Final Response for AMT-2024-145

Peter Somkuti[1,2], Gregory McGarragh[3], Christopher O'Dell[3], Antonio Di Noia[4,5,6], Leif Vogel[4,5,7], Sean Crowell[8], Lesley Ott[2], and Hartmut Bösch[4,5,6]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[2]Global Modeling and Assimilation Office, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD, USA
[3]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA
[4]University of Leicester, Leicester, UK
[5]National Centre for Earth Observation, Leicester, UK
[6]Institute of Environmental Physics (IUP), University of Bremen FB1, Bremen, Germany
[7]now at Kaioa Analytics, Mundaka, Viscay, Spain
[8]LumenUs Scientific, Oklahoma City, OK, USA

Correspondence: Peter Somkuti (peter.somkuti@nasa.gov)

Below are the point-by-point replies to review comments RC1 and RC2 for submission AMT-2024-145. We separated the two sets of comments into sections 1 (RC1) and 2 (RC2), and subsections within that address each comment individually. For each subsection, the original review comment is stated in **bold** text at the beginning of the subsection, followed immediately by our reply. If changes to the manuscript were made in the form of added text, they are put inside a distinct frame such as this:

> Example text added to the manuscript.

New figures or tables are generally added right below the blue frame that surrounds newly added text. Deleted text passages will not be explicitly displayed, but will be seen in the difference document, uploaded later on. Figure numbers are generally organized as follows: in text passages newly added to the document (blue-shaded frames), the figure number represents the number in the revised manuscript; in replies, the figure number is usually followed by e.g. "in the original manuscript" to clarify that the figure number represents the number as it appears in the initial submission; some figures inside this document are being referenced, and those are specifically named "RFIG" (and clickable in this PDF).

1 Review 1 - RC1

1.1 Comment 1

Clarify how BRDF wavelength dependence is treated in the OSSE forward simulation, as it is a key difference between forward and inverse models.

Indeed, in our forward model simulations, the BRDF wavelength dependence is based on a linear interpolation of BRDF parameters from the MODIS MCD43A1 dataset, which is spectrally given for the first 7 MODIS bands. Since the CH_4 band in our simulations and retrievals at 2.3 µm is outside of the wavelength range spanned by those bands, we use the values from band 7 (roughly 2.15 µm) across the band, without any spectral dependence. Therefore, in the forward simulations, the surface reflectance is spectrally flat. We added text at the end of section 2.1 to clarify this point, as well as make corresponding notes in the discussion section. Please also note comment 2.2 that further discusses this matter.

> The surface for the CH_4 band is spectrally flat since the MODIS instruments do not cover the shortwave-infrared region beyond ≈ 2.15 µm. Thus, we take the BRDF coefficients from band 7 and use them for all wavelengths within the CH_4 window, without any spectral variation.

1.2 Comment 2

Do ISCCP cloud observations cover the simulated period? When combined with the CAMS there could be inconsistencies from where the chemical transport model simulates clouds. For instances where there are clouds in CAMS but not ISCCP, the AOD may be overestimated because hygroscopic growth is being accounted for.

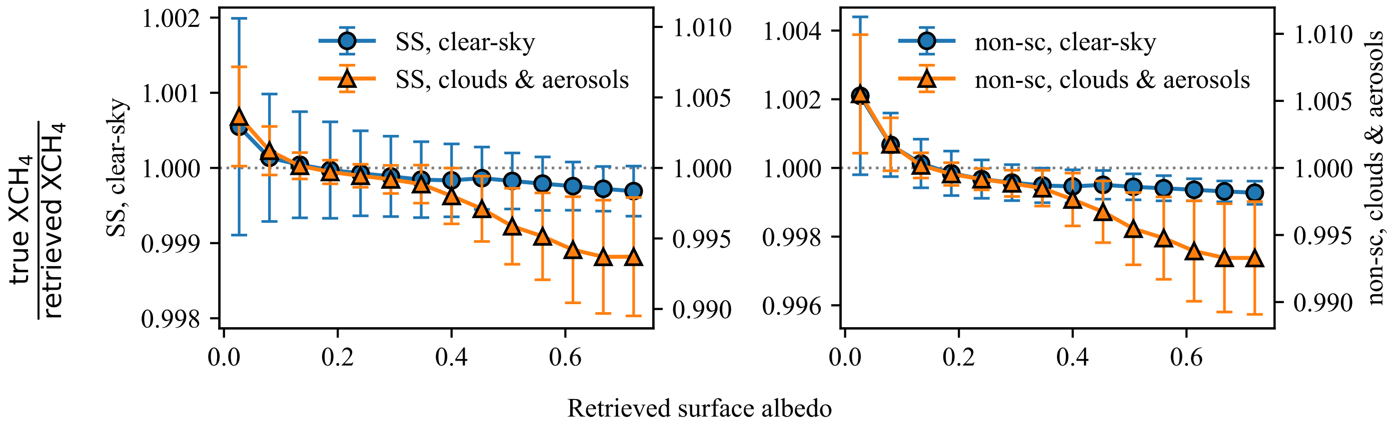
We use CAMS reanalysis for the forward simulation; as stated in Rémy et al. (2019), the CAMS system assimilates AOD from MODIS collection 6, and the Polar Multi-Angle Product, which itself is based on observations from GOME-2, AVHRR and IASI. The observations in the ISCCP dataset, as can be read in Young et al. (2018), themselves are a composite of various polar-orbiting and geostationary platforms that are aggregated into a combined dataset.

We have sampled both the ISCCP cloud dataset and the CAMS aerosol reanalysis at the same time/location pairs for each scene - so they indeed should be compatible in the sense that any biases occurring in the reanalysis due to extended cloud cover would impact scenes that likely contain clouds in our simulations, as informed by ISCCP.

We have not, however, gone through an exercise to validate any aspects that could relate to a potential mismatch between the ISCCP cloud flags (or other parameters we utilize, such as cloud top height) and errors in the CAMS reanalysis. In our study, we are primarily concerned with producing a realistic and geo-spatially reasonable atmospheric state that is informed by both models and observations. We then reference our retrieved results against this truth. We agree that it could be possible that the overall distribution of clear-sky aerosol loadings are skewed with respect to reality, but we believe that this would have negligible impact on our results.

1.3 Comment 3

Earlier in the section it was stated that Rayleigh scattering was insignificant, but is this still true for the small magnitude of the changes being considered in Fig. 7? For instance the negative bias over the dark surfaces seems consistent with atmospheric scattering, since photon paths from light scattered from the atmosphere will have a greater contribution to the total radiance at the sensor relative to the brighter surfaces. To first order this would be radiation from the solar beam directly scattered into the path of the sensor, which would effectively shorten the light path, making a forward model that does not account for it reduce the CH4 column, as is shown in the figure. Since the magnitude of the biases



RFIG 1. Fig. 9 in the originally submitted manuscript.

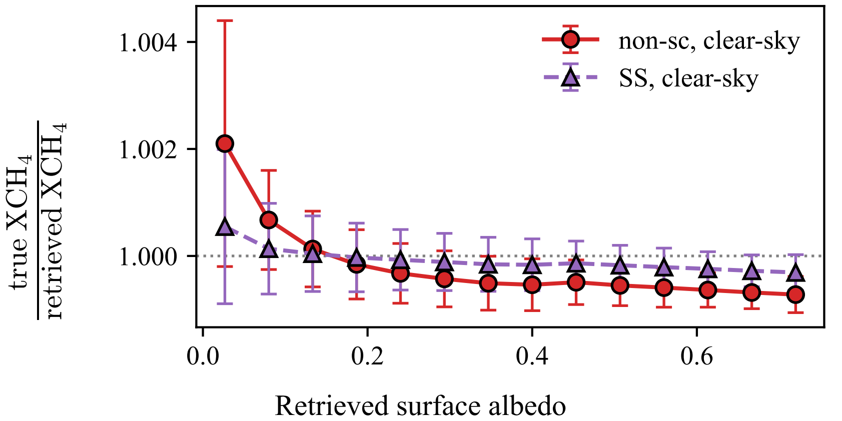
reduces between the non-sc and SS cases, this supports Rayleigh being potentially important for biases at this level. It should also be noted that a 2-5 ppb bias shown in the Figure may actually be significant depending on the application. E.g. A lot of diffuse agricultural sources produce enhancements around this magnitude. If anything the results show how close to perfect a retrieval would have to be to quantify these.

We thank the reviewer for this highly observant comment, they have indeed pointed out an aspect of our simulations and retrievals that we failed to observe, or rather, failed to realize the significance of accounting for Rayleigh scattering at 2.3 µm. In order to assess the importance of accounting for Rayleigh scattering, we have performed the retrievals again in several different configurations and are focusing on the impact it has on the surface bias.

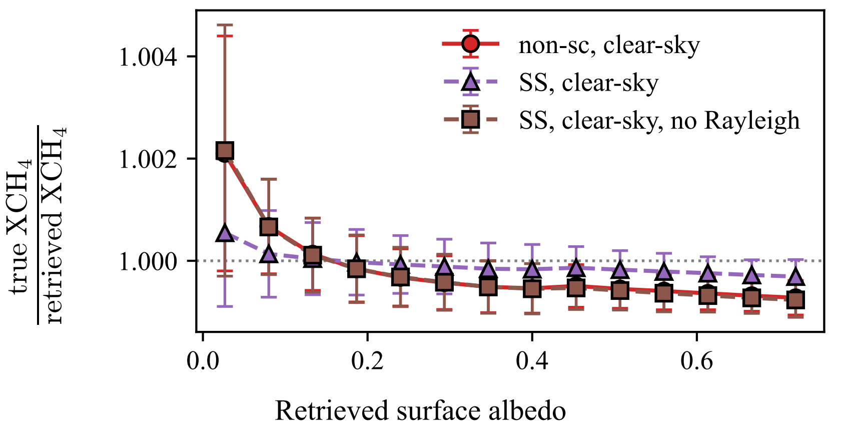
First, we re-visit the originally offending data that corresponds to the two experiments named (CS1) and (CS2) - retrievals done on clear-sky simulations with the single-scatter RT module (CS1) and the non-scattering Beer-Lambert type RT (CS2). Fig. 9 in the original manuscript does not show those two experiments on the same panel, they are the blue curves with round markers on both panels in RFIG. 1.

We can overlay the two curves representing (CS1) and (CS2) into a single figure, shown here in RFIG. 2. One can clearly see the difference between both curves that represent this surface bias: the experiment (CS1) with the non-scattering RT exhibits a larger magnitude, although the overall structure is the same.

Initially, we assume this discrepancy to be the result of minute details related to the different RT modules. After all, the RT module used for (CS2) is a fully-fledged RT framework capable of utilizing multiple solvers in both scalar and vector mode. For example, our retrieval algorithm, when using that XRTM module, enables the pseudo-spherical approximation by default - and even though we do not tend to be in the regime of viewing and solar angles where that would be a major driver for errors, one could imagine small discrepancies become evident in an analysis such as the one we are performing - this was in fact the reason why we used both RT models: to make sure that our results are robust.



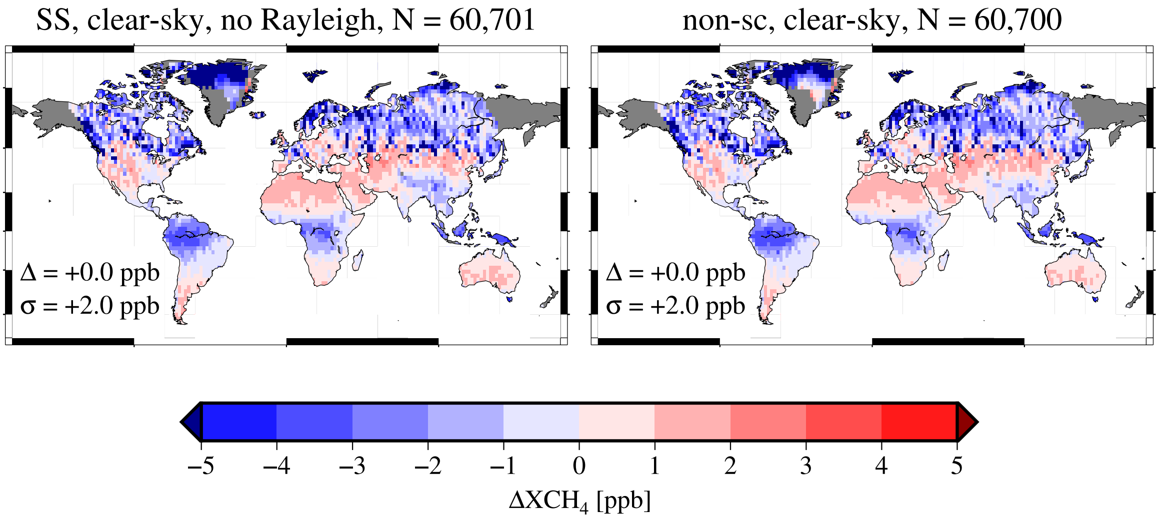
RFIG 2. Overlaying the two curves (blue, round markers) from both panels from RFIG. 1. Note the difference in colors.



RFIG 3. The surface bias curves with the modified experiment (SS, clear-sky, no Rayleigh) to show that the difference between experiments (CS1) and (CS2) arises due to Rayleigh scattering and its proper include in the RT.

Now, we are setting up the retrievals once more - with the main difference that we set the optical depth due to Rayleigh scattering to zero everywhere, the single-scatter albedos are therefore also zero. This allows the XRTM module to still execute correctly, however there will be no contributions due to scattering. When we analyze the bias curves again, we see that the results from (CS2), now without having Rayleigh scattering contributions in the retrieval part, lines up very closely with the results from (CS1).

This result demonstrates clearly that ignoring Rayleigh scattering in these single-band retrievals in clear-sky conditions has a significant impact on the magnitude of these XCH_4 biases, and that indeed Rayleigh scattering is the driving difference between



RFIG 4. The maps corresponding to the modified experiment without Rayleigh scattering on the left, and the original experiment (CS1) on the right. Apart from minor differences, the spatial distribution and magnitude of the errors are practically the same.

experimental set-ups (CS1) and (CS2). For the sake of completeness, we can also see that in spatial context, the two retrieval set-ups yield closer results, which is shown in RFIG. 4.

We edited the following sections in the manuscript to reflect our new understanding of how Rayleigh scattering impacts our results:

1. In Section 2.3 (which describes the aerosol set-up), we removed the statement that Rayleigh scattering at 2.4 µm has negligible contributions to the simulated TOA radiance.

2. In Section 2.4 (which describes the retrieval algorithm approaches), we call out Rayleigh scattering in a more pronounced way to highlight the results to come.

> Here we want to emphasize that the non-scattering RT model will account for extinction due to Rayleigh scattering, as it is calculated as part of the layer-resolved total optical depths. Using the XRTM library, however, the retrieval algorithm forward model will include the contributions from Rayleigh scattering in addition to the extinction.

3. In Section 4.1 (which discusses the results of the clear-sky case), we removed the mention of Rayleigh scattering being negligible, and also re-phrased that paragraph to clarify how Rayleigh scattering is (or is not) accounted for in the simulations and the retrievals. Further, we added a paragraph towards the end of the section to explain the impact of Rayleigh scattering on the results discussed in this section. The figure shown here (RFIG. 3) with accompanying text is added to a new appendix section, as we felt that the additional figure would distract from the overall flow of the manuscript within Section 4.1.

[Added to Section 4.1] The underlying cause for the difference between the results of the experiments (CS1) and (CS2) is the correct accounting for Rayleigh scattering in the retrieval forward model. The non-sc configuration does not produce scattered contributions to the TOA radiance, and thus shows larger errors overall, which can be observed by the scatter of the data in Fig. 7 (SS: $\sigma = 1.2$ ppb vs. non-sc: $\sigma = 2.0$ ppb). We have confirmed the impact of Rayleigh scattering by performing retrieval experiment (CS2) with a modified set-up in which the optical depth due to Rayleigh scattering was forced to be zero, details on that modified experiment can be found in Appendix A.

[Added to (new) Appendix A] In this appendix section we are clarifying the impact of Rayleigh scattering on the clear-sky experiments, as mentioned in Section 4.1.

Observing again Fig. 10, we notice that the magnitude of the bias seen in the two experiments (CS1) and (CS2) (see Fig. 6 for the meaning of those labels) is different. The only change between experiments (CS1) and (CS2) is the used RT scheme in the retrieval: (CS1) uses the Beer-Lambert-Bouguer law for an absorption-only atmosphere, (CS2) uses the single-scattering solver through the XRTM library (McGarragh, 2020). In both experiments, the retrieval forward model computes all necessary contributions from Rayleigh scattering and produces the same total optical parameters: optical depth, single-scatter albedo, and the appropriate phase function. In experiment (CS1), however, only the optical depth is passed on to the RT routine which computes the TOA radiances. Rayleigh scattering contributions are always present in the synthetic observations (the simulation forward model).

It can be shown that the difference in magnitude of the bias can be fully attributed to Rayleigh scattering. We perform a control experiment, which is a modified run of experiment (CS2). We then manually set the optical depth due to Rayleigh scattering to zero everywhere, which automatically leads to the single-scatter albedos being zero everywhere as well. Therefore, we have a retrieval forward model that is does not account for Rayleigh scattering at all.

After performing the same actions on the resulting dataset as before (basic quality screening and bias correction), we compare the results of this modified run with the results of experiments (CS1) and (CS2). In Fig. A1, we can now observe that the modified run is near-identical to experiment (CS1). We note that this is a surprising result, as Rayleigh scattering often is ignored in studies when wavelengths > 1 µm are concerned (e.g. Jongaramrungruang et al. (2021)). Our control experiment shows that the inclusion of Rayleigh scattering can make a significant difference and warrants consideration.

1.4 Comment 4

The physical explanation for the curves in Fig. 9/10 should be discussed - e.g. see section 3.1 of Aben et al. (2007). https://doi.org/10.1016/j.jqsrt.2006.09.013.

We added a paragraph in Section 4.2 which discusses the explanation of the shapes of the curves according to Aben et al. (2007).

> A phenomenological explanation for the shape of the bias curves in Fig. 10 has been stated in Aben et al. (200). When aerosols are present, some fraction of the incident light is scattered into the field of view of the instrument. For scenes with low surface reflectivity, there is a relatively larger amount of light that has a shorter total light path from contributions which are scattered towards the instrument before reaching the surface. A retrieval algorithm that does not account for aerosols can thus only reduce the methane abundance to match the observed radiances. In Fig. 10, an underestimation of XCH_4 is equal to a ratio of true to retrieved XCH_4 larger than 1. On the other extreme, for very bright surfaces, the fraction of light that travels the full path through the atmosphere twice, is comparatively larger. In addition, contributions from multiple scattering due to tropospheric aerosols further increase the effective light path of photons. Without accounting for aerosols, the retrieval algorithm can only increase amount of CH_4 in the atmosphere to match the observed absorption lines, causing an overestimation. This overestimation shows up as a ratio of true to retrieved XCH_4 smaller than 1. We further note that this explanation should hold true for an aerosol-free atmosphere in which Rayleigh scattering occurs. Despite the fact that the total-column optical depth due to Rayleigh scattering at 2.3 µm amounts to only $\approx 10^{-4}$ (Tomasi et al., 2005), the impact is significant and can be observed in these bias curves (see also Appendix A).

1.5 Comment 5

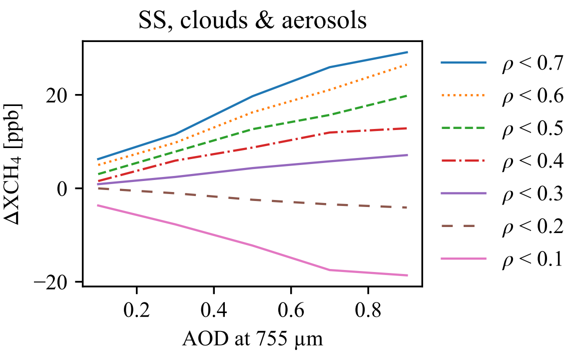
Also these relationships between albedo/aerosol loading are well known. I would possibly reframe this section as a quantification of this type of bias for CH4 in the 2.35 micron band. At the moment it reads like somewhat of a discovery, but these effects have been written about since at least SCIAMACHY e.g. Houweling et al. (2005) https://doi.org/10.5194/acp-5-3003-2005

We added relevant text in the Introduction section (1) as well as additional discussion of the results of Howeling et al. (2005) and Aben et al. (2007) into the Discussion & Conclusions section (5).

> **[added to Section 1]** Aerosols have been identified as a cause for systematic biases in retrievals since the early days of

space-based missions that allowed for the estimation of greenhouse gas total columns. While not exclusively dedicated to greenhouse gases, the SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY) instrument (Bovensmann et al., 1999), was used to retrieve both CO_2 and CH_4 from the near-infrared (NIR) or short-wave infrared (SWIR) part of back-scattered spectra. Biases related to aerosols have been observed over the Sahara by Houweling et al. (2005) and were further studied by Aben et al. (2007) in a more comprehensive simulation exercise. Their conclusions are highly relevant to our study, as they also investigate a single-band retrieval configuration and observe the interaction between aerosol loading and surface reflectivity. We will contextualize our results in that regard in the Discussions section later on (Section 5). For the first dedicated CO_2 and CH_4 mission, the Greenhouse gases Observing SATellite (GOSAT) (Kuze et al., 2009), aerosols were also understood as a cause of bias in the retrieved total columns (Wunch et al., 2011; Uchino et al., 2012; Cogan et al. 2012). However, the way how aerosols interact with the various retrieved quantities is different for retrievals from GOSAT, compared to those of SCIAMACHY. GOSAT provides measurements of two separate absorption bands of CO_2, at 1.6 μm and 2.06 μm (in addition to thermal bands which are not relevant here), which provides effective de-coupling of the surface from the CO_2 concentration. As such, significant surface-related biases have not been observed in GOSAT retrievals for CO_2. The Orbiting Carbon Observatory missions (Crisp et al., 2004; Crisp, 2015; Eldering et al., 2019) use roughly the same configuration in terms of observed spectral windows as the GOSAT mission, and retrieval algorithms used for either instrument are mostly interchangeable. Kulawik et al. (2019) found using another simulation study that the retrieved aerosol optical depth and retrieved surface albedos were indeed correlated when they inspected the posterior covariance matrices. However, as is seen in most related studies (e.g. O'Dell et al. (2018)), the major drivers of biases are retrieved surface pressure as well as the retrieved CO_2 profile shape, and retrieved aerosols and surface albedo contribute much less to the total bias correction (OCO-2 Science Team, 2023). The utilization of 3-band retrievals from GOSAT, OCO-2 and OCO-3 have reduced possible biases due to surface-aerosol interactions such that they are no longer a dominant contribution to the total observer errors.

[added to Section 5] These types of aerosol-driven biases in retrieved trace gas columns have been studied in the past



RFIG 5. [Fig. 13 in the revised manuscript] A re-ordered view of the results from Fig. 10. Each curve represents a sub-set of scenes whose retrieved, apparent surface albedo (ρ) falls into some bin. This way of ordering the results mimics Fig. 3 from Aben et al. (2007) and shows a qualitative match. The order of the legend items is the same as they appear in the figure: the lowest curve (pink, solid) contains scenes with $\rho < 0.1$, the curve above (brown, dashed) contains scenes with $0.1 < \rho < 0.2$, and so forth. Further, the quantity plotted is not the ratio of true over retrieved XCH_4, but the difference of retrieved minus true XCH_4.

and are not exclusive to TROPOMI. We return to the notable studies of high relevance to ours by Houweling et al. (2005) and Aben et al. (2007), which both explored the topic in the context of the SCIAMACHY instrument. In Houweling et al. (2005), they find significant XCO_2 biases in the Sahara region due to high aerosol loadings, which Aben et al. (2007) further elaborated on via a sensitivity study with retrievals from simulated observations. By re-ordering our results in a different manner in Fig. 13, we can reproduce their findings qualitatively, specifically Fig. 4 in Houweling et al. (2005) and Fig. 3 in Aben et al. (2007). One can consider our results to be an extension of their studies to XCH_4 in the 2.3 µm absorption band (rather than XCO_2 from the 1.6 µm band). The stark difference in magnitude of the effect is likely due to the instrument characteristics. For example, SCIAMACHY's spectral resolution ($\approx$ 1.35 nm) is over five times lower than that of TROPOMI ($\approx$ 0.25 nm). The deciding common aspects of our study and those of Aben et al. (2007) are the following: (1) both studies use a single spectral window to retrieve a trace gas, (2) the absorption features within the chosen retrieval window are such that there is no clear continuum to sufficiently de-couple surface reflectance from gas concentration, (3) the studied atmospheric states contain weakly scattering aerosols up to total optical depths of ≈ 1. Our study does not, however, include thin high-altitude cirrus clouds.

1.6 Comment 6

I suspect that the 1.65 micron band is susceptible to the same aerosol/albedo related biases as the 2.3 band when retrieving the absolute column because the physical light path shortening/lengthening effects are still at play. For retrievals

based on the CO2-proxy method they may be lessened for moderate aerosol scenes (assuming that the albedo difference between the CO2 and CH4 bands is similar).

Similar to Comment 2.1, while our simulation framework is capable of performing that task, it would be a significant amount of work to produce the simulations including the 1.6 micron band. We would expect the effect to appear, but much less in magnitude, mostly due to the nature of the spectral features. Methane absorption lines at 1.64 μm do not show broader continuum absorption as those in the 2.3 μm, so the naïve thought here would be that the de-coupling between surface and trace gas amount would be sufficient.

1.7 Comment 7

Nit - "surface albedo" should be "Lambertian surface albedo" (assuming that is what is being fit).

We have corrected this in the text.

1.8 Comment 8

What are the wavelength(s) corresponding to the radiances that are used to estimate the albedo

We use all points from the measured radiance which lie inside the bounds of the retrieval window. We added a short sentence near Equation 3 to clarify this.

> Note that we use the all points from the measurement that fall inside the retrieval band to calculate $\max(I)$.

1.9 Comment 9

Where do the trace gas profiles come from?

We have referenced our earlier study (Somkuti et al., 2021) which contains more details of the simulation set-up. However, we admit that this information might be interesting to readers, so we added a sub-section with a table detailing the source of the trace gas fields. Note also comment 2.7 - we further added the spatial resolution of trace gas fields and the other data sets we use as inputs into our simulations.

> In Table 1 we summarize the various data sets that feed into our simulations and also note the spatial resolution at which those data sets are provided.

1.10 Comment 10

SS and non-sc don't need to be in brackets as they were defined in Section 4.

In this case, the intention is not to introduce the meaning of the short-hand, but merely as a reminder and we would like to keep the current formulation which uses the parentheses.

Table 1. Source and spatial resolution for the key datasets used in the simulations

Data	Source	Spatial resolution
Surface BDRF parameters	MODIS MCD43A1 (Schaaf and Wang, 2015)	$\approx$ 500 m
CO_2, CH_4 and CO profiles	custom GEOS-5† (Molod et al., 2015)	$0.625° \times 0.50°$
Aerosol profiles	CAMS reanalysis (Bozzo et al., 2020)	$0.75° \times 0.75°$
Cloud parameters	ISSCP (Young et al., 2018)	$0.1° \times 0.1°$
Meteorology	ECMWF ERA5 (Hersbach et al., 2020)	$\approx$ 31 km

† Trace gas profiles were sampled from a custom GEOS-5 run at 50 km spatial resolution, written out to $0.625° \times 0.50°$.

1.11 Comment 11

Since there are only two sets does it make more sense to say “non-sc and SS produce offsets of x and y ppb respectively” rather than “each set exhibits an offset between -6 and -3ppb.

We have changed the sentence.

2 Review 2 - RC2

2.1 Comment 1

I understand that interesting questions may be out of scope of this study, but I would encourage the team to pursue investigations on a couple of fronts: (1) retrievals in the 1.65um, where historically based on GOSAT, scattering biases have been deemed unsignificant. Given current and upcoming missions planning their retrievals from the 1.65um band, it would be incredibly valuable to have this analysis done with those missions in mind. (1) For scenes measured in sun-glint mode, given the relevance of methane measurements offshore, greatly affected by aerosol transport from onshore, to characterize methane emissions.

We very much appreciate the supportive sentiment and comment. At the time when we created the simulations, the "weak CO2 band" is included in the forward radiative transfer runs; however unfortunately that band does not extend beyond 1.623 µm as it was originally set up to reflect the band passes of the GeoCarb instrument. If time and resources permit, we might consider running the forward simulations again with an extended set-up that includes methane absorption at 1.65 µm.

Our simulations did include synthetic measurements for sun-glint observation geometries (as prescribed by real OCO-2 viewing angles in glint-mode), however we were not able to reconcile the results observed for that sub-set. As we describe in the Discussion section (5) of the manuscript, sun-glint pointing measurements for both land and ocean surfaces have behaved out-of-family such that we chose to exclude them from the analysis. As with the extended 1.65 µm band, we will consider how to incorporate those scenes into a future analysis.

2.2 Comment 2

Given that a qualitative change in TROPOMI retrievals involved modifying the order of the polynomial that captures spectral dependence of the surface albedo, has the team considered analyzing this aspect? Additionally, it would be useful to discuss the findings in light of Jongaramrungruang et al. (2021) and the spectral differences between GeoCarb and TROPOMI, as mentioned in the discussion

In our simulation (forward) model, the surface reflectance is computed in the following way. We sample the MODIS BRDF product (MCD34A1) at the scene locations and times, which provides BRDF parameters at the wavelengths of the MODIS bands 1 through 7. At the edges of each of the simulated bands we obtain the BRDF parameters via linear interpolation (in spectral wavelength space) from the closest MODIS bands.

For the CH_4 band, however, which ranges between ≈ 2.300 µm and ≈ 2.345 µm, there is no corresponding MODIS band beyond the upper band boundary. In cases like this, we use flat extrapolation from the closest valid point, which are the BRDF parameters from MODIS band 7 near ≈ 2.1 µm. Thus, the BRDF parameters in the forward simulations are spectrally flat for the CH_4 band. While we do retrieve a slope parameter for the surface reflectance, we believe it is mostly compensating for the spectral variation of scattering properties in the forward simulation. Adding another order to the polynomial does not have any meaningful impact - we have tested this now and also added an appropriate paragraph to the Discussion section.

> In Lorente et al. (2023), the bias takes on a slightly different shape when the spectral dependence of the retrieved Lambertian surface albedo is changed from a second- to a third-order polynomial. This is not a feature that we can investigate with our simulations since the surface model in our simulations, which produce the synthetic observations, is spectrally flat. This is solely a constraint of the underlying observation-based dataset (Schaaf and Wang, 2015) which does not provide measurements beyond ≈ 2.15 µm. Therefore, we also cannot investigate the impact of adjusting the order of the retrieved Lambertian surface albedo polynomial.

The study by Jongaramrungruang et al. (2021) focuses on broader retrieval windows with a clear focus on instruments such as AVIRIS-NG, and they are discussing errors that are highly relevant to plume quantification. While the results of our study, in principle, are valid with respect to those types of applications, it is not clear what we can learn from our global dataset applied to those spatially highly localized scenarios. Since plume detection and plume source quantification is a topic of importance, we have added additional text to the discussions in which we mention the few possible conclusions that could be drawn from our study.

> For point source-related applications, such as emission rate quantification, our results remain transferable only in a limited

sense. In a clear-sky environment, the surface variation will already imprint onto the retrieved XCH_4 field, which can impact the estimation of emission rates. This has been studied to some extent by Jongaramrungruang et al. (2021) for various instrument configurations, focused on spectral windows much wider than that used in our study, and more realistic surfaces. They find, in general, lower precision errors when retrieving from the 2.3 µm band, compared to the 1.6 µm one, however they do not account for scattering from either Rayleigh scattering or aerosols. In scenarios with substantial background aerosols, the surface imprint onto the retrieved XCH_4 would show higher magnitude (compared to the same scene on a day without aerosols). A more difficult scenario would be the co-emission of aerosols from the methane point source, as an empirical correction using retrievals outside of the main plume might not fully capture the bias inside the plume.

2.3 Comment 3

Given all the aerosols scenarios available in the simulations, are there any other insights besides the aerosol extinction optical depth, maybe also considering spatial distribution of aerosols?

We did look at this particular aspect, and noted in the discussion section that "[...] the bias seems to be driven by the total aerosol extinction optical depths in the scene, rather than the vertical distribution or how absorbing the aerosols are." We did not see any notable dependence of the biases observed as a function the aerosol types used. We understand, however, that our set-up is not best equipped to answer this question. Due to the nature of our simulations, we are much more likely, for example, to see scenes with high amounts of mineral dust over bright surfaces (deserts). So the aerosol type abundances are closely tied to surface reflectance due to their geographical locations, as they occur in nature. A study with systematic experiments is better suited to quantify the impact of various aerosol types over the full range of possible surfaces.

2.4 Comment 4

Has the team conducted any simulations for non-nadir viewing geometries? This may be particularly relevant for new instruments with higher spatial resolution, which focus on plume retrievals and also thinking on bidirectional reflectance distribution function (BRDF) effects. Exploring whether these factors impact retrieval biases could provide valuable insights.

Our simulations were informed by down-sampled OCO-2 observation geometries, which generally fall under either nadir-viewing or sunglint-following. Unfortunately, we therefore have no way of mimicking plume mappers with the simulations we have currently in hand.

2.5 Comment 5

It would be helpful if the authors provided more insights into why they believe the bias is already present in the absorption-only atmosphere. Additionally, the difference in the sign of the bias between Figure 6 and Figure 7 is not fully discussed. Some areas with a positive bias in Figure 6 appear as stronger negative biases in Figure 7. When averag-

ing the maps into the curves that depict surface reflectance bias, some details about the underlying spatial distribution of the biases may be lost and are not explicitly discussed. Addressing these aspects would improve clarity and strengthen the conclusions.

We discuss this very briefly at the beginning of Section 4.1 in which we present the results from clear-sky experiments. The model that generates the synthetic observations is different from the retrieval (forward) model, hence minor biases are expected. Further, there is strong correlation between retrieved surface reflectance and the retrieved CH_4 that arises due to the similarity between the corresponding Jacobians. This is a feature inherent to the spectral window in both TROPOMI and GeoCarb instruments. We added explanatory text at the end of Section 4.1 along with a new figure (RFIG 6) showing the two Jacobians for an example scene.

> We assume the surface-dependent errors to emerge due to an inherent link between the retrieved XCH_4 and the apparent surface reflectance. This can be easily observed by analyzing the relevant entries in the Jacobian matrix of our forward model. Overlaying the Jacobians for the Lambertian surface albedo polynomial order 0 and the CH_4 profile scale factor, as we did in Fig. 11, we see that they match to some extent in their shapes, and their similarity can be stated with an overall correlation coefficient of $R = 0.91$ for an example case with surface albedo of ≈ 0.1. A more appropriate quantification of the similarity of those two state vector elements would be the construction of a correlation via the posterior covariance matrix $\hat{\mathbf{S}}$: $\mathrm{C}_{ij} = \hat{\mathrm{S}}_{ij} / \sqrt{\hat{S}_{ii} \cdot \hat{S}_{jj}}$. This quantity $\mathbf{C}$ does not just represent the similarity of two entries of the Jacobian matrix i and j, but also accounts for the instrument noise. Again, for this particular example displayed in Fig. 8, the relevant entry in $\mathbf{C}$ is approximately 0.51. The correlation is strong for this retrieval setup since the absorption features of methane in this wavelength range do not show a distinct continuum. In plain terms, if such a correlation is seen in a retrieval forward model, the inversion will generally produce a weighted adjustment between the two offending state vector elements in order to minimize the cost function and minimize the mismatch between modeled and measured radiances. It is important to note that the strength of a correlation is not indicative of the magnitude of the effect on the retrieved quantities. So despite observing such a high correlation between retrieved surface reflectance and CH_4 scale factor, the overall impact is shown to be only a few parts per billion, or a few tenths of a percent in relative terms.

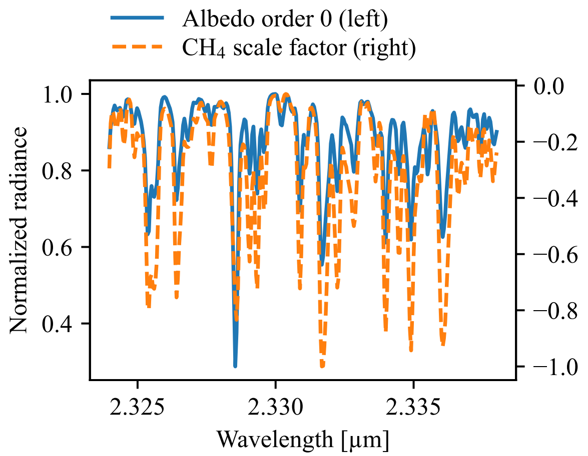
2.6 Comment 6

One remark is to make the figures a little bit bigger, particularly Fig. 8, 9 for the final version of the manuscript.

We have new scaled figures to be more legible in a final 2-column layout, and hope that AMT's editorial office will assist with the proper setting for the final version.

2.7 Comment 7

What is the spatial resolution of the simulations?



RFIG 6. [Fig. 8 in revised manuscript] An illustration of the similarity between two Jacobians: Lambertian surface albedo polynomial order 0 (blue, solid, left axis) and CH_4 scale factor (orange, dashed, right axis). Both are shown normalized, but on separate ordinate axes to show the strong similarity of their shapes. The Pearson correlation coefficient of these two Jacobians is $R = 0.91$.

Consistent with the notion of a 1-dimensional radiative transfer scheme (so no 3D RT effects), our (forward) simulations for TOA radiance have no inherent spatial structure, and can be seen as point-like in the spatial dimension. The underlying datasets which feed into the various parameters of the simulations, however, do have different spatial resolutions. We amended the text in Section 2.1 to add a table which states the data sources as well as their spatial resolutions. See Comment 1.9 for the table.

2.8 Comment 8

Have you looked at scenes covered by snow?

We do not have an explicit way of determining which of our scenes are covered by snow. We take observations from MODIS instruments (the aforementioned MCD43A1 product) and ingest them as long as they are considered "good" via their quality flag included in the data product. There are likely snow-covered surfaces ingested into our dataset.

We can, however, estimate snow-covered surface via the approach used by Wunch et al. (2011) (https://doi.org/10.5194/acp-11-12317-2011). Via this "blended albedo" approach, we take the effective surface albedo (the equivalent surface albedo needed to produce the TOA radiance as seen after the RT calculations) for the O2 A-band (0.76 µm) and the "Strong CO2" band (2.05 µm), which are produced in our simulations, but not used in the analysis, and screen for scenes with a blended albedo > 1. There are a total of 5,000 scenes (out of 60,000) which would be flagged as snow-covered with this method, so roughly 12%. We have found nothing particular about these scenes, likely because the underlying data from MCD43A1 does not employ a snow-specific BRDF model.

2.9 Comment 9

Are low albedo values filtered in the analysis? Asking particularly after seeing the remanent bias for the lowest bin (green square, Fig. 11), and recalling the surface albedo threshold of 0.02 in the TROPOMI quality filtering or 0.05 used in many applications.

Indeed we exclude scenes with a true (effective) surface albedo lower than 0.01. The purpose of this filter is to simply exclude very dark scenes, and only 3,500 scenes of the entire data set match that criterion. Considering only the retrievals which pass the most basic of our quality filter (solar zenith < 75, retrieval converged and valid), less than 40 remain. We thus only remove possible outliers by removing these scenes with very dark surfaces.

2.10 Comment 10

Can you, from this analysis, conclude on the usefulness of the oxygen A-band? Is the solution for TROPOMI to ingest "real" aerosol information in the retrieval? Can this be operationalized?

We mention the limits of our analysis in the Discussion & Conclusions section, however we admit the question of the single- vs. multi-band retrieval was not extensively discussed.

The retrieval set-up we used for this study does not allow for multi-band retrievals as we use a dedicated single-band algorithm. For us to perform a similar analysis that includes the oxygen A-band (and a set of different configurations) would require a significant effort which we currently cannot support. We further believe that our analysis cannot make any meaningful conclusions with respect to the utility of using the oxygen A-band. However, we have performed retrieval tests for a related exercise which happened to use the same family of scene inputs (CAMS aerosols, instrument model, surfaces) but used a full 4-band retrieval set-up using the ACOS algorithm adapted for the GeoCarb instrument. Even in those retrievals, where aerosols are co-retrieved, we can observe the surface bias for the retrieved XCH_4. While this could not be considered strong proof, it strongly suggests that both adding the oxygen A-band as well as co-retrieving aerosols (whose priors are different than the truth) do not solve the issue fully.

References

Bozzo, A., Benedetti, A., Flemming, J., Kipling, Z., and Rémy, S.: An aerosol climatology for global models based on the tropospheric aerosol scheme in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 13, 1007–1034, https://doi.org/10.5194/gmd-13-1007-2020, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Jongaramrungruang, S., Matheou, G., Thorpe, A. K., Zeng, Z.-C., and Frankenberg, C.: Remote sensing of methane plumes: instrument tradeoff analysis for detecting and quantifying local sources at global scale, Atmospheric Measurement Techniques, 14, 7999–8017, https://doi.org/10.5194/amt-14-7999-2021, 2021.

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., and Landgraf, J.: Accounting for surface reflectance spectral features in TROPOMI methane retrievals, Atmospheric Measurement Techniques, 16, 1597–1608, https://doi.org/10.5194/amt-16-1597-2023, 2023.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2, Geoscientific Model Development, 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Rémy, S., Kipling, Z., Flemming, J., Boucher, O., Nabat, P., Michou, M., Bozzo, A., Ades, M., Huijnen, V., Benedetti, A., Engelen, R., Peuch, V.-H., and Morcrette, J.-J.: Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1), Geoscientific Model Development, 12, 4627–4659, https://doi.org/10.5194/gmd-12-4627-2019, 2019.

Schaaf, C. and Wang, Z.: MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global - 500m V006, https://doi.org/10.5067/MODIS/MCD43A1.006, 2015.

Somkuti, P., O'Dell, C. W., Crowell, S., Köhler, P., McGarragh, G. R., Cronk, H. Q., and Burgh, E. B.: Solar-induced chlorophyll fluorescence from the Geostationary Carbon Cycle Observatory (GeoCarb): An extensive simulation study, Remote Sensing of Environment, 263, 112 565, https://doi.org/10.1016/j.rse.2021.112565, 2021.

Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., and Rossow, W. B.: The international satellite cloud climatology project H-Series climate data record product, Earth System Science Data, 10, 583–593, https://doi.org/10.5194/essd-10-583-2018, 2018.