We thank the reviewer for the thoughtful review and kind words. Our responses to both the general and minor comments, which we hope adequately address the reviewer's concerns, are below.

General comments:

P11L324 – P11L330: The technique that utilized instrument-A (here RSP) to cross calibrate instrument-B (here eMAS) is very valuable to the community. I suggest [NOT MANDOTORY] adding an appendix to describe the technique in detail (for example, how to deal with different sensor geometries that might involve different attitude correction [uncertainty might become larger than radiometric uncertainty], instrument line shapes etc.) to make the technique 1) citable and 2) generally applicable to cross-calibration for other instruments.

> This is a good suggestion, though we note that the RSP cross-calibration provides only the temporal degradation (darkening) trend in the eMAS radiometric calibration and is not used as an absolute radiometric benchmark. Instead, the absolute radiometric benchmark was co-located observations from MODIS obtained during targeted under-flights of Aqua with the ER-2 flying along the nadir Aqua ground track. Those eMAS-MODIS comparisons were limited to the near-nadir observations from both imagers for which the observation angles were in rough agreement, and specifically utilized comparisons of the optical property retrievals that inherently account for sensor-specific characteristics (e.g., spectral response functions). The RSP cross-calibration used reflectance comparisons that did not account for spectral response differences and thus cannot provide an absolute radiometric benchmark. We've made some text additions to section 3.2 that we hope provide more clarity. We also are in the process of drafting a calibration readme document that will be posted in the LAADS data archive along with the L1B, as has been done for previous campaigns.

Based on the AMT reference guideline, If the author's name is part of the sentence structure only the year is put in parentheses ("As we can see in the work of Smith (2009) the precipitation has increased") E.g., P8L242 (Platnick, 2000) to Platnick (2000), P8L248 (Gupta et al., 2022b), P9L269 (Gupta et al., 2022b) etc. Additionally, please make the reference order consistent throughout the paper (e.g., when multiple references are used, oldest reference goes first).

> Good catch. We relied on the Copernicus style template in Zotero for our references and citations, which greatly eased reference management but that obviously does not completely adhere to the AMT style guide. We've modified what we think are all the problematic citations, but can't guarantee that we didn't miss a few.

From measurement perspective, accounting for the vertical heterogeneity (z direction) of clouds while constraining the horizontal heterogeneity (xy direction) is quite challenging (which might explain the different results between sawtooth and ramp cases). I wonder what the author think is the better approach? Will the square spiral into the clouds help the constraint of horizontal heterogeneity of clouds?

> This is an excellent question. As with any in situ sampling approach, trade-offs abound. The "sawtooth" (or porpoise) sampling shown in our first case study maximizes the vertical sampling statistics, though the horizontal sampling is limited to the direction of flight and, depending on the heterogeneity of the cloud, may not be adequate. A square spiral may improve horizontal sampling, though the sampling of different vertical layers likely will be spatially decoupled. Ultimately, the sampling approach should be dictated by the science objectives. For evaluating remote sensing retrievals, particularly from sensors having coarser spatial resolution (e.g., spaceborne imagers, etc.), the question to be answered is can a single aircraft provide a representative sample of the retrieval domain, which is complicated by spatial and temporal changes in the cloud. We think this remains an open question.

Minor comments:

P5L134: What does "effective pixel size" mean? Field of view at different angles?

Correct, “effective pixel size” acknowledges the differences in the projected size of the instantaneous field of view that is a function of viewing angle and the altitude at which co-location of the angular observations occurs (e.g., ground, cloud-top).

P5L154: change to “have heritage with the cloud products … from MODIS science team” or something similar

Done.

P8L226: “sun/satellite” to “sun/sensor”

Done.

P8L227: Since at P41L886 the above-cloud gaseous absorption corrections were discussed, it would be good if some clarifications on the definition of “top-of-cloud reflectance” (whether gases and/or aerosols are taken into account) here.

Good suggestion. We’ve added the definition to P8L238-239, and also have added a reference to “top-of-cloud reflectance” to the atmospheric correction discussion on P42L999-1000.

P11L308: change “with some degree of confidence” to “(with some degree of confidence)”

Done, though using commas instead of parentheses (P11L319).

P15L385: In Figure 4, rightmost panel for CER (3.7 mm), a very different pattern (bright potion, relatively low retrieval uncertainty) shown up on the upper side, what can be the causes?

Good observation! That bright pattern corresponds to the region of smaller effective radius retrievals in all three spectral channels apparent in the corresponding right three panels of Figure 3. Since the retrieval uncertainties essentially are proportional to the sensitivity of the solution space to perturbations in the observed top-of-cloud reflectance (with the size of the perturbations determined by the magnitude of the impacts of the various error sources), the 3.7µm solution space is less sensitive to reflectance errors (perturbations) at these small sizes than it is at larger sizes, and this difference in sensitivity across sizes is much larger than it is in either the 2.13 or 1.62µm solution spaces.

P21L475: It looks like the $CAS_{shifted}$ was moving into the right direction to agree with PDI. Is it possible to use a different constraint to shift size bins so CAS matches PDI and claim that the in situ can achieve agreement? Or the constraint will become nonphysical?

$CAS_{shifted}$ does move in the right direction with respect to PDI, but it’s important to remember that PDI also has inherent uncertainties/biases such that neither probe can be considered an unambiguous truth, as we state in the paper. Different probes see different collections of drops in the same cloud due to numerous factors, many of which depend on measurement principle (sizing and counting uncertainty) and probe design (drop size range sampled, size of sample volume, etc.). Thus it’s extremely difficult to “correct” for these differences to achieve agreement between the probes, since as the reviewer suggests, agreement may be achieved for one moment of the DSD but unphysical results for another.

P33L751: “misleadingly intensifying the appearance of scene heterogeneity”: I think this statement is valid for visible. The appearance of heterogeneity should be solid as they come from the NIR or SWIR channels. In return, the cloud optical thickness (visible) will not contain large 3D bias, but CER (NIR) might still suffer from the 3D effects.

Agreed, there is indeed heterogeneity in the SWIR channels for this scene that influence the CER retrievals. The statement in question here was made simply to point out that we stretched the color scale for each spectral channel in this false color RGB such that we enhance the visibility of the glory, yet at the expense of enhancing also the contrast between relatively darker and brighter pixels outside the glory – to the extent that the “darker” cloud pixels look like cloud-free regions rather than just less bright clouds.

P36L807: "10mm" to "10 mm" (adding a space in-between)

Good catch! This is fixed.