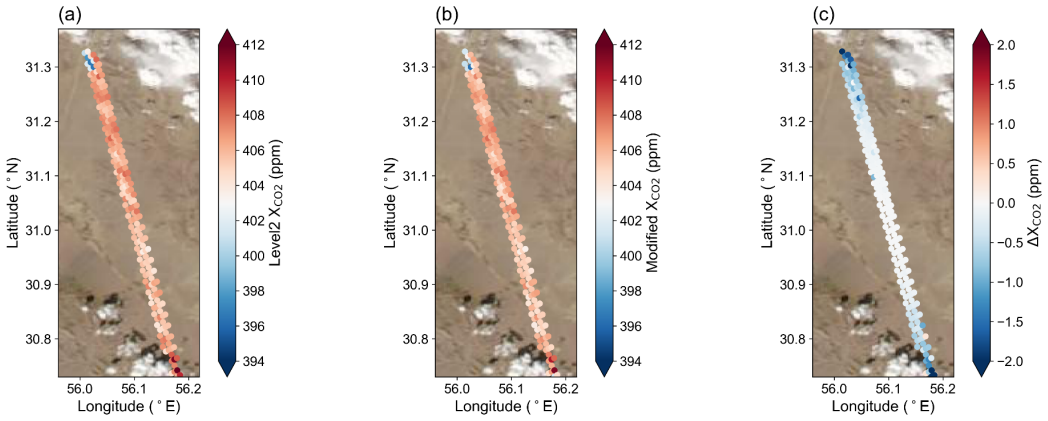


Figure R3. Satellite true-color imagery of MODIS Aqua from NASA Worldview on 5 October 2019 with (a) X_{CO2} in OCO-2 level 2 data, (b) mitigated X_{CO2} retrieved from the adjusted spectra and (c) difference between the mitigated and original X_{CO2} values.

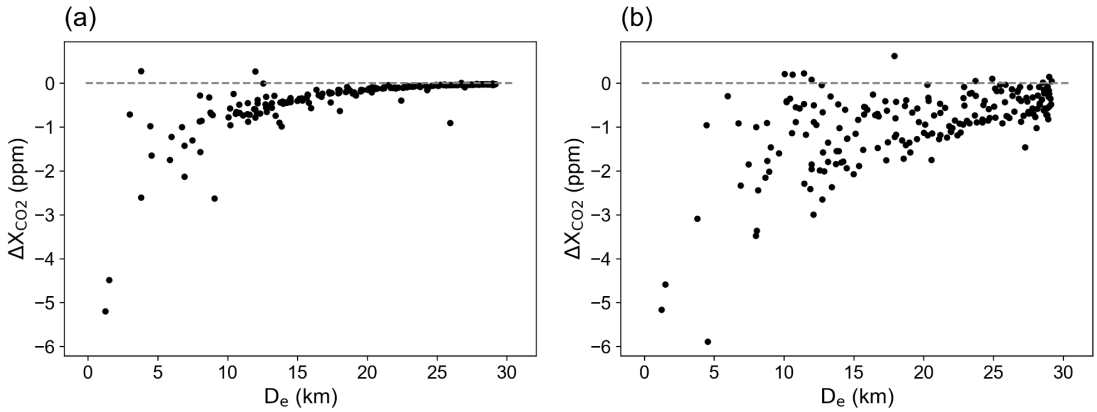


Figure R4. (a) Relationship of ΔX_{CO2} with D_e based on parameterized slopes and intercepts from the bypass method in Table 2. (b) Corresponding relationship using slopes and intercepts derived from the baseline approach for Fig. R3.

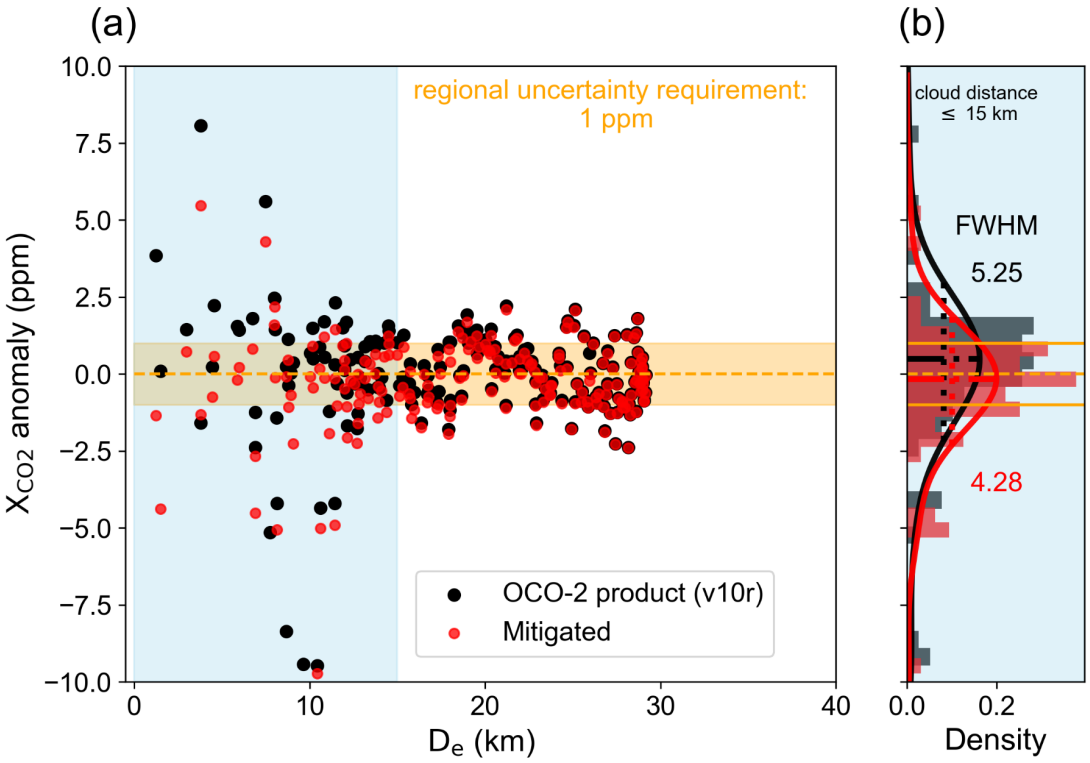


Figure R5. (a) Scatter plot comparing the X_{CO2} anomaly of the OCO-2 L2 product (in black) to its value post-spectra adjustment (in red) for the case shown in the figure above, plotted against D_e. The X_{CO2} anomaly is defined as retrieved X_{CO2} – true X_{CO2}, with the true X_{CO2} defined by the average X_{CO2} of footprints with a D_e greater than 15 km (405.96 ppm in this case). The orange shade indicates the 1 ppm mission requirement. (b) Histograms and probability density functions (PDFs) for the X_{CO2} anomaly of the OCO-2 L2 product (in black) and post spectra adjustment (in red) within a 15 km D_e. This corresponds to the blue-shaded region in (a). The FWHM values of the PDFs of v10r and adjusted data points are 5.25 and 4.28, and the PDF averages are 0.93 and 0.18, respectively. The average change in X_{CO2} after the spectra adjustment for De less than 15 km is -0.86 ppm.

Additional case 2:

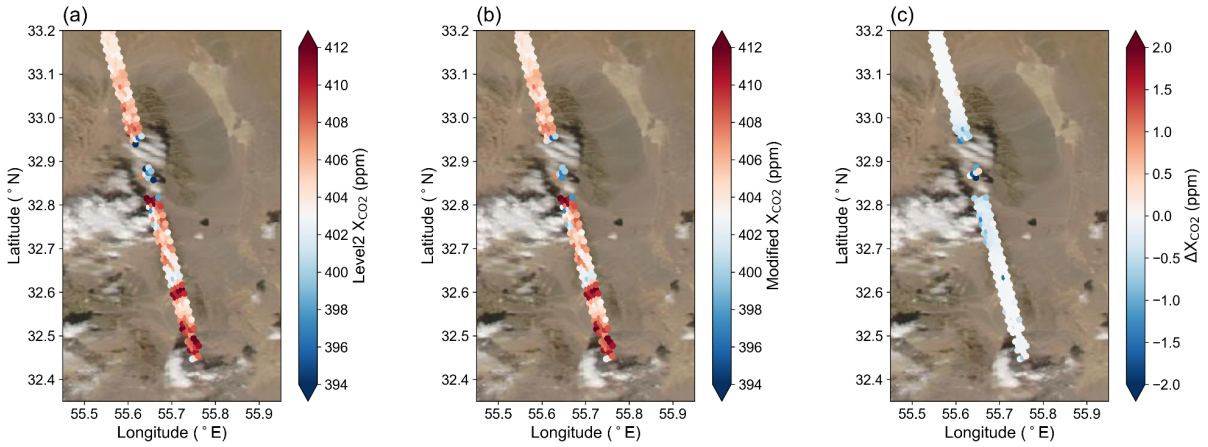


Figure R6. Satellite true-color imagery of MODIS Aqua from NASA Worldview on 18 October 2018 with (a) X_{CO2} in OCO-2 level 2 data, (b) mitigated X_{CO2} retrieved from the adjusted spectra and (c) difference between the mitigated and original X_{CO2} values.

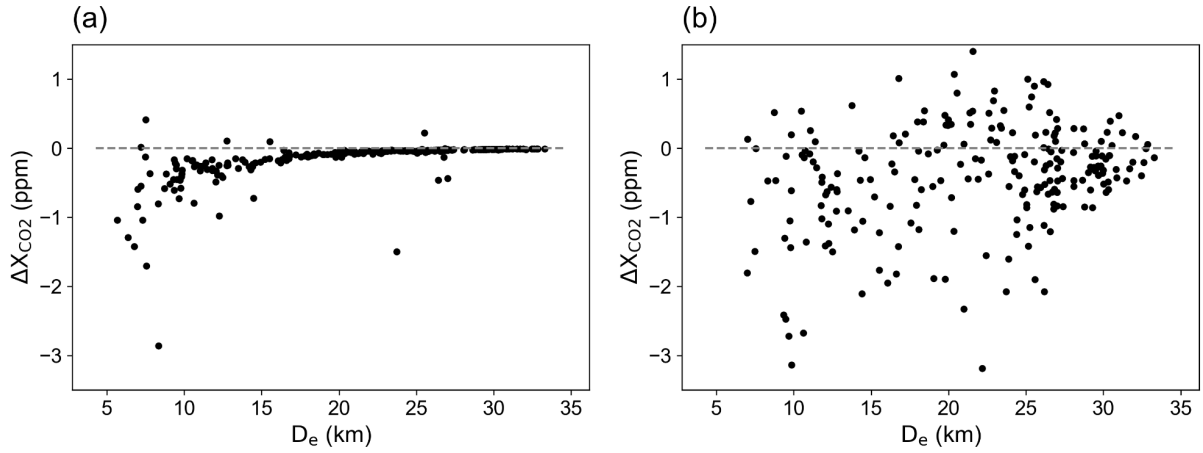


Figure R7. (a) Relationship of ΔX_{CO2} with D_e based on parameterized slopes and intercepts from the bypass method in Table 2. (b) Corresponding relationship using slopes and intercepts derived from the baseline approach for Fig. R6.

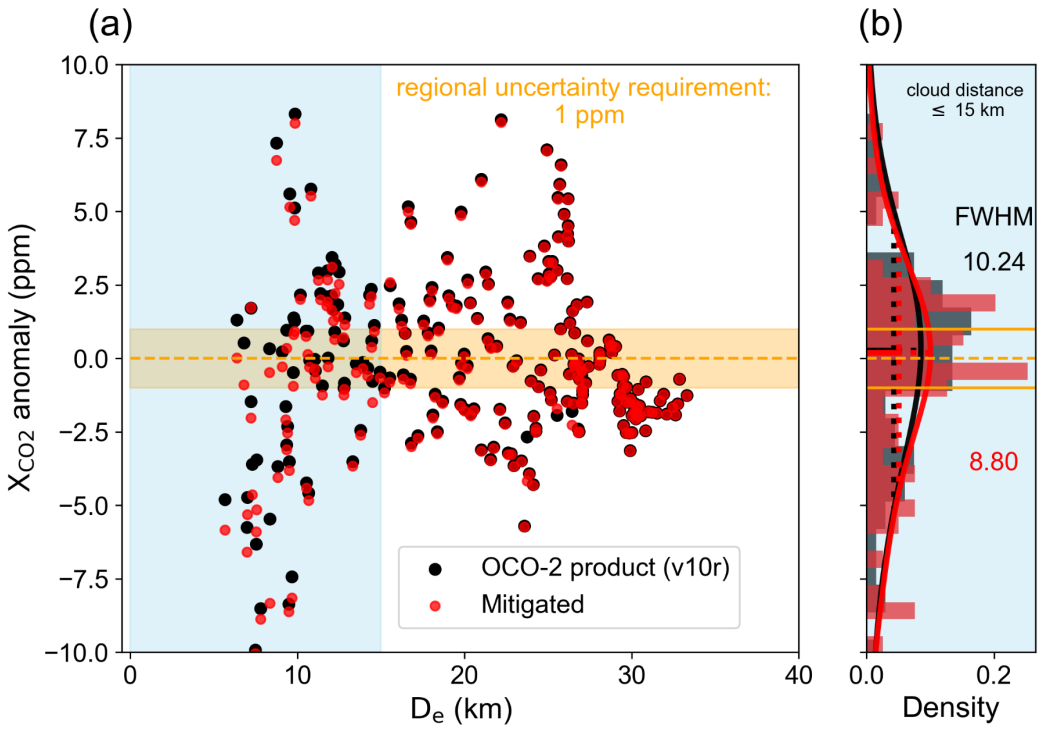


Figure R8. (a) Scatter plot comparing the X_{CO2} anomaly of the OCO-2 L2 product (in black) to its value post-spectra adjustment (in red) for the case shown in the figure above, plotted against D_e. The X_{CO2} anomaly is defined as retrieved X_{CO2} – true X_{CO2}, with the true X_{CO2} defined by the average X_{CO2} of footprints with a D_e greater than 15 km (405.69 ppm in this case). The orange shade indicates the 1 ppm mission requirement. (b) Histograms and probability density functions (PDFs) for the X_{CO2} anomaly of the OCO-2 L2 product (in black) and post spectra adjustment (in red) within a 15 km D_e. This corresponds to the blue-shaded region in (a). The FWHM values of the PDFs of v10r and adjusted data points are 10.24 and 8.80, and the PDF averages are 0.27 and 0.20, respectively. The average change in X_{CO2} after the spectra adjustment for De less than 15 km is -0.45 ppm.

Specific comments:

- The abstract is relatively long, it could be focused a more on the two mitigation strategies and technical details about EaR3T-OCO could be shortened

Thank you for the suggestions. We will revise the abstract to make it more concise by focusing on the two mitigation strategies and moving the technical details about EaR^{3}T-OCO to the methodology and appendix, as recommended.

- Revised abstract:

"Accurate and continuous measurements of atmospheric carbon dioxide (CO_2) are essential for climate change research and monitoring of emission reduction efforts. NASA's Orbiting Carbon Observatory (OCO-2/3) satellites have been deployed to measure the column-averaged CO_2 dry air mixing ratio (X_{CO2}) with a designed uncertainty of less than one ppm for the regional average. Although cloudy measurements are screened out, nearby clouds can still cause retrieval biases due to limitations in the forward one-dimensional (1D) radiative transfer (RT) model used in the OCO retrieval algorithm, which does not account for the scattering from clouds near the satellites' footprints. These biases, known as three-dimensional (3D) effects, can be quantified using 3D-RT models, but they are computationally expensive, especially for hyperspectral applications like OCO-2/3. This paper employs a linear approximation for each OCO-2 spectral band to represent the 3D-RT perturbations on OCO-2 spectra and reduce computational demands. We apply these metrics calculated by 3D-RT to spectrally adjust the real measured OCO-2 radiance prior to the operational retrieval to undo cloud vicinity effects without modifying the standard OCO-2 retrieval code. Additionally, a parameterization method is developed to bypass the need for 3D-RT simulations by incorporating effective cloud distance and other scene variables. The spectral adjustment mitigates X_{CO2} retrieval biases in proximity to clouds over land for two cases shown in the study – the first physics-based radiance level correction of 3D-RT effects on OCO-2/3 retrievals. While the proposed method is computationally efficient for operational use, further validation is required for diverse surface and atmospheric conditions."

- l.24 "These biases, referred to as the three-dimensional (3D) effects, can be quantified effectively using 3D-RT calculations, but these are computationally expensive, especially for hyperspectral applications (e.g., OCO-2/3)."

The authors refer later to the ALIS method for spectral Monte Carlo simulations. It is true that 3D RT is generally expensive but with ALIS spectral simulations are almost as fast as monochromatic simulations (see ALIS, Emde et al., 2011)

We agree that techniques like ALIS are valuable advancements for accelerating 3D-RT calculations, making spectral simulations more efficient. Iwabuchi also developed a spectral acceleration technique. In this work, we chose our own, less sophisticated, acceleration method, but the main emphasis of the work was not on acceleration, and more on application to real-world data. In the future, we envision to leave it to the user to use either ALIS, or the Iwabuchi method, or ours, to correct real-world spaceborne spectroscopy data. In addition, another focus of this work is to demonstrate that the bypass method has the potential to operationally mitigate 3D biases *without* the need for running 3D-RT simulations at all, provided sufficient cloud information is available (e.g., effective cloud distance).

- l.44 "For the case we analyzed, both the 3D-RT calculation method and the parametric bypass method successfully corrected XCO2 biases, which exceeded 2 ppm at the footprint level, and reached up to 0.7 ppm in the regional average."

How do you know the the correction is successful, you get a difference but you do not know the true CO2 concentration?

An approach to validate the retrieval is to use synthetic data, as shown in the cited work by Kylling et al. 2022 for TROPOMI. A more systematic study on the mitigation of cloud scattering for TROPOMI is shown in Hu et al. 2022, who use 2D RT simulations to derive fits to correct retrieved airmass factors and validate this approach using realistic synthetic data based on Large Eddy simulations. The synthetic data (Emde et al. 2022) they are using includes also O_2A band spectra and could possibly also be used to validate the mitigation approach for CO_2 retrievals.

> Determining true CO_2 concentrations without simultaneous in-situ measurements, such as airborne observations, poses a significant challenge. Observations from TCCON stations offer an alternative for validating the mitigation method, though precise nadir views at these stations are very uncommon because of the relative locations of the TCCON stations relative to the sunsynchronous satellite orbits. This "unknown truth" issue is a common limitation in OCO-2 3D cloud effect studies due to the scarcity of ground-based and airborne observations available for validation. Consequently, many studies compare footprints near and far from clouds, using those farther from clouds as reference values (Massie et al., 2021, 2022; Mauceri et al., 2023). We use a similar approach, considering footprints with larger effective cloud distances as being relatively free of cloud bias.
>
> We agree that synthetic data is a valuable tool for validating satellite retrieval algorithms. However, it is nearly impossible to perfectly replicate all atmospheric interactions exactly as they occur in the real atmosphere. Therefore, testing the mitigation approach on real observations remains essential for capturing the complexities of actual atmospheric conditions. In addition, Figure R2 (and others) does show that our bias correction goes into the right direction. Quantifying this with ground truth will need to be left for further studies that include target observations and TCCON stations (as discussed in response to an earlier point). Nonetheless, we will keep using synthetic data as we have in previous work, as suggested. Again, the exclusive use of synthetic data only cannot replace the application to real-world data, where unforeseen effects such as sensor performance, unknown aerosol layers, geolocation inaccuracies etc. may play a role that cannot be captured by synthetic observations only. Our own synthetic data (e.g., from Schmidt et al., 2019) convinced us that we had to go to real-world data next, and that is the primary focus of *this* manuscript.

- l. 46 "We find that the biases depend most strongly on the cloud field morphology and surface reflectance, but also on secondary factors such as aerosol layers and sun-sensor geometry."

I assume with cloud field morphology, you mean the weighted distance to the clouds. The paper does not show how the bias depends on surface reflectance and on sun-sensor geometry. The impact of cloud scattering will certainly increase with increasing solar zenith angle but also for slant viewing angles (here only nadir view is shown). Also the cloud geometrical thickness is probably important.

By "cloud field morphology," we refer to the distribution of clouds, including cloud types, cloud top height (CTH), cloud base height, and cloud optical thickness (COT), all of which can influence the magnitude of the 3D cloud effect.

We agree that solar and viewing geometry also play a significant role in determining the extent of these biases. We acknowledge that the impact of other geometrical factors, such as solar zenith angle, is not explicitly shown in the current paper. We will add a section before Section 5.5 (3D effect mitigation) discussing the impact of surface albedo and solar zenith angle (SZA). Additionally, we plan to conduct a more extensive investigation in future work to encompass these and other relevant factors.

To illustrate the effect of solar zenith angle and surface albedo, we provide parameterization figures for the O_2-A band across different conditions:

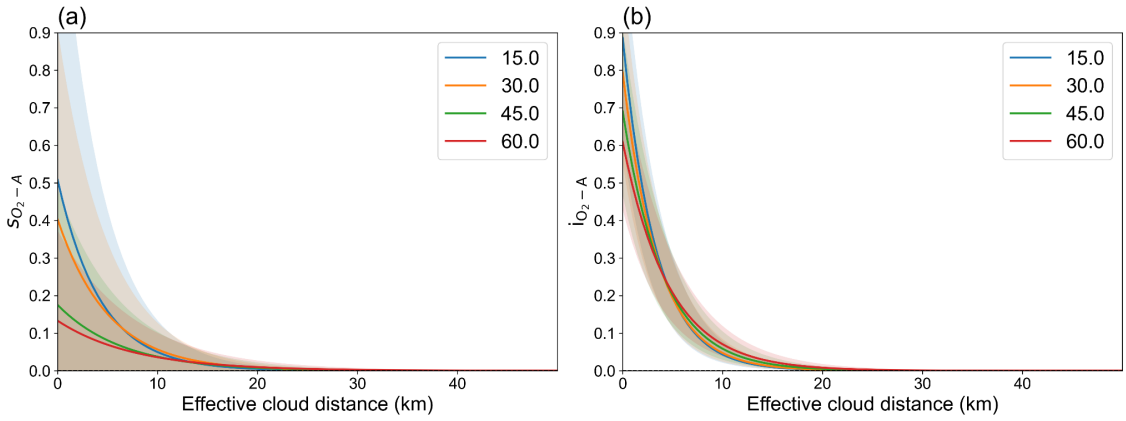


Figure R9. Parameterization of (a) slope and (b) intercept for O_2-A band with effective cloud distance, varied by solar zenith angle, while holding surface albedo and aerosol optical depth (AOD) constant.

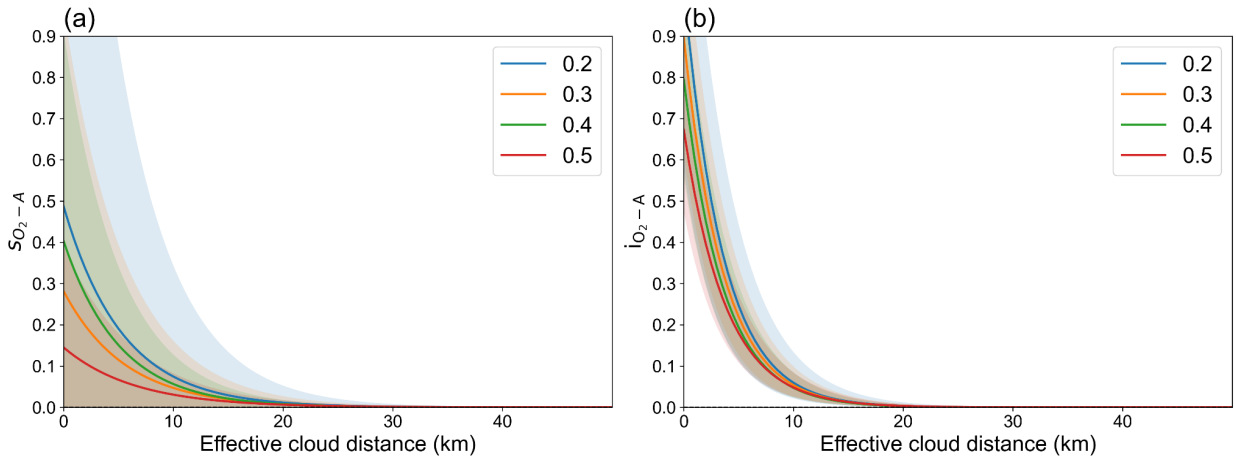


Figure R10. Parameterization of (a) slope and (b) intercept for O_2-A band with effective cloud distance, varied by surface albedo, while holding solar zenith angle and AOD constant.

Combining these results across different solar zenith angles and surface albedo values allows us to develop a two-variable linear parameterization of a_s and d_s (slope parameters) and a_i and d_i (intercept parameters). As summarized in Table R1 below, we find that the amplitude of slope and intercept is inversely proportional to surface albedo and directly proportional to the cosine of SZA (denoted as μ). Additionally, the e-folding distances of the slope are negatively proportional to both surface albedo and μ, while those of the intercept are positively proportional to surface albedo and negatively proportional to μ.

Table R1. The parameterization of a_s and d_s of slope and a_i and d_i of intercept for the three OCO-2 bands.

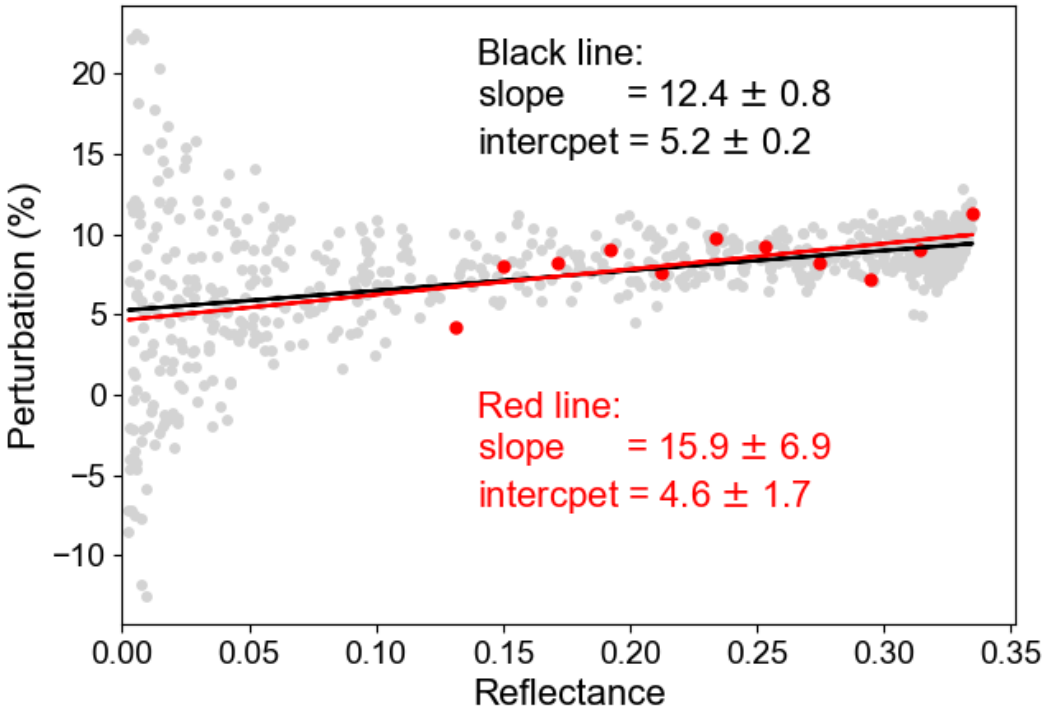
	slope	intercept
O_2-A	$a_s = -0.34 \times alb_{O2\text{-}A} + 0.57 \times \mu - 0.03$ $d_s = -3.2 \times alb_{O2\text{-}A} - 9.9 \times \mu + 14.9$	$a_i = -0.60 \times alb_{O2\text{-}A} + 0.36 \times \mu + 0.72$ $d_i = 0.42 \times alb_{O2\text{-}A} - 2.1 \times \mu + 5.2$
WCO_2	$a_s = -0.15 \times alb_{WCO2} + 0.11 \times \mu - 0.05$ $d_s = -30.7 \times alb_{WCO2} - 7.0 \times \mu + 27.5$	$a_i = -2.07 \times alb_{WCO2} + 1.65 \times \mu + 1.17$ $d_i = 0.63 \times alb_{WCO2} - 1.6 \times \mu + 3.7$
SCO_2	$a_s = -0.18 \times alb_{SCO2} + 0.29 \times \mu - 0.03$ $d_s = -22.6 \times alb_{SCO2} - 21.2 \times \mu + 34.9$	$a_i = -2.77 \times alb_{SCO2} + 2.22 \times \mu + 1.14$ $d_i = 0.51 \times alb_{SCO2} - 1.73 \times \mu + 3.35$

- Eq.1: Remove 100%, because 100%=1. Or multiply by 100 and say that the unit of the perturbation is in per cent.

We will remove 100% in Eq. 1 to simplify the expression, as the perturbation will then be represented in unit form.

- Fig.1: Is the fitted line obtained by fitting the blue dots or the grey dots? You should show that both fits result in the same slope and intercept (you could include the two fitting lines and the corresponding equations).

Thank you for the suggestion. We modify Fig. 1 and have both fitting lines as shown below:



(edited version of Fig. 1) Figure R11. Example of the linear relationship between perturbation and reflectance. The grey dots represent the complete wavelength range, while the red dots indicate the subset selected for the O_2-A band simulation. The black and red lines represent the linear fit of the grey and red dots, respectively,

- l. 146: "The intercept is related to the often-reported increase of reflectance near clouds, or decrease in shadows, whereas the slope accounts for spectroscopic effects."

This interpretation is correct as long as the spectral dependence of scattering can be neglected which is true for small spectral windows.

We will add the description in the sentence that this is true when the spectral window is small, which can be applied to OCO spectral windows as below:

- Revised text, with the main changes underlined:

"The intercept is related to the often-reported increase of reflectance near clouds, or decrease in shadows, whereas the slope accounts for spectroscopic effects for a small spectral range, where the scattering effect can be considered spectral independent."

- Fig.2: It looks as if the main differences between the retrievals are in the cloud shadow region. Could you also show the image without the CO2 concentration included to see whether there is a cloud shadow at the place with higher CO2 concentrations?

Here are the MODIS-Aqua images (a) without and (b) with OCO-2 footprints overlapped. Note that there is a six-minute time difference between the two satellites. Again, we want to emphasize that footprints over cloud shadows are out of the scope of this research.

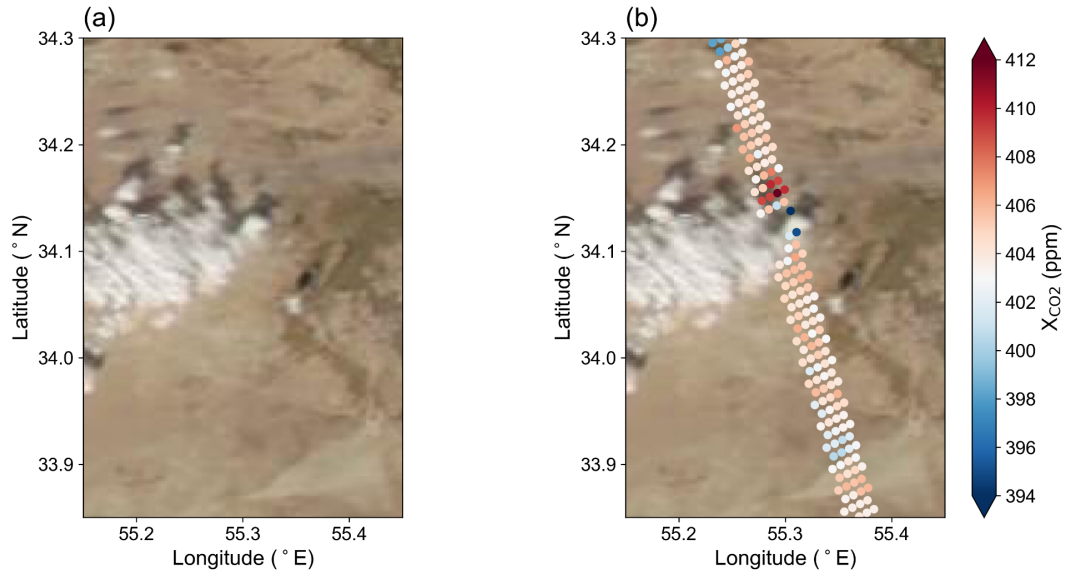


Figure R12. Satellite true-color imagery of MODIS Aqua from NASA Worldview on 18 October 2018, (a) without and (b) with OCO-2 retrieved X_{CO2} overlaid.

- l. 267: "To determine the cloud optical thickness (COT) of each pixel, we run the RT model over several COT and derive the COT-radiance relationship by ourselves to ensure the radiance consistency in 1D-RT simulation."

This is not so clear. It means you do not use the cloud optical thickness as from the MODIS retrieval algorithm but an adjusted optical thickness that is needed as input for the 1D RT simulation to be consistent with the observed radiance? Isn't this exactly the same as the MODIS retrieval?

The retrieved optical thickness is of course biased by 3D effects because generally the reflectance is smaller in 3D simulations compared to 1D for the same vertical optical thickness

due to photon leakage on the cloud sides. That means that the COT is underestimated, but this is not what you mean here?

> Converting radiance to COT follows the method detailed in Appendix C2 of Chen et al. (2023). We do not use the MODIS COT directly because the MODIS cloud identification algorithm may overlook small, isolated clouds, leading to significant radiance discrepancies for nearby OCO footprints. Instead, we apply an optimized cloud detection radiance threshold tailored for each scene and then perform the radiance-to-COT mapping accordingly.
>
> Regarding the concern about the underestimation of COT, we acknowledge that using 1D-RT-based radiance-to-COT mapping can result in biases, where COT may indeed be overestimated for low radiances and underestimated for high radiances due to the 3D effects. Since our primary area of interest is in bright regions, we focus on minimizing radiance discrepancies over these areas rather than on correcting the biases at cloudy pixels. In essence, the clouds in our calculations serve as source of additional diffuse illumination, which is not as strongly dependent on COT than, say, the reflectance. Still, we acknowledge that we may be under- or overestimating the amount of diffuse illumination. Therefore, in the future, integrating COT correction techniques could further refine COT mapping. For instance, Nataraja et al. (2022) developed a neural network-based COT correction using 650 nm reflectance, which could be a potential approach to incorporate.

- l.281: "MCARaTS iteratively traces the path of each photon and calculates the distribution of photons based on the final probability."

What do you mean with "distribution of photons" and "final probability"?

> MCARaTS traces each of the photons (10^9 photons for this study) and records their paths within the simulation domain to determine where they are absorbed, scattered, or transmitted. This process results in a distribution of photon interactions across the simulation area. By dividing the number of photons reaching a particular region by the total number of photons simulated, we obtain a probability distribution, which is then used to calculate the resulting radiance.

- l. 287: "The mean radiance and the standard deviation are then calculated from three runs to estimate the uncertainty."

Three samples are not sufficient to estimate the standard deviation. Why not running more simulations with less photons to get a better estimate?

> We agree that using only three runs may not provide a robust estimate of the standard deviation. However, due to the high computational cost of each MCARaTS simulation, we opted for three runs with a larger number of photons (10^9 per run) to ensure stable

mean radiance values. Running more simulations with fewer photons could increase the noise, affecting the reliability of the radiance estimates. In future work, we will explore optimizing the balance between the number of photons per run and the number of runs to achieve a more precise estimate of the standard deviation while maintaining computational efficiency.

- Eq. 4: Remove 100% which is equal to 1.

We will modify it as suggested.

- l. 328: "Fig. 3a-b presents the 3D-RT simulation and MODIS observation of 650 nm using the COT, CER, and CTH shown in Fig. A3."

It is not fully clear how the cloud input is created from the MODIS data.

Why is CER seems constant (looks like this in Fig. A3)? Please provide the value of CER. How is the cloud vertically constructed? What is the cloud base height? What is the sun-observer geometry?

We apologize for not making this clearer in the manuscript. For this study, we chose to keep the cloud effective radius (CER) constant for simplicity, which is why it appears uniform in Fig. A3c. Specifically, we assigned CER values of 10 μm for low clouds and 25 μm for high clouds. In future work, we intend to incorporate MODIS-derived CER values to better capture spatial variability.

Regarding cloud structure and geometry, clouds are vertically constructed with cloud top height (CTH) derived from MODIS data, as shown in Fig. A3a. The cloud base height (CBH) is set to 1 km for low clouds and 3 km for higher clouds. The sun-observer geometry is matched to OCO-2 observation conditions, using an average viewing angle of 0.31˚ and an SZA of 48.5˚ across all OCO-2 footprints in the area of interest.

In response to your other questions, we have clarified the description in line 265 as follows:

- Original text (Line 265):
  "Once the cloudy pixels are identified, we retrieve the cloud top height (CTH) and cloud effective radius (CER) of the nearest location from the MODIS MYD02 cloud file and assign them to each cloudy grid point."
- Revised text:
  "Once the cloudy pixels are identified, we retrieve the cloud top height (CTH) of the nearest location from the MODIS MYD02 cloud file and assign it to each cloudy grid point. The cloud effective radius (CER) is manually set to 10 μm for low clouds and

25 μm for high clouds in this study. In future versions, we plan to use the actual MODIS CER values to capture more realistic variations."

Can the method also be applied for ice clouds? Is it valid over ocean?

We have not yet tested the method on ice clouds, so its applicability in such cases remains uncertain. Over ocean surfaces, OCO-2 is operated in glint mode, not in nadir mode. We only studied nadir mode in this manuscript. In glint, the CO_2 bias behaves differently due to factors such as ocean glint reflection and specific scattering processes. Our preliminary result (not in this manuscript) shows that the bypass method may also work for ocean cases. However, additional investigations are needed to adapt and validate the method for oceanic conditions.

How is the spectral albedo generated from MODIS data? A dataset and a method to generate hyperspectral surface albedo data from MODIS data is presented in Roccetti et al. 2024, could this be included in your model?

For the surface albedo in this study, we used both the *"brdf_reflectance"* from the OCO-2 Level 2 data and the MODIS MCD43A3 data. The spatial distribution of surface albedo over the area of interest is derived from the MCD43A3 product and then scaled using the OCO-2 BRDF reflectance values. Currently, we assume a wavelength-independent surface albedo within each band range. Incorporating hyperspectral surface albedo data, such as that presented in Roccetti et al. (2024), is an excellent suggestion. Unfortunately the publication was not available when we developed our method. However, we expect that it will enhance the accuracy of our EaR^{3}T-OCO simulator, particularly for surface types with significant spectral variability over small wavelength ranges. We plan on implementing this improvement in future versions of the model. Thank you for the excellent suggestion!

- l. 330: "The heat map in Fig. 3c shows a good agreement between the simulation and observation. As a result, we are confident that the simulation at other wavelengths is able to approach the actual condition."

More tests needed to draw this conclusion. How do other bands compare? At least one image in NIR should be shown.

To address your request, we have included the simulations for all three channels overlaid with OCO footprint observations, as shown below:

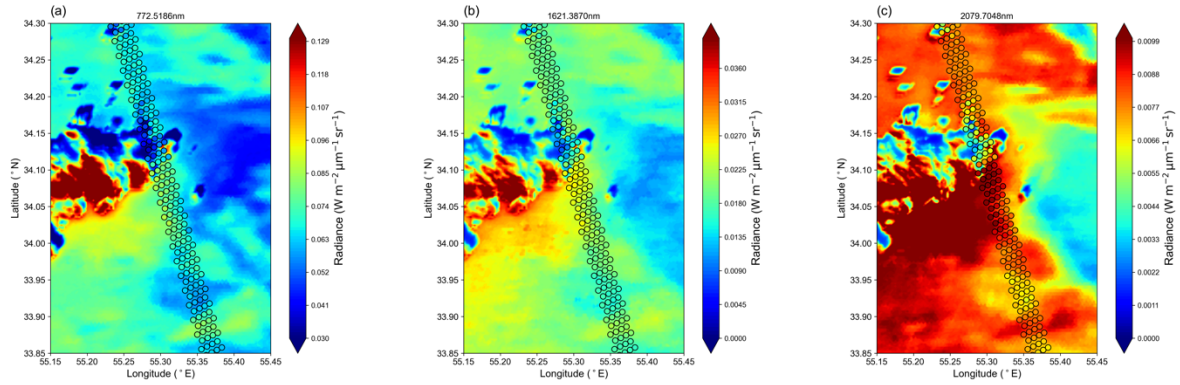


Figure R13. Simulated continuum radiance overlaid with OCO-2 observed radiance (c) for the (a) O_2-A, (b) WCO_2, and (c) SCO_2 band.

Most of the points in the image correspond to clear sky. How is the correlation for the cloudy pixels only? This would show better whether the cloud input is realistic.

We recognize that our radiance-to-COT mapping, which is based on 1D-RT, can lead to an overestimation of COT for low radiance values and an underestimation for high radiance values due to 3D effects. For example, comparing the cloudy pixels between our simulation and observation presents an underestimated radiance. Since our primary focus is on bright areas, where radiance values are typically higher, we prioritize minimizing radiance discrepancies in these regions rather than at the cloudy pixels. As mentioned above, integrating COT correction techniques could further refine COT mapping in the future.

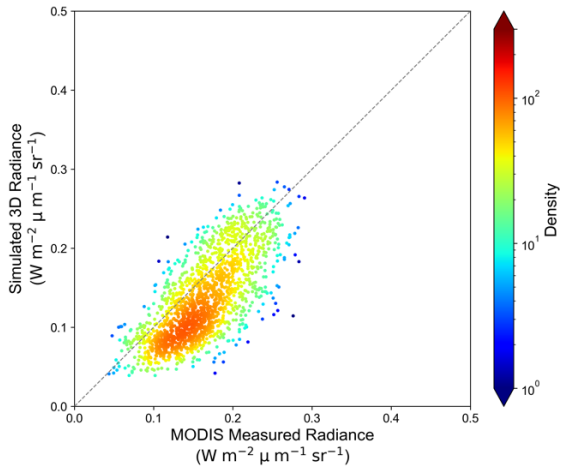
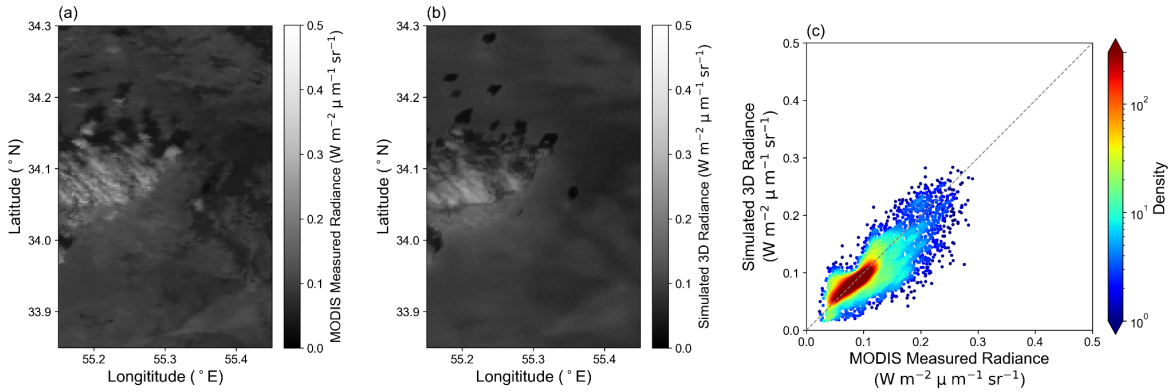


Figure R14. A scatter comparison between Fig. 3a and 3b (Fig. R15 below) for pixels with COT greater than 0.

I think "scatter plot" is a better name than "heat map". Could you also include a colorbar?

Thank you for your suggestion. We will update the Fig.3 as below with changing the heatmap to scatterplot and adding color bar:



(edited Fig. 3) Figure R15. MODIS observation at 650 nm (a) and 3D radiance simulation at 650 nm by EaR^{3}T (b). A scatter comparison between (a) and (b) is depicted in (c).

- l. 345: "Employing a reduced number of wavelengths, uniformly distributed across the reflectance space, effectively minimizes computational demands while still permitting the derivation of $\boldsymbol{s}$ and $\boldsymbol{i}$ for the linear relationships within each band."

How are the wavelengths selected? Do you use the same set of wavelengths for all scenes?

We describe the wavelength selection approach in Lines 340-343. For each band, we first calculate the average transmittance of the region of interest using representative values for solar and viewing geometries, surface albedo, and surface height. To avoid running simulations at wavelengths with extremely low transmittance, we define a minimum threshold for each band, set as the lower of (1) 40% of the band's maximum transmittance or (2) the band's minimum transmittance. As a result, the selected wavelengths for different scenes could be different.

Once the applicable wavelengths are identified, we sort them by transmittance and select a fixed number (e.g., 11) that are uniformly distributed within the transmittance space, as shown in Fig. 4a. This selection ensures that we capture the key spectral variations while minimizing computational cost. The same method is applied for each scene, but the exact set of wavelengths may vary slightly depending on the scene-specific atmospheric conditions.

- Original text (Lines 340-343):
  "To balance computational demands with accuracy, we selected 11 wavelengths evenly distributed over the high 60% transmittance based on sorted clear-sky transmittance for further RT simulation (depicted in Fig. 4 as an example, the transmittance of full spectra is presented in Fig. A4). The transmittance is calculated based on the atmosphere profile, θ_s, etc."

Have you tested whether this relationship remains linear in different cloud situations? Clouds can shield the lower atmosphere so that due to the presence of clouds the amount of CO2 absorption is significantly decreased. I would expect non-linearity effects and would like to understand better why this relation should always be linear.

Thank you for the question. We have observed that when reflectance becomes too low, the linearity of the relationship can deviate significantly. In other words, the main linear part would start from a certain reflectance, the mechanism of which remains unclear. This is one of the reasons why we established a transmittance threshold to exclude wavelengths associated with extremely low reflectance, thereby minimizing non-linear effects.

In cloud-shadowed areas, clouds can shield the lower atmosphere and reduce the total reflected radiance, introducing a negative intercept term. The ratio of CO_2 absorption to the total reflected radiance would remain relatively similar to clear-sky conditions, though the slope behaves differently under shadowed conditions. In some cloud-shadowed areas, however, we also observe non-linear behavior, which will require further investigation to fully understand. Nevertheless, our study primarily focuses on bright areas where the linear relationship remains more stable. The dynamics within cloudier or shadowed regions fall outside the scope of this analysis and would require a separate investigation to address potential non-linearities fully.

- l. 365: 1km² -> 1.25km²?

The area is 1 km² because the 5x5 grid points correspond to a 1 km x 1 km region, with each side of the grid representing 1 km in length.

- l. 366: "We excluded the data if the 25 nearest grid points contained cloud pixels used in the RT simulation."

Does this mean you consider only completely clear sky pixels? What about partially cloudy pixels for which CO2 retrievals are also performed, if the cloud fraction is not too high?

Partially cloudy footprints can pass the pre-screening process, but they are much fewer in number compared to clear-sky footprints close to clouds. In this paper, we focus primarily on footprints located in clear-sky regions to isolate the impact of 3D cloud scattering. Partially cloudy footprints are more complex, as they can be affected

by additional factors beyond scattering, such as cloud fraction and sub-pixel cloud variability. Addressing these would require further investigation, which is outside the scope of this study.

- l. 381: "Though it is instructive to discuss both cloud brightening and cloud shadowing effects, Massie et al. (2023) determined that there are relatively few cloud shadow retrievals in the OCO-2 Lite files, since many observations impacted by shadowing are screened by the OCO-2 pre-retrieval cloud screening algorithms. Thereafter, bright area analyses are the primary focus of our study."

As mentioned before, in the specific scene that you present it seems that you obtain largest differences in a cloud shadow region?

While some large negative-bias footprints do indeed appear in shadowed regions, the majority of positive-bias footprints are located in clear-sky areas. In our investigations, we observed that the version 11 dataset more effectively screens out these negative-bias footprints compared to version 10. We will move toward version 11 retrieval in the future. Footprints under cloud shadows encounter more complexities beyond cloud-induced photon scattering, such as surface albedo variations and reduced CO_2 absorption, which are not fully addressed by our current bright area mitigation approach. As a result, applying this method to cloud-shadowed regions may unintentionally worsen the bias. Our study focuses on mitigating biases in bright, clear-sky regions, while addressing cloud-shadowed areas would require a distinct and targeted strategy.

- l. 387: "we identified an exponential decay relationship between the 3D cloud effect parameters and the effective horizontal cloud distance (D, Fig. 6)"

What is the "bright area" (mentioned in the caption of Fig. 6) in the images, how do you select which is bright area and which is shadow?

The "bright area" is defined as the region where the reflectance calculated from the 3D-RT model is greater than that from the 1D-RT model. These areas show enhanced reflectance due to cloud-induced brightening effects. In contrast, regions where the reflectance is lower than that from the 1D-RT model are classified as cloud shadows, indicating areas where clouds reduce the reflected radiance.

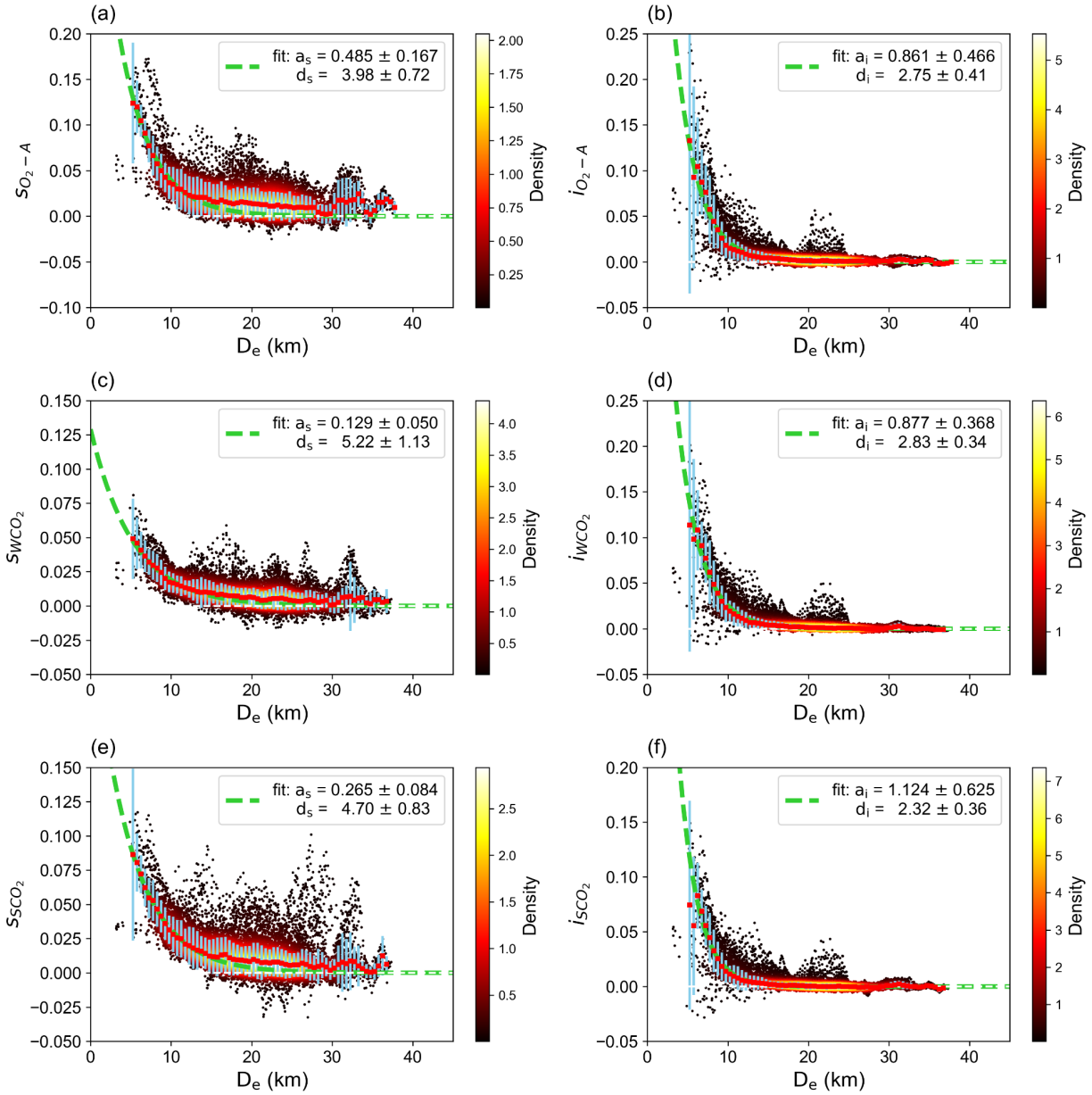
I assume that this exponential decay is only valid in the bright area, because in the shadow there are abrubt changes in reflectance where the shadow ends. Is this the reason, why you say that your method is focused on the bright area?

The exponential decay pattern can also be observed in shadow areas to some extent, but the relationship behaves in the opposite direction, with a decrease in reflectance instead of brightening. Our retrieval method is not optimized for footprints in shadowed

regions, as these areas present additional challenges, including severe changes in reflectance at shadow boundaries, as you mentioned. These complexities would require further investigation to address effectively, so we focus primarily on bright areas in this study.

In Fig. 6 I don't see the "background shading". Could you include colors corresponding to the density of the dots as in Fig.3?

Thank you for pointing out the issue. We understand that the density shading might be confusing. We will replace the density kernel with the density scatter plot as below:



(edited Fig. 6) Figure R16. The same figure as Fig. 6 but changing from heatmap to scatter plot.

- Table 1: These parameters are valid only for the specific scene, correct? This should be clarified. Is the number of digits meaningful?

Yes, the parameters in Table 1 are derived specifically from the results shown in Figures 5 and 6. Considering the uncertainty, the number of digits is not meaningful. We will update the table description to make this clearer as below:

- Original text (Line 433):
  "Table 1. Amplitude and e-folding distances for ***s*** and ***i*** fittings in the O_2-A, WCO_2, and SCO_2 bands."
- Revised text:
  "Table 1. Amplitude and e-folding distances for ***s*** and ***i*** fittings in the O_2-A, WCO_2, and SCO_2 bands for the simulation shown in Fig. 5."

- l. 516: "However, the bypass method is less precise than conducting a 3D-RT simulation with our baseline approach to derive s and i on a pixel-by-pixel basis."

Why is it less precise? I think it could even be more precise because it does not rely on assumptions to produce the input for the 3DRT simulations.

As mentioned earlier, the bypass method is a parameterized approach derived from our baseline method, which itself relies on 3D-RT simulation results. Therefore, it still inherits the assumptions made in the baseline method. While the bypass method offers a computational advantage, its precision can be lower because it uses generalized relationships rather than performing pixel-specific 3D-RT calculations, which can capture more localized variations.

- l. 517: "This bypass approach also disregards the presence of cloud shadows."

Why can the bypass approach not account for the shadows? Shadows are included in the observed radiances.

While it's true that shadows are present in the observed radiances, footprints over cloud-shadowed regions introduce additional complexities beyond the 3D cloud bias that the bypass method is designed to address. Adapting the bypass method to accurately handle shadowed regions would require further development and validation, which lies outside the scope of this study.

- Fig. 12: Are the results shown here based on the basic approach or the bypass approach? There are many shadows in the images. Do you include a cloud mask to exclude the shadows or how are they treated?

The results shown in Fig. 12 are based on the bypass parameterization listed in Table 2. The purpose of this figure is to demonstrate how variations in cloud distribution can lead to different cloud-induced biases. For this analysis, we consider the effective cloud distance for non-cloudy pixels, but we do not apply a cloud mask to exclude shadowed areas. As a result, cloud shadows are included in the images, but shadow-induced biases are not specifically addressed. Investigating these biases would require a more detailed analysis, which is beyond the scope of this study.

- l. 647: "... it allows the parameterization of the six spectral perturbation parameters themselves as a function of macroscopic scene parameters ..."

Which are the macroscopic scene parameters?

The macroscopic scene parameters include solar and viewing geometry (e.g., solar zenith angle, viewing angle), cloud properties (e.g., cloud top height, cloud optical thickness), aerosol properties, and surface albedo. For example, we have demonstrated how variations in surface albedo can influence the spectral perturbation parameters in our analysis.

Couldn't one use only the spectral radiance observations to obtain the perturbation parameters using the bypass method?

We have attempted to retrieve the perturbation parameters directly from the spectral radiance observations, but the results have not been satisfactory so far. We believe this is due to an insufficient degree of freedom for independently retrieving all six perturbation parameters, as they are highly coupled with other factors such as surface albedo and aerosol optical thickness. Decoupling these variables would require additional constraints or assumptions, which complicates the retrieval process.

- l. 656: "While the bypass method does capture the significant modulators of the 3D cloud effects, including surface reflectance and sun-sensor geometry, it is not granular enough to consider detailed scene variables such as cloud top height, cloud morphology, or aerosol load."

Shouldn't the bypass method, since it uses reflectances containing all 3D cloud effects, capture all modulators?

While the bypass method indeed uses reflectance that inherently captures all 3D cloud effects, it simplifies the variation in perturbation parameters by grouping them based solely on effective cloud distances. This simplification does not fully account for the spread in perturbation values that may arise due to more detailed factors, such as variations in cloud top height, cloud morphology, or aerosol distributions, which are not explicitly parameterized in the bypass method.

References:

Schmidt, K. S., Massie, S., and Feingold, G., 2019, June. Impact of Broken Clouds on Trace Gas Spectroscopy From Low Earth Orbit. In Hyperspectral Imaging and Sounding of the Environment (Optica Publishing Group, 2019), paper HW5C-2.

Chen, H., Schmidt, K. S., Massie, S. T., Nataraja, V., Norgren, M. S., Gristey, J. J., Feingold, G., Holz, R. E., and Iwabuchi, H.: The Education and Research 3D Radiative Transfer Toolbox (EaR3T) – towards the mitigation of 3D bias in airborne and spaceborne passive imagery cloud retrievals, Atmos. Meas. Tech., 16, 1971–2000, https://doi.org/10.5194/amt-16-1971-2023, 2023.

Nataraja, V., Schmidt, S., Chen, H., Yamaguchi, T., Kazil, J., Feingold, G., Wolf, K., and Iwabuchi, H.: Segmentation-based multi-pixel cloud optical thickness retrieval using a convolutional neural network, Atmos. Meas. Tech., 15, 5181–5205, https://doi.org/10.5194/amt-15-5181-2022, 2022.

Roccetti, G., Bugliaro, L., Gödde, F., Emde, C., Hamann, U., Manev, M., Sterzik, M. F., and Wehrum, C.: Development of a HAMSTER: Hyperspectral Albedo Maps dataset with high Spatial and TEmporal Resolution, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-167, 2024.

Emde, C., Yu, H., Kylling, A., van Roozendael, M., Stebel, K., Veihelmann, B., and Mayer, B.: Impact of 3D cloud structures on the atmospheric trace gas products from UV–Vis sounders – Part 1: Synthetic dataset for validation of trace gas retrieval algorithms, Atmos. Meas. Tech., 15, 1587–1608, https://doi.org/10.5194/amt-15-1587-2022, 2022.

Yu, H., Emde, C., Kylling, A., Veihelmann, B., Mayer, B., Stebel, K., and Van Roozendael, M.: Impact of 3D cloud structures on the atmospheric trace gas products from UV–Vis sounders – Part 2: Impact on NO2 retrieval and mitigation strategies, Atmos. Meas. Tech., 15, 5743–5768, https://doi.org/10.5194/amt-15-5743-2022, 2022.