Response summary

We have addressed every reviewer comment. From reviewer 2 we did not change the manuscript for two comments:

1) **What is the effect on the OE retrieval? The reviewer noted this might be too much work for this paper and we agree, because we know there are other factors that now need correcting. It would substantially delay things to integrate ReFRACtor into the retrieval, modify the code for the new v8/v9 OCO-2 L1bSc data changes, and re-run the very expensive retrieval when we know the output will still likely have big biases before we fix the vertical structure issue. This is ongoing work.**
2) **Could truncation of the lookup table in P_top cause our biases? We didn't think this was likely since it should have caused biases predominantly at high or low P_top, and Figure 5(c) doesn't show this. Nevertheless, we reprocessed the retrieval with a more restricted P_top range, showing that this doesn't explain our bias. We showed these results in the review response but think it would have been an unjustified distraction in the main paper so did not make changes.**

These are justifiable exclusions and otherwise the reviewers helped us to tidy up the paper nicely, we hope you will agree.

We've put in page breaks and big colourful titles to separate each section, the sections are:

Reviewer 1 response: p2—3
Reviewer 2 response: p4—9
Redlined paper: p10—33.

Reviewer 1 response

The paper is well written, and in principle can be published as it is now.I have two minor questions:

Thank you for your positive review. We answer your specific questions below with our text in bold.

We tried to keep our text changes as short as possible to maintain the manuscript's flow. This response expands in more detail to provide context so you can confidently judge whether our streamlined text is clear & accurate.

Line 156: I am confused with the sentence, that LIDORT assumes semi-infinite atmosphere. What is the benefit in assuming semi-infinite atmosphere?

You're right to be confused, our phrasing was bad so we have changed it. The OCO-2 RT bolts together multiple codes. Multiple scattering is very important for us so we specifically mention LIDORT,[1] which uses an infinite-medium solution method for the radiative transfer equation particular integral. However, the OCO-2 code does not otherwise assume a semi-infinite atmosphere so we removed "semi infinite".

Single scattering is handled separately[2], then there's the second-order-of-scattering code[3] polarisation correction and also the low stream interpolator[4] (which is also 2OS corrected), based on a successive order of interaction[5] approach. ReFRACtor combines these as described in O'Dell et al. (2012)[6]. There are a lot of assumptions and caveats here, and many decisions and supporting evidence underlie the algorithm design. Reporting everything would be unwieldy, so we now point at O'Dell et al. (2012), from which interested readers can reconstruct the full methodology.

The new phrasing is:

"The forward RT simulations used to generate the LUTs are performed with the ReFRACtor RT code, which implements the methodology described in Section 2.2.4 of O'Dell et al. (2012). Of particular relevance for cloudy scenes, multiple-scattering is calculated using LIDORT with a polarisation correction for low orders of scattering (Natraj and Spurr, 2007; Spurr, 2006). This assumes a plane-parallel atmosphere with a correction to the direct beam to account for Earth's sphericity."

We have endeavoured to include all relevant information while maintaining brevity.

1. **Spurr, R. & Christi, M. The LIDORT and VLIDORT Linearized Scalar and Vector Discrete Ordinate Radiative Transfer Models: Updates in the Last 10 Years. in *Springer Seeries in Light***

Scattering **(ed. Kokhanovsky, A. A.) 1–62 (Springer Nature, 2019). doi:10.1007/978-3-030-03445-0_1**

2. **Nakajima, T. & Tanaka, M. Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation. *J. Quant. Spectrosc. Radiat. Transf.* 40, 51–69 (1988).**
3. **Natraj, V. & Spurr, R. J. D. A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering–absorbing media. *J. Quant. Spectrosc. Radiat. Transf.* 107, 263–293 (2007).**
4. **O'Dell, C. W. Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation. *J. Geophys. Res.* 115, D10206 (2010).**
5. **Heidinger, A. K., O'Dell, C., Bennartz, R. & Greenwald, T. The Successive-Order-of-Interaction Radiative Transfer Model. Part I: Model Development. *J. Appl. Meteorol. Climatol.* 45, 1388–1402 (2006).**
6. **O'dell, C. W. *et al.* The ACOS CO 2 retrieval algorithm – Part 1: Description and validation against synthetic observations. *Atmos. Meas. Tech.* 5, 99–121 (2012).**

How do you determine the most meaningful wavelengths?

We think this is still referring to the text following L156 and specifically "Angular output is calculated for a handful of wavelengths". We have expanded on the description:

"Angular output is calculated with 8 streams for predefined bins in gas optical depth while single stream calculations are done for preselected wavenumbers at a mean separation $\Delta\nu$ ~0.04 cm^{-1}, with smaller separation within absorption bands. The high- and low-stream outputs are combined using O'Dell (2010)'s low-stream interpolation to rapidly and accurately reproduce high-stream output at all wavenumbers. These are then interpolated onto a uniform $\Delta\nu$=0.01 cm^{-1} grid and convolved with the instrument line shapes (ILS) to obtain channel radiances."

We used a standard configuration provided by the OCO-2 team based on their optimised sampling. We had wanted to keep all discussion in wavelength rather than wavenumber, but the code implements this part in wavenumber so we drop consistency for precision, i.e. "uniformly at $\Delta\nu$=0.01 cm^{-1}" over "close-to-uniformly at $\Delta\lambda$~0.00059 nm".

Reviewer 2 response

Firstly, thanks for having read this paper so attentively and providing a review that both catches minor mistakes and is thought provoking with regards to the main science and context. We genuinely appreciate your efforts and respond to each point in detail below. Our text continues in bold, we leave yours unbolded, and we insert the updated figures at the end of this document.

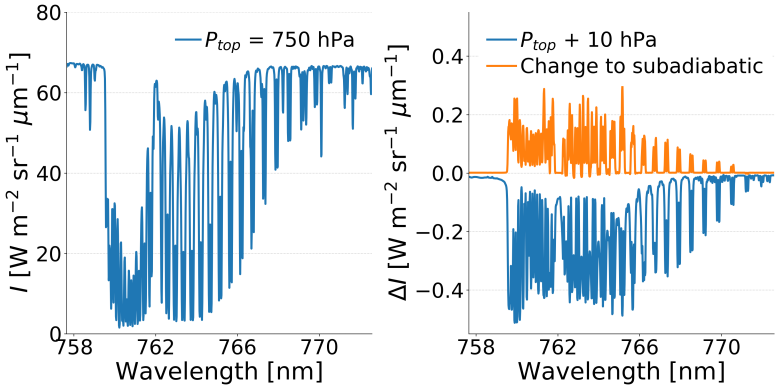
"A new OCO-2 cloud flagging and rapid retrieval of marine boundary layer cloud properties" by **Mark Richardson[1,2], Matthew D. Lebsock[1], James McDuffie[1], Graeme L. Stephens**

The manuscript documents a new cloud flagging and retrieval pre-processor for OCO-2 marine boundary liquid cloud retrieval algorithm. This new pre-processor contains two steps: step one uses a machine learning perceptron network to classify low liquid cloud, and step two that simultaneously retrieves optical thickness (t), effective radius and cloud top pressure (Ptot) with a three-parameter LUT using CO2, and O2 continnum radiances and A-band absorption to continuum ratio. The new retrievals and throughput are compared with MODIS and CLIPSO retrievals and original CLD-LIDAR-AUX as needed. In general, the new classifier/retrieval without CLIPSO show promising results. The residual biases in optical thickness, effective radius and cloud top pressure are discussed that could be potentially linked to spatial variability and in-cloud vertical structure. I would like to see how changes in pre-processor retrieval of tau, Re and Ptot affect the eventual OE retrieval beyond the impact of qualifying throughput. But that could be another major undertake, if the authors are planning to change the cloud model within the OE retrieval system and that could be content of the next paper.

We are also excited to see the updated OE retrieval but want to do a comprehensive job. We first need to derive a correction for the cloud vertical structure bias we mention in this submission. After that we have to (i) recalculate the S_y covariance matrix accounting for vertical structure and given our new, smaller r_e uncertainty and (ii) repeat the channel selection algorithm of Richardson & Stephens (2018) to determine whether to continue using the 75 channels selected for OCO2CLD-LIDAR-AUX. We judge this to be too much content to add to this paper.

We expect the vertical structure issue to be important enough that doing more retrievals without accounting for it wouldn't be a good time investment.

The figure below shows (left) τ=10 cloud simulated *I* and (right) changes in *I* when properties are the same but the vertical extinction structure is converted to subadiabatic (orange) or when the cloud is lowered by 10 hPa (blue). 10 hPa is multiple times our idealised OE posterior P_{top} uncertainties (Richardson & Stephens, 2018, again), and the vertical structure effect on radiances is similar in magnitude to that, so we want to finish development of our correction method.



The paper is well written and the addition of Table 1 is very helpful. I have a few more minor comments as follows:

120 P1L25-33. The description of the first paragraph is a bit misleading as if A-band alone allows all the retrieval, i.e., Tau, Ptop, H and Nd. But A-band only helps with cloud flag and cloud top height, other channels are needed to retrieval Tau, Re, and H.

Agreed, leaving the r_e issue to the end of the paragraph makes it easy to misinterpret. We removed the last sentence and added:

125 **", provided coincident information about effective radius (r_e) from other channels."**

P3L83. Why do you limit to nadir only orbits?

We have added the following justification:

"…we use nadir only orbits to provide complementary vertical information on clouds that are too low or thin to be adequately profiled by CloudSat's nadir-view radar. Glint-view footprints would
130 **preclude our use of the nadir-only CALIPSO lidar data and atmospheric photon path lengths would be longer, thereby reducing the retrieval sensitivity. Given our retrieval's computational expense we limit to nadir orbits to optimise the likelihood of good retrievals."**

Additional reasons include how this is largely a CloudSat project; the original CloudSat proposal included an A-band spectrometer for this purpose. We hope to use the ~5 years of coincident
135 **measurements by combining CloudSat's exquisite sensitivity to precipitation with an OCO-2 N_d product plus ancillary information to do aerosol-cloud-precipitation science.**

That said it would be great if we (or someone) had the time & resources to explore the potential to use the glint data.

Figure 3. A-band ratio will be less sensitive for low clouds. The 403 hPa cloud shown is way above the boundary clouds. More relevant to the retrieval of low clouds is whether you are able to separate clouds from 827 hPa to something like 750 km.

We assume this refers to Figure 1 and have changed to footprints containing clouds at 943 hPa and 740 hPa. This is a good suggestion to keep everything consistently focussed on low clouds.

We argue against showing smaller P_{top} gaps than this: this figure's purpose is to help readers to visualise spectrum differences due to changing photon path length. The ~200 hPa gap makes this visible. The rest of our analysis and past work shows sensitivities and error statistics. The new figure 1 and caption are at the end of this document.

P7L221: Are you sure the middle parameter is Ic,st/Ic,O2, not Ic,st? In Figure 4a, LUT shows Ic,o2 and Ic,st as controlling parameters.

We are absolutely not sure. That was a typo and has been corrected, thanks for your detailed reading.

This was a leftover from an earlier version when we used that ratio: it "stretches out" the vertical component of the Nakajima-King table so that instead of looking like a droopy slice of pizza it looks more like a wavy pizza al taglio.

The original interpolation code performed better with this approach at low τ, but updates mostly removed the performance advantage so we decided to stick to the standard droopy pizza slice to avoid having to describe & justify an alternative.

Table 1. It is better to spell out that the input to machine learning classifier includes both radiances and radiance ratios.

Agreed, text changed:

"…based on OCO-2 radiances and radiance ratios"

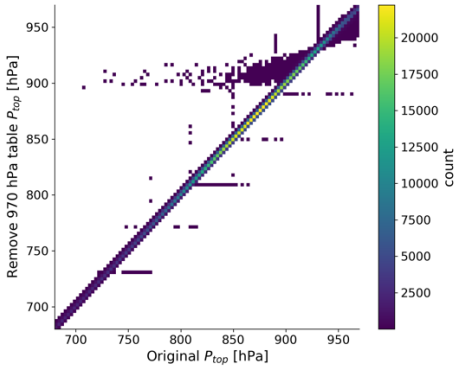
P7L220. Not clear if you process all pixels with LUT, or only pixels that identified as cloud from classifier to further process with LUT. If so, please mention here.

Added another step:

"(ii) Apply the classifier to identify appropriate cloudy footprints, pass only these to the next step"

P9L280. Could the asymmetric bias of Ptot be related to truncation of LUT, i.e., the same A- band ratio could occur in higher cloud but smaller optical depth? How will the retrieval change if you have LUT that covers more pressure levels, i.e., from 500hPa to 1010 hPa?

We did not think about this but have now tested. Extending the LUT to 1,010 hPa would put cloud top below our surface pressure. We'd need to generate full new LUTs as well as reprocessing the retrievals. Our alternative is to cut down the current LUTs: we tried removing the 970 hPa entries, then removing the 650 hPa entries, then removing both. The retrievals were then reprocessed. Note that we can still retrieve lower/higher values because our interpolator can "fill in" P_{top} values in some cases. Here's a 2d histogram of the retrieved P_{top} when the 970 hPa inputs are removed, versus our main results:



We first address the horizontal and vertical striping: it's where LUT P_{top} values are defined. We couldn't identify the exact reason. It could be due the Nakajima-King table "point" at low tau/reff where CO_2-band radiances span a small range and could present interpolation problems. These "stripes" contain optically thin clouds (median τ=3.5 versus population τ=9.7) with small droplets (median r_e=5.8 μm versus population r_e=12.0 μm), i.e. they are in the point of the N-K pizza slice. This is <0.02 % of the sample so we don't consider it to be worth further development at this point.

Back to the rest of that 2d histogram: the top right cluster shows disagreement where LUT information has been removed. The interpolation method still generates some retrievals there, and gives high or low biases depending on the original P_{top}.

This suggests to us that the limitations placed on the LUT P_{top} values are not related to our P_{top} bias relative to CALIPSO. Referring to Figure 5(c), the large biases occur for P_{top} values well within the table rather than at the edges as we'd expect if it were related to our truncated table.

Here are the median and 10—90 % ranges of retrieval differences when we restrict the input LUT P_{top} values: the median bias ranges from -0.8 hPa to 0.1 hPa, versus our ~12 hPa median bias.

LUT Ptops removed	Median [10%,90%] new minus full table P_{top}
970 hPa	0.07 [-1.42,2.29]
650 hPa	-0.78 [-3.98,6.84]
650, 970 hPa	0.00 [-0.08,0.09]

Our LUT intentionally extends to 650 hPa, i.e. beyond the ISCCP "low cloud" threshold of 680 hPa, which was an attempt to minimise truncation issues. This was done even though we intend to focus on the lower altitude clouds. Given the apparently small effect we have not made any changes to the paper in response to this comment.

Figure 7. I would rename the "passed", "failed LUT","old but not new" into "passed both","passed classifier","failed both".

This is nicer and clearer. However, since we don't pass the non-classified cases to the LUT [now clarified in response to P7L220 comment] it feels a bit weird to say the final group "failed both". We reproduced the figure using "pass both", "passed classifier", "failed classifier" and changed the legend. We also increased label font size a bit. The figure and its caption are at the end of this document.

P13L380 Do you mean "inhomogeneity"?

Another typo, changed.

P13L384. You have previously scaled the Ic,st channel by 0.9804. Is this scaling sufficient? If you scale by a smaller value, the mean bias of Re might be reduced.

This is true and it's unfair to apparently "accuse" calibration when it could be due to one of our choices. We had thought about this but attentive readers should definitely question this, so now we provide our justification by replacing that sentence with:

"We also ran the LUT retrieval with a -1 % shift in $I_{c,st}$, which shifts median r_e by +0.2 μm. Such a radiance shift could be necessary due to errors in calibration or in our derived scaling factor of 0.9807, which we used to relate the L1bSc file $I_{c,st}$ to our lookup table channels. We could therefore reduce our r_e bias by further scaling the $I_{c,st}$ radiances, but the scaling was derived from directly comparing the channel radiances rather than as a *post hoc* correction to improve retrieval results. If the r_e bias is due to other factors then this *post hoc* correction could result in compensating errors which hide other flaws in the retrieval."

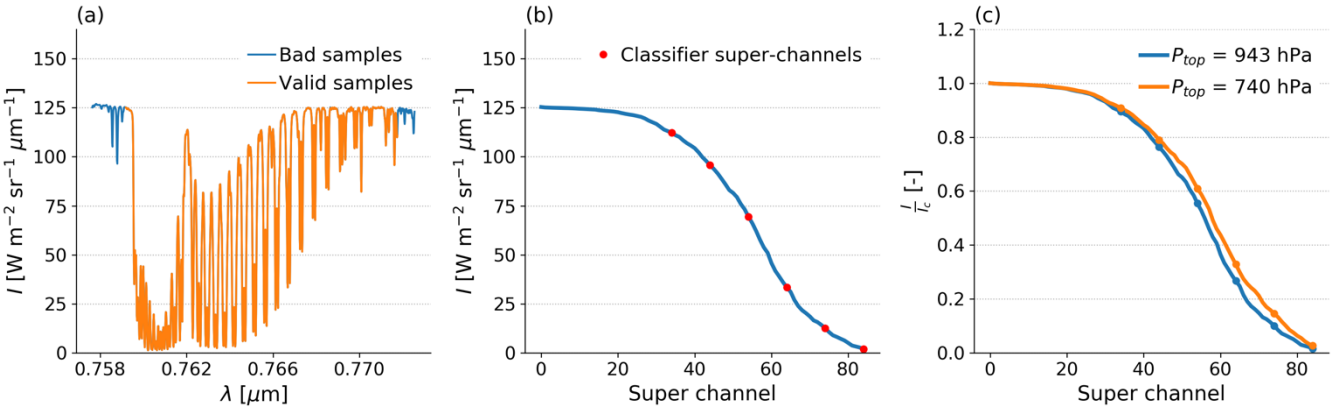


Figure 1.(a) OCO-2 A-band spectrum for a cloud with CALIPSO P_{top} = 943 hPa, with the used channels in orange and non-used (e.g. due to FPA pixel damage) in blue, (b) the smoothed super-channel spectrum, where the channels are ranked in brightness and then non-overlapping 10-channel means are taken. The super channels used in the classifier are shown in red. (c) Comparison of the ranked ratio I/I_c between this cloud, and one at higher altitude (P_{top} = 740 hPa). The SZA is within 0.01° and the continuum radiance within 1 % between each spectrum, the differences are largely due to the shorter path length resulting in less absorption for the higher altitude cloud.

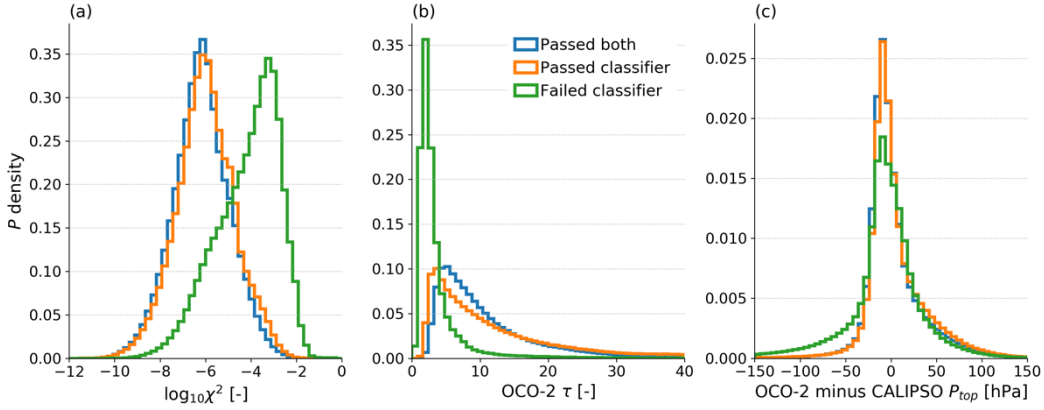


Figure 2. Normalised histograms of OCO2CLD-LIDAR-AUX outputs where SZA < 45° separated into whether the soundings pass the new pre-processor flag and retrieval or not. The "passed both" set are those that returned valid cloud properties from the LUT along with T_{top} > 0 °C, the "passed classifier" case gave invalid cloud values or had T_{top} < 0 °C, and the "failed classifier" set are those that were attempted in OCO2CLD-LIDAR-AUX but are not passed by the new classifier. (a) Logarithm of χ^2, (b) retrieved τ, (c) P_{top} minus the closest CALIPSO retrieved P_{top}.

A new OCO-2 cloud flagging and rapid retrieval of marine boundary layer cloud properties

Mark Richardson[1,2], Matthew D. Lebsock[1], James McDuffie[1], Graeme L. Stephens[1,2,3]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[2]Colorado State University, Fort Collins, CO 90095, USA
[3]Department of Meteorology, University of Reading, RG6 7BE, UK

Correspondence to: Mark Richardson (markr@jpl.nasa.gov)

Abstract. The Orbiting Carbon Observatory-2 (OCO-2) carries a hyperspectral A-band sensor that can obtain information about cloud geometric thickness (H). The OCO2CLD-LIDAR-AUX product retrieved H with the aid of collocated CALIPSO lidar data to identify suitable clouds and provide *a priori* cloud-top pressure (P_{top}). This collocation is no longer possible since CALIPSO's coordination flying with OCO-2 has ended, so here we introduce a new cloud flagging and *a priori* assignment using only OCO-2 data, restricted to ocean footprints where solar zenith angle < 45°. Firstly, a multi-layer perceptron network was trained to identify liquid clouds over ocean with sufficient optical depth ($\tau > 1$) for a valid retrieval, and agreement with MODIS-CALIPSO is 90.0 %. Secondly, we developed a lookup table to simultaneously retrieve cloud τ, effective radius (r_e) and P_{top} from A-band and CO_2 band radiances, with the intention that these will act as the *a priori* in a future retrieval. Median P_{top} difference versus CALIPSO is 12 hPa with interdecile range [-11,87] hPa, substantially better than the MODIS-CALIPSO [-83,81] hPa. The MODIS-OCO-2 τ difference is 0.8 (-3.8,6.9) and r_e is -0.3 [-2.8,2.1] μm. The τ difference is due to optically thick and horizontally heterogeneous cloud scenes. As well as an improved passive P_{top} retrieval, this *a priori* information will allow a purely OCO-2 based Bayesian retrieval of cloud droplet number concentration (N_d). Finally, our cloud flagging procedure may also be useful for future partial column above-cloud CO_2 abundance retrievals.

1. Introduction

Hyperspectral O_2 A-band measurements near $\lambda = 0.78$ μm, such as those taken by the Orbiting Carbon Observatory-2 (OCO-2), may provide unique new information about boundary layer clouds by retrieving their geometric thickness (H) or droplet number concentration (N_d), provided coincident information about effective radius (r_e) from other channels. They are able to do this because the spectrum responds to the photon path length between the Sun, Earth and the sensor. Increased H or decreased N_d with all other cloud properties held constant leads to increased distance between within-cloud scattering events, and therefore a longer photon path length and decreased transmittance in wavelengths where O_2 absorbs. This leads to

spectrally varying changes in observed A-band spectra that can allow joint retrievals of cloud optical depth (τ), cloud top pressure (P_{top}) and H, provided there is sufficient spectral resolution and low enough noise (O'Brien and Mitchell, 1992; Richardson and Stephens, 2018). ~~Equivalently, with knowledge of a relevant effective radius (r_e), the N_d could in principle be retrieved along with τ and P_{top}.~~

The basic principle of A-band absorption for cloud height is well established (Fischer and Grassl, 1991; Rozanov and Kokhanovsky, 2004; Yamamoto and Wark, 1961) and numerous spaceborne A-band instruments retrieve cloud properties (Koelemeijer et al., 2001; Kokhanovsky et al., 2005; Lindstrot et al., 2006; Loyola et al., 2018; Preusker et al., 2007; Vanbauce et al., 1998), but most lack the spectral resolution or noise characteristics to obtain H (e.g. Schuessler et al. (2014)). Others rely on multi-angle (Ferlay et al., 2010) or combined A- and B-band information (Yang et al., 2013), although these tend to contain little information on low-altitude and relatively thin clouds like marine stratocumulus (Davis et al., 2018; Merlin et al., 2016).

An OCO-2 based retrieval of τ, P_{top} and H has been developed (OCO2CLD-LIDAR-AUX, available at www.cloudsat.cira.colostate.edu/data-products/level-aux/oco2cld-lidar-aux), which uses lidar-based retrievals from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite to help identify cloudy scenes and constrain prior P_{top} (Richardson et al., 2019). This retrieval is targeted at single-layer liquid clouds over the ocean whose response, both to warming and aerosols, are a major source of uncertainty in climate simulations (e.g. Bony et al. (2005); Bodas-Salcedo et al., (2019); Zelinka et al. (2020)). Independent information about cloud structure may help to address timely questions where other sensors which rely on different retrieval approaches and assumptions can lead to apparently contradictory conclusions (Rosenfeld et al., 2019; Toll et al., 2019).

With CALIPSO leaving the A-Train constellation in 2018, collocation between OCO-2 and CALIPSO footprints is no longer possible. Our future retrievals require a new cloud flagging method plus *a priori* cloud top information for our iterative Bayesian optimal estimation (OE) retrieval (Rodgers, 2000). This paper describes a new pre-processor for OCO-2 based liquid cloud property retrievals that provides the requisite cloud flagging and *a priori* information. Details of OCO2CLD-LIDAR-AUX are summarised in Table 1, which also lists the main changes introduced in this study.

We do not use the published OCO-2 cloud flag as it was not developed for ocean nadir scenes (Taylor et al., 2016), since they were considered too dark for OCO-2's main mission of column CO_2 (XCO2) retrievals (Crisp, 2008; Crisp et al., 2004; Eldering et al., 2016). Therefore we train a multi-layer perceptron network to rapidly identify liquid cloud scenes using collocated CALIPSO and Moderate Resolution Imaging Spectroradiometer (MODIS) retrievals. For the prior cloud property retrieval we develop lookup tables (LUTs) that jointly retrieve τ, r_e and P_{top} using OCO-2 O_2 A-band and strong CO_2 band (λ ~ 2.06 μm) radiances. These are similar to the Nakajima-King tables used in MODIS cloud retrievals (Nakajima and King, 1990) but add an A-band absorption ratio that is sensitive to P_{top}.

Our OCO-2 OE retrievals are computationally expensive due to the complex radiative transfer (RT), so we aim to avoid footprints which are unlikely to yield good retrievals. The cloud flagging and prior LUT retrieval developed here are a necessary step in excluding these footprints, and we further exclude those where solar zenith angle, SZA > 45° based on

OCO2CLD-LIDAR-AUX's retrieval statistics. It is possible that a future partial-column (i.e. above cloud) XCO2 retrieval could be developed, which would likely be targeted at columns above optically thick clouds, so the pre-processor developed here could find wider use (Schepers et al., 2016; Vidot et al., 2009). A further development is that our past retrievals used a fixed r_e and the addition of varying r_e is eased by a new Python RT interface using the ReFRACtor (Reusable Framework for the Retrieval of Atmospheric Composition) software described in Section 2.3. Our new LUT retrieval of a prior r_e will allow a more appropriate r_e to be assumed in the iterative OE.

The paper is organised as follows: Section 2 describes the relevant OCO-2 details, data selection and radiative transfer calculations before detailing the methodology. Section 3 reports the performance statistics of the classifier, compares LUT retrieved cloud properties versus MODIS and CALIPSO where the instrument footprints overlap, and compares the final pre-processor throughput against that of OCO2CLD-LIDAR-AUX. Section 4 discusses and contextualises the results as well as proposing actionable future work that could address identified biases and Section 5 concludes.

2. Methods and data

2.1. Instruments and data selection

The OCO-2 measurement approach and instrumentation are detailed in Bösch et al. (2017), the L2FP RT's application to clouds in Richardson et al. (2017), and the MODIS-CALIPSO-OCO-2 matchup data are as used in Taylor et al. (2016). The datasets used here are listed in Table 2, in particular, from the OCO-2 Level 1b Science (L1bSc) data we obtain calibrated radiances and RT inputs such as solar zenith angle (SZA) and instrument characteristics.

The OCO-2 satellite flies in the Sun-synchronous A-Train constellation (L'Ecuyer and Jiang, 2010), and measures during the daytime ascending node with an equator crossing time near 1:30 pm. Its orbits are committed primarily to either glint or nadir view, and we use nadir only orbits to provide complementary vertical information on clouds that are too low or thin to be adequately profiled by CloudSat's nadir-view radar. Glint-view footprints would preclude our use of the nadir-only CALIPSO lidar data and atmospheric photon path lengths would be longer, thereby reducing the retrieval sensitivity. Given our retrieval's computational expense we limit to nadir orbits to optimise the likelihood of good retrievals.

OCO-2~~. It~~ carries three co-boresighted grating spectrometers centred over the O_2 A-band (λ~0.78 μm), weak CO_2 band (λ~1.68 μm) and strong CO_2 band (λ~2.06 μm).

The satellite operates in a pushbroom fashion with a swath of 8 footprints whose orientation relative to the track rotates through the orbit as the satellite angles to optimise solar power generation. The subsequent parallelogram-like footprints are nominally near 1.4 km×2.2 km at nadir. The channels' wavelengths vary across the track due to the manner in which the optics focus light onto the focal plane array (FPA), and wavelength also drifts throughout an orbit due to Doppler shift. This causes issues for a LUT developed from a fixed set of channels, since the wavelengths sampled by those channels will differ between each measurement. Furthermore, some sensor pixels are damaged and we only include channel indexes where all 8

swath soundings are classed as good, which reduces the A-band sample from 1,016 to 853 channels. Section 2.2 describes how we use a channel averaging approach to reduce the consequences of this wavelength shift in the cloud classifier and Section 2.4 our related channel selection for the LUT.

For classifier training and validation, we require spatial overlap between OCO-2, MODIS and CALIPSO data. The ascending OCO-2 ground track is approximately 200 km to the east of Aqua's and therefore within the MODIS swath, so we select the 1 km MODIS retrieval footprint whose centre is closest at the surface to the centre of the OCO-2 footprint. However, CALIPSO only measures once at nadir so only one OCO-2 footprint in each swath can be collocated. Furthermore, even during formation flying the satellites drifted within their control boxes and some CALIPSO measurements occurred outside the OCO-2 swath. We only include footprints with a CALIPSO-OCO-2 matchup distance of <1.5 km at the surface. Finally, the dataset was further restricted to footprints with surface type = water, SZA < 45° and with valid radiances. Between 2014-09-06 and 2018-04-30 the MODIS-CALIPSO-OCO-2 matchup dataset has 5,909 nadir orbits of which 4,743 contain valid matchups. This is reduced to N = 3,907 orbits through 2016-12-31 when we also require an OCO2CLD-LIDAR-AUX retrieval.

2.2. Cloud Classifier Data Selection and Training

For the first step of rapidly identifying footprints that contain liquid clouds over the ocean we select a machine learning classifier which is trained on a set of collocated MODIS-CALIPSO footprints before being validated against an independent set of MODIS-CALIPSO data. The footprints which pass this classifier will be forwarded to the LUT estimator to generate the *a priori* cloud property estimate.

We generated independent sets of training (N=100,000) and validation (N=250,000) footprints by randomly selecting orbits and taking all their valid footprints until we had those sample sizes. We assign a cloud flag value of 1 to a footprint when the following conditions are all met, else it is 0:

(i) CALIPSO Feature_Classification_Flag = 2 (cloud present and is liquid)

(ii) CALIPSO retrieves a single layer

(iii) MODIS Cloud_Optical_Thickness > 1 (cloud present and is sufficiently optically thick)

As input we take the continuum radiances (I_c) from all 3 OCO-2 bands and correct for illumination geometry via $\mu_0^{-1} I_c$ where μ_0=cos(SZA), plus a number of A-band ratios described below. From Python's sklearn package we selected a multi-layer perceptron network (sklearn.neural_networks.MLPClassifier) with hidden layer sizes of (100,50,25). These selections are justified in Supplementary §1.

In these bands the ocean is dark and reflectance increases monotonically with τ, so the $\mu_0^{-1} I_c$ help to identify optically thick clouds. Ice also absorbs more strongly than water in the higher wavelength bands, which aids phase discrimination.

We calculate A-band absorption ratios by dividing a non-continuum (i.e. absorbing) channel radiance $I_{abs,O2}$ by $I_{c,O2}$. Clouds tend to increase $I_{abs,O2}/I_{c,O2}$ since photons scattered from the clouds encounter fewer O_2 molecules than those that travel all the way to the surface. This principle has been exploited to improve detection of clouds over bright snow & ice surfaces with

the A- and B-band channels of the Earth Polychromatic Imaging Camera (EPIC) on board the Deep Space Climate Observatory (DISCOVR) (Zhou et al., 2020). Consider:

$$\frac{I_{abs,O2}}{I_{c,O2}} = \exp\left(-\int k_{O2}(z)\, dz\right), \quad (1)$$

Where $k_{O2}(z)$ is the O_2 absorption coefficient, which is then integrated over the photon path $\int dz$. Considering a δ-function distribution of photon path lengths along the beam that is scattered from a single layer with constant $k_{O2}(z) = k_{O2}$, then at nadir the path can be decomposed into the path from TOA to the layer top, $\mu_0^{-1}\Delta z$, and from the layer top to TOA Δz:

$$\frac{I_{abs,O2}}{I_{c,O2}} = \exp(-k_{O2}(\mu_0^{-1} + 1)\Delta z), \quad (2)$$

making $k\Delta z$ the subject:

$$k\Delta z = -\ln\left(\frac{I_{abs,O2}}{I_{c,O2}}\right)(\mu_0^{-1} + 1)^{-1}, \quad (3)$$

We take the right-hand side of Eq. (3) as our observable. If we select channel combinations with near-constant k_{O2}, then the observable is proportional to Δz. Lower values should be associated with high (i.e. more likely ice clouds), and high values with clear scenes. This assumes similar scattering properties for the $I_{abs,O2}$ and $I_{c,O2}$, which is justified by the A-band's small wavelength range.

The k_{O2} sampled by individual channels varies for three main reasons:

(i) The central wavelength of each channel depends on the cross-track position due to the way in which the optics focus light on the FPA,

(ii) The wavelengths sampled change due to Doppler shift induces by relative Earth-satellite and Earth-Sun motion,

(iii) The strength of O_2 absorption varies due to line broadening induced by atmospheric conditions.

We use a method from Richardson et al. (2017) to address these factors. The 853 undamaged channels are ranked from brightest to darkest and a non-overlapping 10-channel mean is taken, resulting in 85 full "super channels". These are combined with $I_{c,O2}$ and μ_0 using Equation (3), and we selected every 10th super channel from the 35th onwards (Supplementary §1 shows little improvement from additional super channels).

This is illustrated in Figure 3, with Figure 3(a) showing an example cloudy spectrum and the damaged channels, Figure 3(b) the ranked super channels and those used in the classifier, and Figure 3(c) compares $I_{abs,O2}/I_{c,O2}$ for the original spectrum (CALIPSO P_{top} = 827 hPa) and for a spectrum with similar μ_0 and $I_{c,O2}$ but with CALIPSO P_{top} = 403 hPa. The brightest super channels show little response to scattering layer altitude, so they contain little information and they are excluded from the classifier. The higher altitude cloud has brighter $I_{abs,O2}$ due to the shorter mean path length. As stated previously, this aids in the phase classification, and also to discriminating between cloudy and clear scenes since very low $I_{abs,O2}/I_{c,O2}$ is more likely associated with photons scattered from the surface.

2.3. Radiative Transfer Simulations and ReFRACtor Interface

The forward RT simulations used to generate the LUTs are performed with the ReFRACtor RT code, which implements the methodology described in Section 2.2.4 of O'Dell et al. (2012). Of particular relevance for cloudy scenes, multiple-scattering is calculated using ~~which is based on~~ LIDORT with a polarisation correction for low orders of scattering (Natraj and Spurr, 2007; Spurr, 2006). This ~~code makes a semi-infinit~~assumes a ~~e~~ plane-parallel ~~assumption~~atmosphere~~,~~ with a correction to the direct beam to account for Earth's sphericity. Angular output is calculated with 8 streams for predefined bins in gas optical depth while single stream calculations are done for preselected wavenumbers at a mean separation $\Delta\nu$ ~0.04 cm^{-1}, with smaller separation within absorption bands. The high- and low-stream outputs are combined using O'Dell (2010)'s low-stream interpolation to rapidly and accurately reproduce high-stream output at all wavenumbers. These are then interpolated onto a uniform $\Delta\nu$=0.01 cm^{-1} grid and convolved with the instrument line shapes (ILS) to obtain channel radiances. ~~Angular output is calculated for a handful of wavelengths with 8 streams, with the rest of the spectrum interpolated using a single stream using the method of O'Dell (2010), which reliably reproduces the higher-stream output.~~ The selected~~is~~ numbers of streams ~~was~~ were found to reproduce cloudy scene radiances given MODIS and CALIPSO cloud properties (Richardson et al., 2017) and also matches the selection in Vidot et al. (2009).

OCO2CLD-LIDAR-AUX used OCO-2 L2FP RT and required input L1bSc and meteorology files plus a file containing pressure level and cloud information. Each footprint's output was saved to a file for every OE iteration, adding to a read-write bottleneck. Further inefficiency arose as if any footprint in an orbit included a type of scatterer (e.g. water clouds with r_e = 10 μm, which we term wc_010), its scattering properties had to be assigned for every profile in the orbit. For example, if one footprint contained a wc_010 cloud, every other footprint in the orbit that didn't contain a wc_010 cloud would need an assigned wc_010 profile with extinction = 0.

Here we use ~~a~~ the new ReFRACtor, which handles footprints as individual objects. Inputs are assigned uniquely to that object and it stores the RT output and updated properties internally, so no external reading or writing is required for intermediate OE iterations.

For LUT input we take an ocean footprint near 25 °S from the L1bSc file for orbit 16094a on 2017-07-11 for instrument and satellite properties, although we manually vary SZA. We used the mean OCO-2 cloudy profiles for tropical (20 °S—20 °N) footprints from Richardson and Stephens (2018). The high latitude case is excluded as its surface temperature is near 0 °C, so will mostly represent ice and mixed-phase clouds, and using the midlatitude (20—50 °S/N) case had little effect on the retrieval statistics.

The RT code takes input on levels and then linearly interpolates to generate vertically homogeneous layers. We use 16 pressure levels: 3 assigned linearly in P from TOA to 500 hPa, then 10 from 500 hPa to P_{top}, 2 from P_{top} to cloud bottom (P_{bottom}), and the final level is the surface. This was found to reliably reproduce OCO-2 L2FP RT standard outputs which use 20 levels, but with faster processing.

The cloud extinction is assigned to the level at the cloud centre, whose neighbouring levels are at P_{top} and P_{bottom} and the layer interpolation results in a vertically homogeneous cloud with constant $\tau(z)$ and $r_e(z)$. Rozanov and Kokhanovsky (2004) showed that a vertically uniform assumption may introduce radiation biases, relative to our target marine boundary layer clouds which tend to be vertically stratified with increasing extinction towards cloud top (Bennartz, 2007; Grosvenor et al., 2018; Painemal and Zuidema, 2011), but quantifying such a bias requires extensive testing that we intend to perform separately. For now, the cloud H is calculated as in Szczodrak et al. (2001) where $H \propto \sqrt{\tau r_e}$, and is converted to ΔP by assuming $\Delta z/\Delta P \approx 10$ m hPa^{-1}. Where this would result in $P_{bottom} > P_{surf}$, the cloud is compressed while maintaining the same P_{top}.

For surface reflection ReFRACtor does not currently allow for a Cox-Munk surface, so we assume Lambertian with albedo that varies by band and SZA. The band- and SZA-dependent values are derived from a set of OCO-2 radiances as described in Supplementary §2 and range from 0.010—0.054.

Gaseous absorption is from the absorption coefficient (ABSCO) version 5.0 tables used in OCO-2's latest XCO2 retrieval, version 9. These tables account for line changes due to temperature, pressure and water vapour. Cloud properties are pre-calculated using Mie theory at integer micron values of r_e following an assumed Gamma droplet size distribution with width parameter $\gamma = 1/9$. This follows the standard OCO-2 XCO2 retrieval aerosol input file, but with an update to correct an error in water absorption in the CO_2 bands [Aronne Merrelli, pers. comm.].

2.4. Lookup table development and retrieval

The LUT is designed to produce prior cloud property estimates for our future OE retrieval, which specifically targets marine boundary layer clouds and aims to provide additional information about their H or N_d. We therefore limit the range of the LUT properties to cover the majority of these clouds, with properties τ from 1—50, r_e from 4—32 μm and P_{top} from 650—970 hPa, and SZA spans 20—45° inclusive (see Supplementary Table 3 for selected values). The simulated outputs are $I_{c,O2}$ in the O_2 A-band, $I_{c,st}$ in the strong CO_2 band and an A-band ratio $I_{abs,O2}/I_{c,O2}$.

We take the mean of 5 channels for each of $I_{c,O2}$ and $I_{c,st}$ and 10 channels for $I_{abs,O2}$, and fixed channel indices are required to consistently convert the RT simulated spectra into LUT radiances. The selected channels minimise the root mean squared error (RMSE) across a large sample of footprints against the L1bSc continua (for $I_{c,O2}$ and $I_{c,st}$) and the 60th super-channel for $I_{abs,O2}$ (as defined in Section 2.2). The 60th super channel is picked as it showed the greatest sensitivity to CALIPSO P_{top} in Richardson et al. (2017). The selection algorithm is described in Supplementary §4, the error statistics are in Supplementary Table 4 and the channel indices in Supplementary Table 5. The error statistics show that our selection is valid for a range of meteorological conditions, illumination geometries, Doppler shifts and for all 8 cross-track sounding positions.

The LUT channels are highlighted in Figure 4, which shows mean spectra from a large sample of cloudy footprints. The channel means with 2σ ranges are shown as shaded bands and are compared with the truth as solid lines. The truth for I_c is the mean of the sample L1bSc radiance continua, which represent the brightest channels in each footprint and whose channel

indices may change with footprint while the $I_{abs,O2}$ truth is the spectrum's 60th super channel. The estimators are consistent with the truth in each case, with the best agreement for $I_{c,O2}$ and a negative bias in $I_{c,st}$. We found that scaling the L1bSc $I_{c,st}$ value by 0.9804 resulted in similar error statistics to using our selected channels, so we use scaled L1bSc $I_{c,st}$ in our LUT retrieval since it those radiances are already loaded for the classifier. The individual $I_{abs,O2}$ channels show a large spread, but the channel selection algorithm accounts for anti-correlation in their radiances such that the 10-channel mean is consistent with the 60th super channel across all test footprints.

For each SZA, Refractor is run for all combinations of input cloud properties to generate A-band and strong CO_2 band radiances for these selected LUT channels. A LUT is generated at each 5° in SZA from 20—45° inclusive, and the retrieval works as follows:

(i) From an L1bSc file, load the SZA plus radiances to get $I_{abs,O2}$, $I_{c,O2}$ and $I_{c,st}$

~~(i)~~(ii) Apply the classifier to identify appropriate cloudy footprints, pass only these to the next step

~~(ii)~~(iii) Convert these into the LUT observables $I_{c,O2}$, $I_{c,st}$~~/$I_{c,O2}$~~, and $I_{abs,O2}/I_{c,O2}$

~~(iii)~~(iv) Scale observables onto the nearest LUT SZA using the appropriate μ-related scaling,

~~(iv)~~(v) Interpolate within the LUT to simultaneously estimate τ, r_e and P_{top}.

If the observed radiances are outside the LUT values then a NaN is returned and the footprint is flagged as not retrievable. The footprint is also flagged as likely to contain ice if L2Met $T(P_{top,retrieved}) < 0$ °C. We refer to NaN or $T_{top} < 0$ °C outputs as not being passed by the LUT, since these footprints will not be attempted in our future OE retrieval.

2.5. Pre-processor prior validation

The pre-processor is run on the 3,907 orbits used in OCO2CLD-LIDAR-AUX from September 2014—December 2016 where the new L1bSc and L2Met Version 8 files are available along with the collocated MODIS and CALIPSO files, and where there are any ocean footprints with SZA < 45°. For validation of the LUT we consider only those footprints where the CALIPSO matchup distance < 1.5 km as in Section 2.1, where the MODIS, CALIPSO and LUT retrievals are within the valid LUT property range and where derived OCO-2 $T_{top} > 0$ °C (N = 1,264,449). The primary analysis is in the pairwise differences between the LUT retrieved properties and MODIS τ or r_e, and CALIPSO P_{top}. The MODIS P_{top} is also evaluated against that of CALIPSO.

2.6. Comparison with OCO2CLD-LIDAR-AUX pre-processor

The OCO2CLD-LIDAR-AUX matchups are separated into three sets: those that are not flagged by the classifier, those that are flatted but do not pass the LUT retrieval (due to out-of-range cloud properties or implied $T_{top} < 0$ °C), and those that fully pass the pre-processor. Throughput and agreement are compared with the OCO2CLD-LIDAR-AUX cloud flag, retrieval χ^2, retrieved τ and P_{top} discrepancy versus CALIPSO. The pre-processor performs well if it successfully passes those footprints with small posterior χ^2 and P_{top} discrepancy while avoiding those with larger values.

The OCO2CLD-LIDAR-AUX cloud flag was based on simple thresholds in $\mu_0^{-1}I_{c,O2}$ and $\mu_0^{-1}I_{c,wk}$ combined with a phase discrimination based on their combination, and finally a requirement for valid CALIPSO single-layer clouds with $P_{top} > 680$ hPa occurring within 10 km. This flag did not have a SZA cutoff at 45°, so we will also specifically consider comparisons between the outputs of the two pre-processors where SZA < 45°.

3. Results

3.1. Cloud classifier test statistics

As in Section 2.2 the classifier output is 1 when we expect a single layer liquid cloud with $\tau > 1$ and 0 otherwise and the validation data, which we also term "truth", is the MODIS-CALIPSO classification. We use the following terms:

(i) True Positive (TP), classifier = 1, truth = 1

(ii) False Positive (FP), classifier = 1, truth = 0

(iii) False Negative (FN), classifier = 0, truth = 1

(iv) True Negative (TN), classifier = 0, truth = 0

Which are normalised such that TP + FP + FN + TN = 100 %. These can be summarised in a confusion matrix, as is done in Figure 5(a) for the N = 250,000 non-training sample. Its trace is the accuracy score of 90.0 % and the off-diagonal elements represent potential misclassifications. Figure 5(b) shows that the FNs are largely clouds of lower MODIS τ than those identified by the classifier, with 29.4 % of FNs having MODIS $\tau < 3$, compared with 7.3 % of TPs.

Some of these "missed" clouds may be due to collocation error, for example a cloud may average $\tau > 1$ over the 1 km MODIS footprint, bur not over the larger OCO-2 footprint. The classifier will also have errors: it maximises the accuracy score, and detecting lower τ clouds may require passing darker scenes which could increase the prevalence of FPs.

Figure 5(c) shows the distribution of CALIPSO T_{top} where retrieved, and shows far more cold-topped clouds in the FP case compared with the TPs, although there is also a $T_{top} < 0$ °C peak in the FN case. This suggests that the classifier misidentifies some ice clouds as liquid, and also that some of the FNs may in reality be mixed phase clouds that CALIPSO has nevertheless identified as liquid. For example, 24.6 % of FNs have $T_{top} < -10$ °C, compared with 7.6 % of the TP sample.

Among the false positives, we expected that there would be a larger occurrence of broken or multi-layered clouds, where thick broken clouds were sufficiently bright to trigger detection or where overlying thin ice clouds have too little effect on the radiances to be flagged as ice. We describe a scene as broken when the MODIS partially cloudy retrieval exists (Cloud_Optical_Thickness_PCL > 0) and a scene as multi-layered when CALIPSO retrieves more than 1 cloud layer, although strictly this can only detect multiple layers when the upper layer does not fully attenuate the lidar. While 11.3 % of the full sample is multi-layered, 40.1 % of the FP cases are, and while 12.2 % of scenes are partially cloudy, 30.4 % of FP footprints are. Overall, 69.4 % of FPs are associated with multi-layer or broken clouds, or both.

3.2. Lookup table matchup performance

Figure 6(a) shows simulated $I_{c,O2}$ and $I_{c,st}$ at SZA = 30° and P_{top} = 810 hPa for all τ and r_e while Figure 6(b) contains the median and spread of $I_{abs,O2}/I_{c,O2}$ at each fixed P_{top}. Most of the $I_{abs,O2}/I_{c,O2}$ variance is explained by P_{top}, with spread largely due to changes in within cloud scattering. For example, for P_{top} = 970 hPa, the optically thickest clouds were artificially compressed to prevent them from extending below the surface thereby reducing the in-cloud path and increasing the maximum $I_{abs,O2}/I_{c,O2}$.

The OCO-2 LUT retrievals are compared with those of MODIS and CALIPSO in Figure 7(a—c) and the MODIS and OCO-2 P_{top} differences relative to CALIPSO are in Figure 7(d). We consider the OCO-2 value minus the other product's value, and report median [10^{th}, 90^{th} percentiles] instead of standard deviation as these distributions are commonly non-Gaussian. There is good correlation between OCO-2 and other products, with a τ difference of 0.77 [-3.77,6.93], an r_e difference of -0.25 [-2.78,2.13] μm and a P_{top} difference of 12 [-11,87] hPa. As can be seen in Figure 7(d), the LUT P_{top} retrieval outperforms that of MODIS, whose difference relative to CALIPSO is -17 [-83,81] hPa, i.e. the OCO-2 interdecile range is approximately 40 % smaller than that of MODIS.

We also divide the τ and r_e differences by the MODIS reported uncertainty ($\sigma_{\tau,MODIS}$, $\sigma_{re,MODIS}$). If the OCO-2 and MODIS retrievals were independent Gaussian with equal variance then the standard deviation of OCO-2-MODIS differences would be $\sqrt{2} \approx 1.41\sigma_{MODIS}$. We find values of $1.26\sigma_{\tau,MODIS}$ and $0.37\sigma_{re,MODIS}$, indicating that the r_e retrievals are not independent and that our differences are within the MODIS-reported uncertainties.

We acknowledge discrepancies in the median retrieved τ and r_e, and refer to these as biases. The τ bias grows both with OCO-2 retrieved τ and with the horizontal variability of the scene as displayed in Figure 8. For this figure, the samples were split into deciles according to the LUT retrieved τ or the MODIS sub-pixel index at λ = 0.66 μm, which is the standard deviation of the 250 m footprint radiances with a 1 km cloud retrieval, divided by the mean of those radiances. Spatial variability and greater optical depths appear to drive much of the τ bias but we could not identify a dominant factor consistently correlated with the small r_e bias. These issues are further discussed in Section 4.2.

3.3. Pre-processor throughput

The multi-layer perceptron classifier passes 5.5 % of all OCO-2 footprints as $\tau > 1$ liquid clouds, of which 0.9 % return invalid cloud properties from the LUT and a further 0.8 % have implied $T_{top} < 0$ °C, resulting in a final throughput of 3.8 %. This is smaller than OCO2CLD-LIDAR-AUX, which attempted to retrieve 14.1 % of all soundings. However, most of the difference is due to SZA, and when we restrict the denominator to all footprints with SZA < 45° the throughputs are 13.1 % for OCO2CLD-LIDAR-AUX and 11.7 % for the new classifier, or 8.1 % after the LUT thresholds.

Figure 9 displays histograms of selected OCO2CLD-LIDAR-AUX outputs for SZA < 45° retrievals split into footprints where the new pre-processor passes the footprint (blue), where the LUT returns invalid properties or $T_{top} < 0$ °C (orange), or

the classifier does not pass the footprint (green). The new classifier identifies "better" retrievals that ended with smaller fit errors: median $\chi^2 = 7.2\times10^{-7}$, versus 1.3×10^{-4} for those not passed (among those with SZA < 45°). The LUT filtering further improves the statistics, with median $\chi^2 = 9.8\times10^{-7}$ for those not passed by the LUT retrieval versus $\chi^2 = 6.6\times10^{-7}$ for those successfully retrieved with T_{top} > 0 °C. The perceptron network also tends to pass clouds that are more optically thick (median τ = 8.6 vs. 2.4) and to show smaller spread in the difference between OCO-2 and CALIPSO P_{top} (standard deviation of differences, σ = 33 hPa vs. 55 hPa).

The OCO2CLD-LIDAR-AUX footprints that are excluded by the new pre-processor are consistent with optically thinner clouds and with poorer quality retrievals. Among the footprints that are passed by the new pre-processor, 17.1 % were not attempted in OCO2CLD-LIDAR-AUX.

4. Discussion

4.1. Cloud classifier and pre-processor throughput

The cloud classifier's agreement of 90.0 % with MODIS-CALIPSO is similar in performance to the original OCO-2 operational cloud flagging for ocean glint used in the L2FP XCO2 retrieval (Taylor et al., 2016). Furthermore, the multi-layer perceptron network is lightweight (size <250 kB) and fast. It throughputs 11—13 % of ocean soundings where SZA < 45°, of which under a quarter are poor retrieval candidates according to MODIS-CALIPSO. These cases are consistently (~69 %) broken or multi-layered cloud scenes, while the missed MODIS-CALIPSO cloud scenes are commonly optically thinner (~4 times likely to be τ < 3) and colder (~3 times likelier to have T_{top} < -10 °C) than the hit cloud scenes. These thinner and colder samples are also likely to be poor candidates for our target future retrieval of droplet number concentration in warm topped clouds.

Applying the LUT retrieval further reduces the number of footprints that are taken to be liquid clouds with τ > 1. The OCO2CLD-LIDAR-AUX retrieval attempted 13.1 % of SZA < 45° footprints, the new classifier-LUT pre-processor passes 8.1 %. Figure 9 showed that the excluded footprints tended to be more optically thin, have larger discrepancy in retrieved P_{top} relative to CALIPSO, and to have higher χ^2. This suggests that the new pre-processor will pass better retrieval candidates to the OE code, thereby improving efficiency. Of those that are now passed, 17 % were not passed by OCO2CLD-LIDAR-AUX. These likely include cases of mis-identification that will result in poor quality retrievals, but may also include true cloud cases that were not identified in OCO2CLD-LIDAR-AUX. For example, retrievals were previously classified using the nearest CALIPSO footprint up to 10 km away, and if a cloud was in the OCO-2 field of view but not the CALIPSO field of view, it would not previously have been passed. Overall, the new pre-processor shows good performance in terms of identifying scenes which likely contain liquid clouds with sufficient τ.

4.2. Lookup table cloud property retrieval

The LUT retrieval shows good correlation with MODIS τ and r_e plus CALIPSO P_{top} in Figure 7. Versus CALIPSO, the LUT-based P_{top} retrievals have a smaller-magnitude bias and 40 % smaller inter-decile range than MODIS. The 12 hPa P_{top} bias represents OCO-2 retrieved clouds that are lower in the atmosphere than retrieved by CALIPSO. These statistics may include cases of broken cloud, either above a lower cloud or above the surface. Three-dimensional (3d) cloud effects, or combined scattering from multiple cloud layers could lead to longer mean photon path lengths and thereby a larger OCO-2 P_{top}, assuming that CALIPSO tends to identify the highest layer. We consider full 3d radiative transfer treatments to be beyond the scope of this study but point readers to a wide literature on this topic (Davis and Knyazikhin, 2005; Heidinger and Stephens, 2002; Kokhanovsky et al., 2007; Várnai and Marshak, 2002).

Aerosol is ignored in these simulations, as previous analysis using CALIPSO aerosol products showed no change in OCO2CLD-LIDAR-AUX P_{top} bias in response to CALIPSO-identified aerosol (Richardson et al., 2019). Furthermore, above-cloud scattering aerosol would tend to reduce photon path length and therefore have an opposite effect on P_{top} to our observed bias.

Retrieved P_{top} could also change due to the assumed cloud vertical structure and meteorological profile used in the LUT development. If the cloud vertical structure used in the RT differs from reality, then this could lead to incorrectly simulated within-cloud photon paths. Firstly, if the simulated cloud is too geometrically thin (low H) for a given τ, r_e then the within-cloud path length will be too small and the above-cloud path must be lengthened to compensate, resulting in a positive P_{top} bias, and vice versa for too-high simulated H. This study improves on the OCO2CLD-LIDAR-AUX prior realistically varying H with r_e in addition to τ, but a bias may remain. In particular, shallow marine clouds tend to have extinction weighted towards the top which affects the exiting radiance and may introduce P_{top} biases which vary with geometry and cloud properties. We intend to perform a separate and more detailed analysis of how realistic vertical cloud profiles affect simulated OCO-2 radiances, and determine how to account for such a vertical-structure bias.

With regards to meteorology, a warmer and moister profile broadens the O_2 absorption lines and we expect stronger resultant absorption in the selected $I_{abs,O2}$ channels. Our tropical meteorology may lead to too-strong absorption in non-tropical scenes such that the retrieved cloud is lifted (i.e. lower P_{top}) to compensate, but the observed bias is opposite to this. We also retrieved using a LUT developed with the Richardson and Stephens (2018) midlatitude meteorology where surface temperature is approximately 10 °C cooler. The retrieved P_{top} distribution shifts as expected with median P_{top} bias increasing from 12 hPa to 15 hPa.

Overall the OCO-2 LUT gives better P_{top} retrieval statistics than MODIS for these shallow marine clouds, where for these clouds MODIS retrievals rely on brightness temperature at $\lambda \sim 11$ μm and so may mis-assign P_{top} when a temperature inversion is present (Baum et al., 2012). However, OCO-2 has a larger footprint, smaller swath and only retrieves during nadir view orbits. The P_{top} bias relative to CALIPSO is concerning for a future optimal estimation retrieval, since biased prior properties may subsequently bias the posterior retrieved state in unpredictable ways. We confidently exclude aerosol

and meteorology as the main factors in the observed bias, and propose that the main candidate processes are a combination of horizontal variability, OCO-2-CALIPSO collocation error and potentially vertical cloud structure. In the future OE retrieval we would expect horizontally non-uniform clouds to produce spectra that are more difficult to match under our RT assumptions, so such cases may be identified by the posterior χ^2 statistics. For vertical structure biases, we plan a detailed future investigation.

Retrieved τ and r_e show good correlation with those from MODIS and the variance of the differences is smaller than implied by the MODIS-reported uncertainties, if the LUT and MODIS uncertainties are independent Gaussian with MODIS' reported variance. However, OCO-2's $I_{c,O2}$ instrumental noise is lower than MODIS' (single channel signal-to-noise ratio, SNR of 300—1200, versus the MODIS band 4 specified SNR of 228), so the instrumental uncertainty contribution to the error budget should be smaller for OCO-2. There are also common characteristics between the retrievals, such as the use of fixed droplet size distribution variances, so individual footprint error will covary between the two. Such covariance should further reduce the inter-satellite difference in retrieved τ and r_e. A quantitative analysis would require a thorough calculation controlling for individual terms in the error budget, we simply conclude that there is no evidence of substantial unexpected variance in our retrieved τ and r_e.

Of greater concern is the residual OCO-2 minus MODIS differences of 0.77 in τ and to a lesser extent -0.25 μm in r_e. For τ the bias increases both with horizontal inhomogeneity and with τ, and we expect to be able to identify these clouds scenes using retrieved τ, and either OCO-2 developed metrics of spatial variability or future retrieval χ^2.

For the r_e bias we briefly assessed several factors. Horizontal variability tends to increase retrieved r_e (Werner et al., 2018; Zhang et al., 2012) but we found no evidence of a strong dependence on spatial variability according to MODIS SPI. ~~For a hypothetical calibration bias we ran the LUT retrieval with a -1 % shift in $I_{c,st}$, and this shifts median r_e by +0.2 μm.~~ We also ran the LUT retrieval with a -1 % scaling of $I_{c,st}$, which changes median r_e by +0.2 μm. Such a radiance shift could be necessary due to errors in calibration or in our derived scaling factor of 0.9804, which we used to relate the L1bSc file $I_{c,st}$ to our lookup table channels. We could therefore reduce our r_e bias by further scaling the $I_{c,st}$ radiances, but the scaling was derived from directly comparing the channel radiances rather than as a *post hoc* correction to improve retrieval results. If the r_e bias is due to other factors then this *post hoc* correction could result in compensating errors which hide other flaws in the retrieval. Instrumentally, the MODIS band 7 used in these r_e retrievals begins at $\lambda = 2.105$ μm, outside the strong CO_2 band. Changes in CO_2, or, more likely, temperature and vapour-driven broadening or vapour absorption could affect retrieved r_e. When retrieving with the midlatitude profile LUT described above, the median retrieved r_e increases by 0.17 μm. Given that the r_e discrepancy is small, we make no further efforts to explain or reduce it.

5. Summary and Conclusions

Here we developed a new pre-processor for a future optimal estimation retrieval using the OCO-2 A-band to provide new estimates of droplet number concentration in marine water clouds. This future retrieval aims to address limitations in the previously published OCO2CLD-LIDAR-AUX product, by (1) removing the requirement for collocated CALIPSO data now that the satellites are no longer formation flying and (2) adding OCO-2 information about r_e to extend the analysis to droplet number concentration. The pre-processor must identify footprints that likely contain liquid clouds of sufficient τ, and provide prior properties for the future cloud retrieval. It may also be useful for identifying appropriate footprints on which other researchers could conduct partial column XCO2 retrievals.

The pre-processor first flags potentially cloudy scenes using a multi-layer perceptron network fed with continuum radiances across all 3 OCO-2 bands plus a set of absorption band radiances from the O_2 A-band. The next stage of the retrieval is to use a three-dimensional lookup table that that jointly retrieves τ, r_e and P_{top} using radiances from two bands plus an A-band absorption ratio. Footprints whose radiances are inconsistent with the lookup table, or whose implied P_{top} occurs where $T < 0$ °C can also be excluded from future retrievals. These footprints were associated with worse fit statistics in OCO2CLD-LIDAR-AUX, implying that the new pre-processor will minimise the waste of computational resources on poor quality retrievals.

This pre-processor flag shows excellent agreement with MODIS and OCO-2, and the lookup table τ and r_e compare well with MODIS, while its P_{top} shows better retrieval statistics than MODIS, when taking CALIPSO as the truth. Many of the inter-satellite differences are associated with known factors: false positives from the classifier occur when scenes contain broken or multi-layered clouds, and the τ retrieval bias grows with the horizontal heterogeneity of the scene.

A main concern is that the median OCO-2 retrieved P_{top} is closer to the surface than CALIPSO's, by approximately 12 hPa (~120 m). The assumed mean cloud extinction or its profile will affect photon paths lengths and so could introduce a bias in retrieved P_{top}, and we propose that a detailed analysis of cloud vertical structure is the next and final step before the development of a new OCO-2 cloud retrieval. If successful, this new retrieval would add independent information on cloud droplet number concentration, allowing attempts to resolve apparent disagreements about low cloud processes.

Data sources sklearn is on github at: https://github.com/scikit-learn/scikit-learn

Refractor is on github at: https://github.com/ReFRACtor/framework

OCO2CLD-LIDAR-AUX can be downloaded from the CloudSat Data Processing Center at: http://www.cloudsat.cira.colostate.edu/data-products/level-aux/oco2cld-lidar-aux

Acknowledgements This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged. MR gratefully

acknowledges Dr. Aronne Merrelli for his updated cloud scattering property tables that corrected an error in CO_2 band absorption and removed resultant biases in retrieved r_e. MR also thanks Annmarie Eldering and Michael Gunson for helpful input related to the OCO-2 mission.

675

References

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K. and Yang, P.: MODIS Cloud-Top Property Refinements for Collection 6, J. Appl. Meteorol. Climatol., 51(6), 1145–1163,
680 doi:10.1175/JAMC-D-11-0203.1, 2012.

Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, J. Geophys. Res., 112(D2), D02201, doi:10.1029/2006JD007547, 2007.

Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R. and Elsaesser, G. S.: Strong Dependence of Atmospheric Feedbacks on Mixed-Phase Microphysics and Aerosol-Cloud Interactions in HadGEM3, J.
685 Adv. Model. Earth Syst., 11(6), 1735–1758, doi:10.1029/2019MS001688, 2019.

Boesch, H., Brown, L., Castano, R., Christi, M., Connor, B., Crisp, D., Eldering, A., Fisher, B., Frankenberg, C., Gunson, M., Granat, R., McDuffie, J., Miller, C., Natraj, V., O'Brien, D., O'Dell, C., Osterman, G., Oyafuso, F., Payne, V., Polonsky, I., Smyth, M., Spurr, R., Thompson, D. and Toon, G.: Orbiting Carbon Observatory (OCO)-2 Level 2 Full Physics Algorithm Theoretical Basis Document, Pasadena, CA. [online] Available from:
690 https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_L2_ATBD.V8.pdf, 2017.

Bony, S. and Dufresne, J.-L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophys. Res. Lett., 32(20), L20806, doi:10.1029/2005GL023851, 2005.

Crisp, D.: NASA Orbiting Carbon Observatory: measuring the column averaged carbon dioxide mole fraction from space, J. Appl. Remote Sens., 2(1), 023508, doi:10.1117/1.2898457, 2008.

695 Crisp, D., Atlas, R. ., Breon, F.-M., Brown, L. ., Burrows, J. ., Ciais, P., Connor, B. ., Doney, S. ., Fung, I. ., Jacob, D. ., Miller, C. ., O'Brien, D., Pawson, S., Randerson, J. ., Rayner, P., Salawitch, R. ., Sander, S. ., Sen, B., Stephens, G. ., Tans, P. ., Toon, G. ., Wennberg, P. ., Wofsy, S. ., Yung, Y. ., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D. and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Adv. Sp. Res., 34(4), 700–709, doi:10.1016/j.asr.2003.08.062, 2004.

700 Davis, A. B. and Knyazikhin, Y.: A Primer in 3D Radiative Transfer, in 3D Radiative Transfer in Cloudy Atmospheres, pp. 153–242, Springer-Verlag, Berlin/Heidelberg., 2005.

Davis, A. B., Merlin, G., Cornet, C., Labonnote, L. C., Riédi, J., Ferlay, N., Dubuisson, P., Min, Q., Yang, Y. and Marshak, A.: Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: An optimal estimation approach, J.

Quant. Spectrosc. Radiat. Transf., 216, 6–16, doi:10.1016/j.jqsrt.2018.05.007, 2018.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A. M., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D. and Yoshimizu, J.: The Orbiting Carbon Observatory-2: First 18 months of Science Data Products, Atmos. Meas. Tech. Discuss., 1–30, doi:10.5194/amt-2016-247, 2016.

Ferlay, N., Thieuleux, F., Cornet, C., Davis, A. B., Dubuisson, P., Ducos, F., Parol, F., Riédi, J. and Vanbauce, C.: Toward New Inferences about Cloud Structures from Multidirectional Measurements in the Oxygen A Band: Middle-of-Cloud Pressure and Cloud Geometrical Thickness from POLDER-3/ PARASOL, J. Appl. Meteorol. Climatol., 49(12), 2492–2507, doi:10.1175/2010JAMC2550.1, 2010.

Fischer, J. and Grassl, H.: Detection of Cloud-Top Height from Backscattered Radiances within the Oxygen A Band. Part 1: Theoretical Study, J. Appl. Meteorol., 30(9), 1245–1259, doi:10.1175/1520-0450(1991)030<1245:DOCTHF>2.0.CO;2, 1991.

Grosvenor, D. P., Sourdeval, O. and Wood, R.: Parameterizing cloud top effective radii from satellite retrieved values, accounting for vertical photon transport: quantification and correction of the resulting bias in droplet concentration and liquid water path retrievals, Atmos. Meas. Tech., 11(7), 4273–4289, doi:10.5194/amt-11-4273-2018, 2018.

Heidinger, A. K. and Stephens, G. L.: Molecular Line Absorption in a Scattering Atmosphere. Part III: Pathlength Characteristics and Effects of Spatially Heterogeneous Clouds, J. Atmos. Sci., 59(10), 1641–1654, doi:10.1175/1520-0469(2002)059<1641:MLAIAS>2.0.CO;2, 2002.

Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W. and de Haan, J. F.: A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, J. Geophys. Res. Atmos., 106(D4), 3475–3490, doi:10.1029/2000JD900657, 2001.

Kokhanovsky, A. A., Rozanov, V. V., Burrows, J. P., Eichmann, K.-U., Lotz, W. and Vountas, M.: The SCIAMACHY cloud products: Algorithms and examples from ENVISAT, Adv. Sp. Res., 36(5), 789–799, doi:10.1016/j.asr.2005.03.026, 2005.

Kokhanovsky, A. A., Mayer, B., Rozanov, V. V., Wapler, K., Burrows, J. P. and Schumann, U.: The influence of broken cloudiness on cloud top height retrievals using nadir observations of backscattered solar radiation in the oxygen A-band, J. Quant. Spectrosc. Radiat. Transf., 103(3), 460–477, doi:10.1016/j.jqsrt.2006.06.003, 2007.

L'Ecuyer, T. S. and Jiang, J. H.: Touring the atmosphere aboard the A-Train, Phys. Today, 63(7), 36–41, doi:10.1063/1.3463626, 2010.

Lindstrot, R., Preusker, R., Ruhtz, T., Heese, B., Wiegner, M., Lindemann, C. and Fischer, J.: Validation of MERIS Cloud-Top Pressure Using Airborne Lidar Measurements, J. Appl. Meteorol. Climatol., 45(12), 1612–1621, doi:10.1175/JAM2436.1, 2006.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V. and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11(1), 409–427, doi:10.5194/amt-11-409-2018, 2018.

Merlin, G., Riedi, J., Labonnote, L. C., Cornet, C., Davis, A. B., Dubuisson, P., Desmons, M., Ferlay, N. and Parol, F.: Cloud information content analysis of multi-angular measurements in the oxygen A-band: application to 3MI and MSPI, Atmos. Meas. Tech., 9(10), 4977–4995, doi:10.5194/amt-9-4977-2016, 2016.

Nakajima, T. and King, M. D.: Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory, J. Atmos. Sci., 47(15), 1878–1893, doi:10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2, 1990.

Natraj, V. and Spurr, R. J. D.: A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering–absorbing media, J. Quant. Spectrosc. Radiat. Transf., 107(2), 263–293, doi:10.1016/j.jqsrt.2007.02.011, 2007.

O'Brien, D. M. and Mitchell, R. M.: Error Estimates for Retrieval of Cloud-Top Pressure Using Absorption in the A Band of Oxygen, J. Appl. Meteorol., 31(10), 1179–1192, doi:10.1175/1520-0450(1992)031<1179:EEFROC>2.0.CO;2, 1992.

O'dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O. and Wunch, D.: The ACOS CO 2 retrieval algorithm – Part 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5(1), 99–121, doi:10.5194/amt-5-99-2012, 2012.

O'Dell, C. W.: Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation, J. Geophys. Res., 115(D10), D10206, doi:10.1029/2009JD012803, 2010.

Painemal, D. and Zuidema, P.: Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements, J. Geophys. Res. Atmos., 116, D24206, doi:10.1029/2011JD016155, 2011.

Preusker, R., Fischer, J., Albert, P., Bennartz, R. and Schüller, L.: Cloud-top pressure retrieval using the oxygen A-band in the IRS-3 MOS instrument, Int. J. Remote Sens., 28(9), 1957–1967, doi:10.1080/01431160600641632, 2007.

Richardson, M. and Stephens, G. L.: Information content of OCO-2 oxygen A-band channels for retrieving marine liquid cloud properties, Atmos. Meas. Tech., 11(3), 1515–1528, doi:10.5194/amt-11-1515-2018, 2018.

Richardson, M., McDuffie, J., Stephens, G. L., Cronk, H. Q. and Taylor, T. E.: The OCO-2 oxygen A-band response to liquid marine cloud properties from CALIPSO and MODIS, J. Geophys. Res. Atmos., 122, 8255–8275, doi:10.1002/2017JD026561, 2017.

Richardson, M., Leinonen, J., Cronk, H. Q., McDuffie, J., Lebsock, M. D. and Stephens, G. L.: Marine liquid cloud geometric thickness retrieved from OCO-2's oxygen A-band spectrometer, Atmos. Meas. Tech., 12, 1–21, doi:10.5194/amt-12-1-2019, 2019.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding Theory and Practice, World Scientific, Singapore., 2000.

Rosenfeld, D., Zhu, Y., Wang, M., Zheng, Y., Goren, T. and Yu, S.: Aerosol-driven droplet concentrations dominate coverage and water of oceanic low-level clouds, Science (80-.)., 363(6427), eaav0566, doi:10.1126/science.aav0566, 2019.

Rozanov, V. V. and Kokhanovsky, A. A.: Semianalytical cloud retrieval algorithm as applied to the cloud top altitude and the cloud geometrical thickness determination from top-of-atmosphere reflectance measurements in the oxygen A band, J. Geophys. Res., 109(D5), D05202, doi:10.1029/2003JD004104, 2004.

Schepers, D., Butz, A., Hu, H., Hasekamp, O. P., Arnold, S. G., Schneider, M., Feist, D. G., Morino, I., Pollard, D., Aben, I. and Landgraf, J.: Methane and carbon dioxide total column retrievals from cloudy GOSAT soundings over the oceans, J. Geophys. Res. Atmos., 121(9), 5031–5050, doi:10.1002/2015JD023389, 2016.

Schuessler, O., Loyola Rodriguez, D. G., Doicu, A. and Spurr, R.: Information Content in the Oxygen A-Band for the Retrieval of Macrophysical Cloud Parameters, IEEE Trans. Geosci. Remote Sens., 52(6), 3246–3255, doi:10.1109/TGRS.2013.2271986, 2014.

Spurr, R. J. D.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Radiat. Transf., 102(2), 316–342, doi:10.1016/j.jqsrt.2006.05.005, 2006.

Szczodrak, M., Austin, P. H. and Krummel, P. B.: Variability of Optical Depth and Effective Radius in Marine Stratocumulus Clouds, J. Atmos. Sci., 58(19), 2912–2926, doi:10.1175/1520-0469(2001)058<2912:VOODAE>2.0.CO;2, 2001.

Taylor, T. E., O'Dell, C. W., Frankenberg, C., Partain, P. T., Cronk, H. Q., Savtchenko, A., Nelson, R. R., Rosenthal, E. J., Chang, A. Y., Fisher, B., Osterman, G. B., Pollock, R. H., Crisp, D., Eldering, A. and Gunson, M. R.: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data, Atmos. Meas. Tech., 9(3), 973–989, doi:10.5194/amt-9-973-2016, 2016.

Toll, V., Christensen, M., Quaas, J. and Bellouin, N.: Weak average liquid-cloud-water response to anthropogenic aerosols, Nature, 572(7767), 51–55, doi:10.1038/s41586-019-1423-9, 2019.

Vanbauce, C., Buriez, J. C., Parol, F., Bonnel, B., Sèze, G. and Couvert, P.: Apparent pressure derived from ADEOS-POLDER observations in the oxygen A-band over ocean, Geophys. Res. Lett., 25(16), 3159–3162, doi:10.1029/98GL02324, 1998.

Várnai, T. and Marshak, A.: Observations of Three-Dimensional Radiative Effects that Influence MODIS Cloud Optical Thickness Retrievals, J. Atmos. Sci., 59(9), 1607–1618, doi:10.1175/1520-0469(2002)059<1607:OOTDRE>2.0.CO;2, 2002.

Vidot, J., Bennartz, R., O'Dell, C. W., Preusker, R., Lindstrot, R. and Heidinger, A. K.: CO 2 Retrieval over Clouds from the OCO Mission: Model Simulations and Error Analysis, J. Atmos. Ocean. Technol., 26(6), 1090–1104, doi:10.1175/2009JTECHA1200.1, 2009.

Werner, F., Zhang, Z., Wind, G., Miller, D. J. and Platnick, S.: Quantifying the Impacts of Subpixel Reflectance Variability on Cloud Optical Thickness and Effective Radius Retrievals Based On High-Resolution ASTER Observations, J. Geophys. Res. Atmos., 123(8), 4239–4258, doi:10.1002/2017JD027916, 2018.

Yamamoto, G. and Wark, D. Q.: Discussion of the letter by R. A. Hanel, "Determination of cloud altitude from a satellite," J. Geophys. Res., 66(10), 3596–3596, doi:10.1029/JZ066i010p03596, 1961.

Yang, Y., Marshak, A., Mao, J., Lyapustin, A. and Herman, J.: A method of retrieving cloud top height and cloud geometrical thickness with oxygen A and B bands for the Deep Space Climate Observatory (DSCOVR) mission: Radiative transfer simulations, J. Quant. Spectrosc. Radiat. Transf., 122, 141–149, doi:10.1016/j.jqsrt.2012.09.017, 2013.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A. and Taylor, K. E.: Causes of higher climate sensitivity in CMIP6 models, Geophys. Res. Lett., 2019GL085782, doi:10.1029/2019GL085782, 2020.

Zhang, Z., Ackerman, A. S., Feingold, G., Platnick, S., Pincus, R. and Xue, H.: Effects of cloud horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case studies based on large-eddy simulations, J. Geophys. Res. Atmos., 117(D19), n/a-n/a, doi:10.1029/2012JD017655, 2012.

Zhou, Y., Yang, Y., Gao, M. and Zhai, P.-W.: Cloud detection over snow and ice with oxygen A- and B-band observations from the Earth Polychromatic Imaging Camera (EPIC), Atmos. Meas. Tech., 13(3), 1575–1591, doi:10.5194/amt-13-1575-2020, 2020.

Table 1. Summary of methods for determining properties in OCO2CLD-LIDAR-AUX, and changes introduced in this study. OCO2CLD-LIDAR-AUX is a full optimal estimation (OE) retrieval that combined CALIPSO and OCO-2 information to obtain its prior state. This study is intended to provide OCO-2 only prior information for a future OE retrieval.

Property	OCO2CLD-LIDAR-AUX	This study
Cloud flagging	1. CALIPSO single layer cloud 2. CALIPSO P_{top} > 680 hPa 3. OCO-2 radiances exceed static thresholds 4. OCO-2 weak/A-band radiance ratio above fixed threshold for given A-band radiance	Multi-layer perceptron network classification based on OCO-2 radiances and radiance ratios
Cloud effective radius (r_e)	Fixed r_e = 12 μm in retrieval	Estimated from OCO-2 radiances via 3d

		lookup table
Cloud top pressure (P_{top})	Prior from collocated CALIPSO 01kmCLay, posterior from OE retrieval	Estimated from OCO-2 radiances via 3d lookup table
Cloud optical depth (τ)	Prior from lookup table map to A-band radiance, posterior from OE retrieval	Estimated from OCO-2 radiances via 3d lookup table
Cloud geometric thickness (H)	Prior from subadiabatic model using prior,	Not reported: assumed subadiabatic where needed for radiative transfer, derived from τ, P_{top}, r_e.
Retrieved properties	τ, P_{top}, H	τ, P_{top}, r_e for prior in future retrieval. Prior H and/or N_d to be derived from subadiabatic model.

Table 2. Summary of datasets used. Note that we use the 01kmCLay and MYD061KM products collocated with OCO-2 as described in Taylor et al. (2016).

Dataset name	Long name	Summary of data used
L1bSc	OCO-2 Level 1b Science data	OCO-2 radiances, geographic information and geometry for radiative transfer, instrument information.
L2Met	OCO-2 Level 2 Meteorological data	Footprint T profiles for T_{top}, mean T and q profiles for LUT RT.
01kmCLay	CALIPSO 1 km cloud layer product	Cloud layer presence, P_{top} and phase from feature classification flag
MYD061KM	MODIS Aqua 1 km cloud product	Cloud effective radius and optical depth
OCO2CLD-LIDAR-AUX	OCO-2 CALIPSO combined cloud retrieval	Preprocessor throughput, retrieval χ^2 and retrieval cloud properties

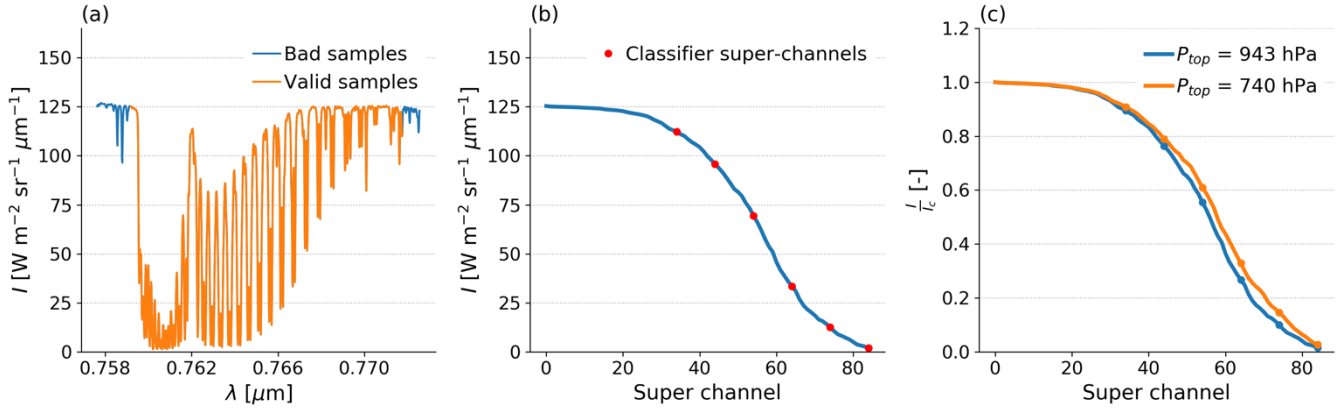


Figure 3.(a) OCO-2 A-band spectrum for a cloud with CALIPSO P_{top} = 943 hPa, with the used channels in orange and non-used (e.g. due to FPA pixel damage) in blue, (b) the smoothed super-channel spectrum, where the channels are ranked in brightness and then non-overlapping 10-channel means are taken. The super channels used in the classifier are shown in red. (c) Comparison of the ranked ratio I/I_c between this cloud, and one at higher altitude (P_{top} = 740 hPa). The SZA is within 0.01° and the continuum radiance within 1 % between each spectrum, the differences are largely due to the shorter path length resulting in less absorption for the higher altitude cloud.

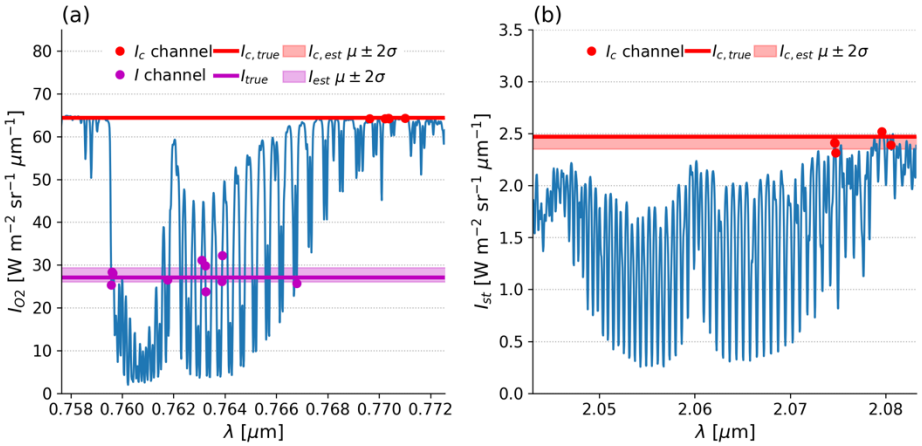


Figure 4. Mean cloudy scene spectra in (a) the O_2 A-band and (b) the strong CO_2 band. The channels used in the lookup tables are shown as points, red for the continuum radiance I_c and magenta for the O_2 absorption band radiance I. Thick horizontal lines represent the "truth", either the L1bSc file's continuum radiance for I_c or the mean of the 600—609th brightest undamaged channels (i.e. the 60th super-channel) for I. The shaded bands of the same colour are the mean±2 standard errors based on the selected channel sample sizes (5 for I_c, 10 for I).

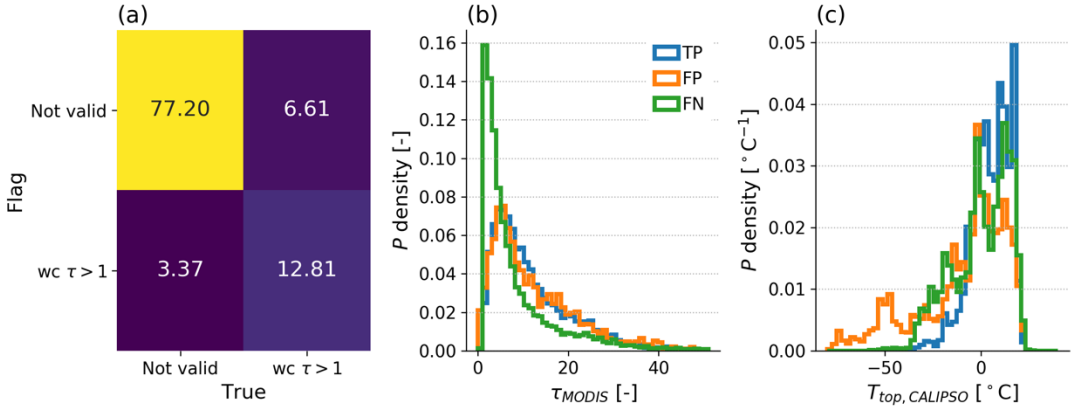


Figure 5. (a) confusion matrix with values in %, comparing trained classifier ("flag") with collocated MODIS-CALIPSO definitions ("truth") and entries being classified as single-layer water cloud (wc) with τ > 1, or "not valid", (b) normalised histograms of collocated MODIS τ where retrieved, for true positives (TP), false positives (FP) and false negatives (FN), (c) normalised histograms of collocated CALIPSO T_{top} where retrieved, colours as in (b).

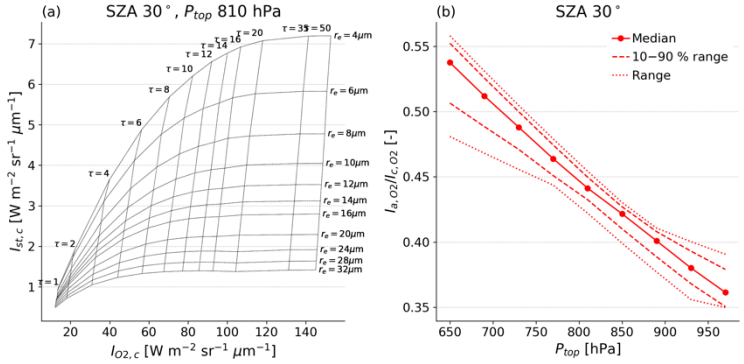


Figure 6. Example lookup table (LUT) properties. (a) radiance in the strong CO_2 continuum as a function of A-band continuum radiance at SZA = 30° for clouds with P_{top} = 810 hPa and τ, r_e as labelled, (b) A-band absorption ratio within the SZA = 30° table as a function of P_{top}. The solid line is the median value within each LUT at a fixed P_{top}, the dashed lines span the 10—90 % range and the dotted lines span the minimum to maximum values.

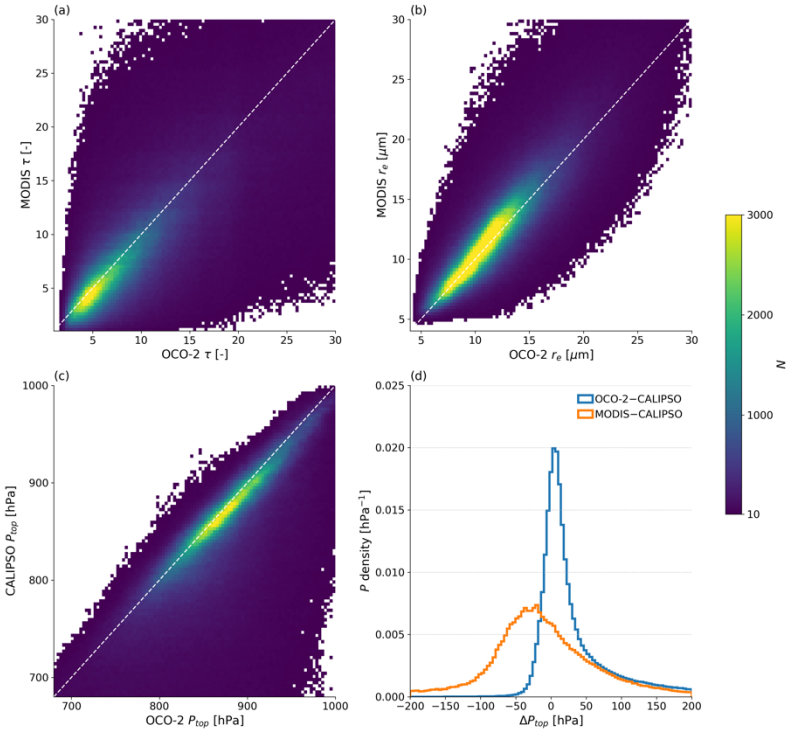


Figure 7. Inter-satellite comparison of retrieved cloud properties, (a) MODIS vs. OCO-2 τ, (b) MODIS vs. OCO-2 re, (c) CALIPSO vs. OCO-2 P_{top} and (d) OCO-2 or MODIS P_{top} minus CALIPSO P_{top}. The colour bar on the right applies to (a)—(c).

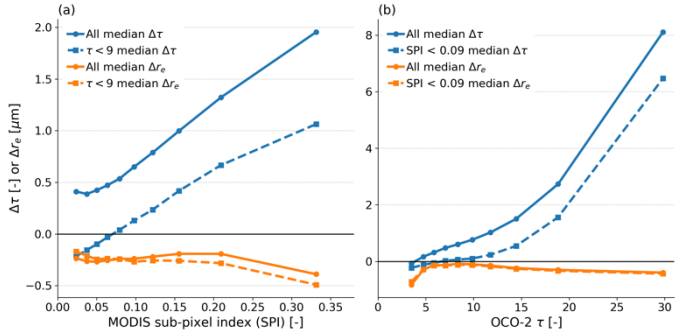


Figure 8. Binned median bias in OCO-2 minus MODIS τ (blue) or r_e (orange) when (a) binned by MODIS sub-pixel index (SPI) derived from the 250 m sampling at λ = 0.66 μm or (b) binned by the OCO-2 LUT retrieved τ. Solid lines are for the full samples, and dashed lines are for the subset (a) below the median OCO-2 τ or (b) below the median MODIS SPI.

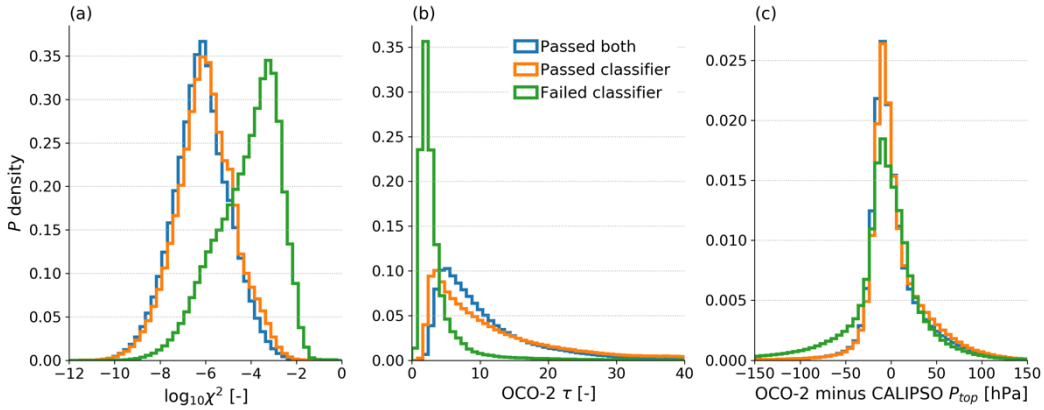


Figure 9. Normalised histograms of OCO2CLD-LIDAR-AUX outputs where SZA < 45° separated into whether the soundings pass the new pre-processor flag and retrieval or not. The "passed both" set are those that returned valid cloud properties from the LUT along with T_{top} > 0 °C, the "~~failed LUT~~passed classifier" case gave invalid cloud values or had T_{top} < 0 °C, and the "~~old but not new~~failed classifier" set are those that were attempted in OCO2CLD-LIDAR-AUX but are not passed by the new classifier. (a) Logarithm of χ^2, (b) retrieved τ, (c) P_{top} minus the closest CALIPSO retrieved P_{top}.