Authors' Response to Reviews of

Sensitivity of bi-spectral retrievals to the fixed effective variance assumption in ship tracks

Iarla Boyce, Alice Cicirello, Edward Gryspeerdt

Atmospheric Measurement Techniques, `egusphere-2026-2326`

RC: *Reviewers' Comment*, **AR:** Authors' Response. Text shown in red indicates manuscript text removed during revision; text in blue indicates text added. Line numbers refer to the compiled revised manuscript.

We thank both reviewers for their careful and constructive reviews. We would draw attention to several overarching changes made in response, referenced below against the individual comments that triggered them:

G.1 Retitled the manuscript from "Assessing retrieval biases in ship tracks" to "Sensitivity of bi-spectral retrievals to the fixed effective variance assumption in ship tracks", and reframed the Abstract and Introduction to present the paper as a controlled sensitivity study of one assumption, with ship tracks as the motivating test case (RC1.1, RC1.4, RC2.C).

G.2 Switched the central retrieval baseline from the legacy $v_{\text{eff}} = 0.13$ to the operational MODIS Collection 6 value $v_{\text{eff}} = 0.10$ throughout every figure, table, and quoted statistic, retaining 0.13 as an explicit secondary comparison (RC1.3).

G.3 Added a new sensitivity sweep across the assumed retrieval v_{eff} (0.07, 0.10, 0.13; new Sect. 3.3, Fig. 8, Table 3) (RC1.4).

G.4 Engaged quantitatively with Lebsock and Witte (2023): grounded the polluted $v_{\text{eff}} = 0.05$ value in their k–N_d relationship, implemented and tested their empirical correction (Discussion), and added a new test of how sensitive the reported biases and the correction's effectiveness are to the assumed strength of the k–N_d coupling itself (new Sect. 3.4, Table 4) (RC2.A, RC2.B).

G.5 Added a Limitations subsection before the Conclusion (RC1.15).

G.6 Revised causal language throughout to be conditional on the assumed degree of spectral narrowing being representative of real ship tracks (RC1.8, RC2.B).

---

1. RC1: Anonymous Referee #3

RC1.1 — Title and framing

RC: *The title may not accurately reflect the main scope of the study. Although ship tracks are an important and typical application, the core issue examined here is the retrieval bias introduced by*

assuming a fixed droplet effective variance. The current title, "Assessing retrieval biases in ship tracks", may imply a broader evaluation of ship-track retrieval uncertainties than the manuscript actually provides. The authors may consider revising the title to emphasize the effective-variance assumption, with ship tracks framed as the motivating or application case.

AR: We agree and have retitled the manuscript accordingly, retaining ship tracks as the named application case since it remains the paper's central focus; as they provide direct comparison of clean and polluted regions. This is essential for constraining and understanding aerosol-cloud interactions. The abstract now states explicitly that this is a controlled, synthetic retrieval experiment isolating the sensitivity to one assumption, with idealised ship tracks chosen as the test case because they provide the strongest realistic contrast between the assumed and true distribution widths.

`Title`

Assessing retrieval biases in ship tracks → Sensitivity of bi-spectral retrievals to the fixed effective variance assumption in ship tracks

RC1.2 — Practical significance relative to other uncertainty sources

RC: *The practical significance of the reported bias should be evaluated relative to other uncertainty sources... it remains unclear whether this bias is large or small compared with uncertainties from other retrieval variables and assumptions... The authors could compare the fixed-veff-induced bias with other known uncertainty terms... Lai et al. (2019, https://doi.org/10.3390/rs11141703) discussed differences among satellite cloud optical thickness and cloud effective radius products.*

AR: As this study focuses on ship tracks primarily, errors common to both clean and polluted regions are expected to cancel when contrasting the two. A full quantitative intercomparison would require implementing each of the other error sources within our synthetic framework, which we consider beyond the scope of a single-assumption sensitivity test. We have instead added an explicit qualitative placement in the new Limitations subsection, anchored specifically to ship-track studies where the clean-to-polluted contrast in v_{eff} is largest.

`Line 391 (Limitations, closing sentence)`

...the fixed-v_{eff} assumption warrants consideration alongside these established sources, particularly in ship-track studies where the clean-to-polluted contrast in v_{eff} is largest.

RC1.3 — $v_{\text{eff}} = 0.13$ as the primary retrieval assumption

RC: *The use of veff = 0.13 as the primary retrieval assumption is a major weakness. The manuscript acknowledges that MODIS Collection 6 uses veff = 0.10 for liquid clouds, yet most of the analysis and headline conclusions are based on 0.13. This makes the reported bias closer to an upper-bound historical case than to the current operational situation. Results for veff = 0.10 should be presented as a central case in the main text, with veff = 0.13 retained only as a sensitivity or legacy comparison.*

AR: Agreed; see G.2. $v_{\text{eff}} = 0.10$ is now the central retrieval baseline in every figure, table, and quoted statistic, with $v_{\text{eff}} = 0.13$ retained as an explicitly labelled secondary comparison. We would also note that this does not meaningfully change the results of the study, as the bias is just shifted, not removed.

`Line 88 (Sect. 2.1)`

...with this study setting this at a constant value $v_{\text{eff}} = 0.13$ to align with the historical assumption employed by the MODIS operational retrieval algorithm... → ...with this study setting this at

a constant value $v_{\text{eff}} = 0.10$ to align with the operational assumption employed by the MODIS Collection 6 retrieval algorithm...

`Line 229 (Sect. 3.2)`

...a mean bias of 24.56%, a large error... the standard deviation of this bias is quite large at ±23.75... → ...a mean bias of 13.91%, a large error... the standard deviation of this bias is quite large at ±22.14... As a sensitivity comparison, adopting the legacy $v_{\text{eff}} = 0.13$ baseline instead produces a substantially larger polluted N_d bias of 24.56% (±23.75%)...

RC1.4 — Justification and sensitivity range for the polluted $v_{\text{eff}} = 0.05$ value

RC: *The polluted-cloud value veff = 0.05 is treated as representative of a strongly narrowed ship-track droplet distribution, but the manuscript does not demonstrate that this value is typical for the cases to which the conclusions are applied. A sensitivity analysis over a realistic range of polluted veff values is needed. At minimum, the authors should show how re, LWP, and Nd biases change for veff values such as 0.05, 0.07, 0.10, and perhaps intermediate values representing aging ship tracks.*

AR: We have addressed this in two ways: (1) grounding $v_{\text{eff}} = 0.05$ quantitatively via the aircraft-derived k–N_d relationship of Lebsock and Witte (2023); (2) adding a new sensitivity sweep across the assumed retrieval v_{eff} (0.07, 0.10, 0.13; new Sect. 3.3). We note this sweep varies the assumed v_{eff} rather than the true polluted value the comment requests, and that directly testing the requested configuration would require new forward radiative-transfer simulations not feasible within the timeframe of this revision. The underlying mechanism is the same regardless of which quantity is varied: bias is driven by the assumed–true mismatch via $k(v_{\text{eff}})$, and Sect. 3.3 already shows bias magnitude tracks this mismatch rather than being an artefact of either specific baseline. Because $k(v_{\text{eff}})$ is nonlinear, the relationship is not perfectly symmetric between the two framings, but the qualitative conclusion, i.e., smaller mismatch, smaller bias, holds regardless of which side is varied. We return to the case of a weaker true v_{eff}–N_d coupling in new Sect. 3.4 (see RC2.B).

`Line 139 (Sect. 2.3)`

This choice is further supported by the empirical k–N_d relationship derived from aircraft observations by Lebsock and Witte (2023): at a typical ship-track droplet concentration of $N_d \approx 100\,\text{cm}^{-3}$, their Combined Fit gives $k \approx 0.84$, which corresponds via Eq. (4) to $v_{\text{eff}} \approx 0.05$, with heavier pollution implying still narrower distributions.

`Lines 265-279 (new Sect. 3.3, Fig. 8, Table 3)`

Sweep across assumed retrieval $v_{\text{eff}} = 0.07, 0.10, 0.13$, showing all four bias metrics scale monotonically with the assumed–true mismatch (polluted N_d bias: 24.56% → 13.91% → 7.31%).

RC1.5 — Fidelity to the operational MODIS retrieval

RC: *The retrieval experiment does not sufficiently mimic an operational MODIS retrieval. Operational products involve exact spectral response functions, multiple absorbing bands, cloud-mask and partly cloudy pixel screening, multilayer flags, retrieval failure logic, uncertainty estimates, and quality filtering... The authors should either implement a more realistic operational-like retrieval workflow or clearly state that the results are a simplified sensitivity experiment rather than a product-level bias estimate.*

AR: This is an idealised study to investigate the potential impact of a single effect in particularly extreme circumstances that are central to our current understanding of aerosol-cloud interactions

- shiptracks. The experiment deliberately isolates the fixed-v_{eff} assumption from all other components of the operational chain so the reported bias is attributable to that assumption alone. The new Limitations subsection states this explicitly.

`Line 386 (Sect. 4.1)`

...this study does not reproduce the full operational MODIS retrieval chain, including cloud masking, multilayer and partly-cloudy pixel flags, and other quality-control logic; the experiment is deliberately restricted to the bias introduced by the v_{eff} assumption alone... the reported bias magnitudes should be interpreted as a bound on the sensitivity to this one assumption under idealised conditions, rather than as a prediction of the net bias in any specific operational N_d product.

RC1.6 — Sensitivity of N_d bias to cloud thickness, adiabaticity, and LWP assumptions

RC: *The manuscript fixes cloud thickness at 500 m and imposes LWP conservation, but real marine boundary-layer clouds and ship tracks vary substantially. Please test the sensitivity of Nd bias to cloud thickness, adiabaticity, LWP, and optical-depth ranges, or discuss why the current fixed assumptions are adequate.*

AR: This connects to our response to RC1.2 and RC1.5. The aim of this study is not to assess every source of error in MODIS N_d retrievals. It is to assess how well the fixed-v_{eff} assumption works specifically for ship tracks, and hence how reliably aerosol-cloud interactions can be assessed from them. The potential errors from cloud thickness, adiabaticity, and LWP assumptions have already been established by other studies, including Grosvenor et al. (2018). We do not repeat that work here. The Limitations subsection notes the fixed cloud-thickness assumption, consistent with our narrower scope.

`Line 295 (Sect. 4.1, Limitations)`

First, the clouds are modelled as single-layer, vertically homogeneous, plane-parallel slabs with a fixed geometric thickness of $H = 500\,\text{m}$... the sensitivity of the reported bias magnitudes to H has not been tested.

RC1.7 — Treatment of N_d outliers exceeding 1200%

RC: *The very large Nd outliers exceeding 1200% require much more careful treatment. These cases may reflect numerical instability, LUT-edge interpolation issues, unrealistic geometry-cloud combinations, or known failure modes that would be removed by standard quality control. Please report the frequency of such cases, whether they pass realistic retrieval filters, and provide statistics both with and without them.*

AR: Quantified in Sect. 3.2. Only 1 of ∼10,000 cases exceeds the 1200% value quoted in the text, and removing it has little impact on the mean. But a broader heavy tail of large, if less extreme, biases drives the ±22.14% standard deviation. That single outlier does not pass standard r_e/τ QC filters; whether the broader heavy-tail cases also survive QC is not separately reported, and dedicated geometric filtering would be needed to remove the outlier regardless.

`Line 254 (Sect. 3.2)`

While these extreme outliers represent a small fraction of the total pixels, they dictate the heavy-tailed nature of the error distribution and drive the standard deviation to ±23.75%, reinforcing the common practice of filtering these geometries prior to analysis. → These extreme outliers are rare: across the full synthetic population of ∼10,000 cases, approximately 0.01% (one case)

exceeds the 1200% threshold, and excluding it shifts the mean polluted N_d bias by less than 0.3 percentage points (from 13.91% to 13.63%), confirming the mean estimate is robust to outlier removal. Nonetheless, the single outlier and the broader heavy tail drive the standard deviation to ±22.14%. Standard quality-control filters on retrieved r_e and τ do not remove this case, as it falls within physically plausible retrieval bounds. Dedicated geometric filtering of extreme forward-scattering angles is required. These geometries are not confined to the synthetic experiment. MODIS views up to viewing zenith angles of approximately 60–67° at swath edge. Combined with solar zenith angles around 60°, common outside near-noon overpasses, this range includes the forward-scattering geometry identified here. A practical filter would flag retrievals with both high viewing zenith angle and low scattering angle, matching the specific combination shown above to produce the largest bias.

RC1.8 — Overstated implications for susceptibility and marine cloud brightening

RC: *The implications for aerosol-cloud susceptibility and marine cloud brightening are overstated. The manuscript's synthetic experiment isolates one retrieval-assumption bias, but MCB efficacy and observed susceptibility depend on emissions, meteorology, cloud adjustments, precipitation, entrainment, and measurement strategy. The discussion should be rewritten to present MCB relevance as a possible implication rather than a conclusion supported by this analysis.*

AR: Agreed; see G.6. All downstream claims have been rewritten from implied-causal to explicitly conditional statements throughout the Abstract, Discussion, and Conclusion.

`Line 423 (Conclusion)`

Consequently, observational estimates of cloud susceptibility and the efficacy of marine cloud brightening may be exaggerated in current satellite products. → Under the same conditions, observational estimates of cloud susceptibility and the efficacy of marine cloud brightening could be exaggerated in current satellite products.

RC1.9 — Connection to real observations

RC: *The manuscript needs some connection to real observations. This does not necessarily require a full observational analysis, but the authors should at least compare their assumed veff values and bias magnitudes with published aircraft, in situ, or satellite-based ship-track studies. Without this comparison, readers cannot assess whether the synthetic scenarios represent common ship-track conditions or only an extreme bounding case.*

AR: Addressed. v_{eff} values are grounded via Miles et al. (2000) in the original text. Lebsock and Witte (2023) has been added as more recent grounding (Sect. 2.3). We are not aware of any published study that isolates the v_{eff} contribution specifically for ship tracks, so for bias magnitudes we instead cite two aircraft-based MODIS evaluations: Witte et al. (2018) find no significant r_e bias in marine stratocumulus, consistent with our near-zero clean-regime bias; Passer et al. (2025) report N_d discrepancies of ±50% or more, indicating our v_{eff}-driven N_d bias (14–25%) is a plausible but non-dominant contributor to total retrieval uncertainty.

RC1.10 — Statement of novelty

RC: *The paper would benefit from a clearer statement of novelty. The fact that droplet dispersion affects retrieved re and Nd is already known in the aerosol-cloud and retrieval literature. The authors should more explicitly identify what is new here: the ship-track-specific quantification, the bias propagation to susceptibility metrics, the geometry dependence, or the implication for MCB monitoring.*

AR: Agreed. The following addition was made to the end of the Introduction. We describe the contribution's relationship to prior work rather than asserting primacy, since the latter is a claim best defended here in the response rather than embedded in the manuscript text: this study is, to our knowledge, the first to quantify this bias for the ship-track clean-to-polluted contrast, propagate it to N_d and susceptibility metrics, characterise its geometry dependence, and test its robustness to the assumed k–N_d coupling strength.

`Line 65`

While the sensitivity of retrieved cloud properties to the assumed droplet size distribution is established in the retrieval literature, this study quantifies the resulting bias specifically for the clean-to-polluted contrast characteristic of ship tracks, where the fixed-variance assumption is particularly likely to break down, and propagates it through to its consequences for N_d and derived susceptibility metrics, characterises its dependence on viewing geometry, and tests its robustness to the assumed strength of the underlying k-N_d coupling. Susceptibility metrics of this kind underpin observational constraints on the radiative forcing from aerosol-cloud interactions (RFaci). Biases in them therefore propagate directly into RFaci uncertainty.

RC1.11 — Grammar and wording

RC: *Please check grammar and wording throughout, including phrases such as 'lead to to', 'overestimation the efficacy', 'cam become unstable', and 'maxmimising'.*

AR: All four corrected, along with several further instances of the same class of error caught during revision (e.g. "systemic"→"systematic" throughout, "induce"→"induces" for subject–verb agreement).

RC1.12 — Inconsistent effective-radius bias figures (abstract vs. text)

RC: *The abstract reports re overestimation of approximately 3%, while later values include 3.43% and 3.31%. Please harmonize the numbers or explain the different averaging choices.*

AR: Traced during revision: 3.43% was the mean polluted r_e bias under the (now legacy) $v_{\text{eff}} = 0.13$ baseline; 3.31% was a median from a separate interpolation-sensitivity variant mistakenly carried into the conclusion. A single, traceable set of numbers is now used throughout, and the abstract has been further reframed to report the clean-to-polluted contrast directly (see RC1.3, G.2) rather than an absolute bias against a single assumption.

`Line 11 (Abstract)`

...causes a systemic overestimation of effective radius (r_e) of approximately 3.4% in the polluted regime... → ...causes retrieved effective radius (r_e) to differ systematically between clean and polluted regimes. The polluted branch is overestimated by approximately 2.3% relative to the near-unbiased clean branch under the current MODIS Collection 6 assumption (rising to 3.4% under the legacy assumption), understating the true clean-to-polluted contrast in r_e...

RC1.13 — LUT grid spacing and interpolation

RC: *Please provide the LUT grid spacing, interpolation method, and treatment of retrievals near LUT boundaries similar to Liu et al. (https://doi.org/10.1016/j.aosl.2023.100337).*

AR: Documented as follows. We have not separately characterised retrieval behaviour specifically at LUT boundaries; our outlier analysis (RC1.7) captures the practical consequence of edge-of-LUT instability for extreme geometries.

`Line 168 (Sect. 2.4)`

The retrieval LUT spans effective radii from 2.0 to 20.0 μm in steps of 0.1 μm, and optical depths from 0.5 to 61, with a finer spacing of 0.25 between $\tau = 0.5$ and $\tau = 10$... and coarser steps of 1.0 for $\tau > 10$... following the variable-resolution LUT design used operationally for other bi-spectral retrievals (Liu et al., 2023).

RC1.14 — Figure caption completeness

RC: *Figure captions should state whether statistics include all geometries, filtered geometries, or outlier-contaminated cases.*

AR: All figure and table captions now state which population is shown and confirm that all statistics quoted in text and tables use the unfiltered population, including outliers.

`Figs. 1-8 captions (e.g. lines 127, 182)`

All viewing geometries are included; points outside the 1st–99th percentile of each bias variable are omitted for display clarity, while quoted statistics use the unfiltered population.

RC1.15 — Limitations subsection

RC: *Please include a concise limitations subsection before the conclusion. This should explicitly list the idealized 1D, homogeneous, LWP-conserving, fixed-veff, and fixed-cloud-thickness assumptions.*

AR: A new Limitations subsection has been added immediately before the Conclusion.

`Lines 292-299 (new subsection)`

Limitations: (1) fixed 1D, homogeneous, plane-parallel, 500 m-thick cloud geometry; (2) isolation from the full operational MODIS retrieval chain; and a closing placement of these biases alongside other established retrieval uncertainty sources, anchored to ship-track studies.

---

2. RC2: Michael Diamond

RC2.A — Missing engagement with Lebsock and Witte (2023)

RC: *Lebsock & Witte (2023), in Atmospheric Chemistry and Physics, have an extensive discussion of the relationship between effective variance (via k) and Nd and propose their own correction to the adiabatic Nd calculation from effective radius (re) and cloud optical thickness retrievals. This should be discussed given the relevance to the present work; I'm very interested in knowing whether the proposed corrections based on in-situ aircraft data would substantially moderate the potential biases found in the present work.*

AR: See G.4. We now engage with Lebsock and Witte (2023) in three ways: (1) their Combined Fit relationship grounds our choice of the polluted $v_{\text{eff}} = 0.05$ value; (2) we implement their correction and apply it to our synthetic retrievals, finding it reduces the polluted N_d bias from 13.91% to 0.19% (central baseline); (3) we test how sensitive both the bias and the correction's effectiveness are to the strength of the assumed k–N_d coupling itself (new Sect. 3.4). We also simplified the practical recommendation following from this: apply the correction only within already-identified polluted regions, not scene-wide, and validate against observational v_{eff} retrievals before wider adoption.

`Line 328 (Discussion, new paragraph)`

Applying this correction (Combined Fit parameters, $k_1 = 0.562$, $k_2 = 0.974$, $N^* = 48.5\,\mathrm{cm}^{-3}$) to the synthetic retrievals here reduces the polluted N_d bias from 13.91% to 0.19% under the central $v_{\mathrm{eff}} = 0.10$ baseline, and from 24.56% to −1.82% under the legacy $v_{\mathrm{eff}} = 0.13$ baseline... However, the same correction applied uniformly to the clean background population does not improve, and in most cases worsens, the retrieval... These results suggest a targeted fix: apply the aircraft-derived k–N_d correction only within already-identified polluted regions, not scene-wide, and validate it against observational v_{eff} retrievals before wider adoption. Importantly, this estimate of the correction's benefit assumes a tight k–N_d coupling, and Sect. 3.4 shows both the bias and the benefit shrink if the true coupling is weaker.

`Lines 279-306 (new Sect. 3.4, new Table 4)`

Entirely new subsection: Replaces the fixed true $v_{\mathrm{eff}} = 0.05$ with a true k drawn from the Lebsock and Witte (2023) Combined Fit curve plus Gaussian scatter calibrated to $r \approx 0.10$. Mean polluted N_d bias falls from 13.91% to 3.80% ($v_{\mathrm{eff}} = 0.10$) and 24.56% to 13.62% ($v_{\mathrm{eff}} = 0.13$); the correction, which nearly zeroes the bias deterministically, instead overcorrects by 5–11 percentage points.

RC2.B — Overly strong/causal language about the k–N_d relationship

RC: *Although Lebsock & Witte (2023) do find a (noisy) relationship between k and Nd as discussed above, more conditional language about the nature of the k-Nd [relationship] may be more appropriate.* If *the relationship is strong and causal (and instantaneous), then the current results about biases hold. If the real relationship is substantially weaker, however, the bias may be much less important. I've personally been surprised by how weak the relationship between k and Nd seems to be based on aircraft data (for example, I did a quick k, Nd correlation on ORACLES low cloud data and get values of r < 0.1, albeit with high statistical significance).*

AR: Agreed; see G.4 and G.6. Causal/deterministic language has been softened to conditional throughout, and we added new Sect. 3.4 to directly test this concern.

`Line 280 (new Sect. 3.4, paragraph 1)`

This is a deliberate idealisation that isolates the retrieval bias from any assumption about how strongly droplet dispersion and concentration actually covary in nature, but it also means the reported bias magnitudes implicitly assume a perfect, noise-free correspondence between a cloud's true N_d and its true v_{eff}.

`Line 316 (Discussion)`

The finding that the standard retrieval may overestimate N_d in excess of 24% in polluted regimes implies that this bias contributes to an overestimation the microphysical sensitivity... → The finding that the standard retrieval may overestimate N_d by approximately 14%... suggests that, if the degree of spectral narrowing assumed here is representative, this bias could contribute to an overestimation of the microphysical sensitivity...

RC2.C — Generality beyond ship tracks

RC: *The authors repeatedly invoke ship tracks, but their argument applies seemingly with equal force to all studies of ACI.*

AR: Agreed; see G.1. Ship tracks remain the chosen test case because the clean/polluted v_{eff} contrast is large enough there to produce a measurable, well-bounded signal, and because they

let us speak directly to a specific open question (the instantaneous LWP response reported by Gryspeerdt et al., 2021).

RC2 — Specific comments

Line 26 ("obscured"):

RC: *Is 'obscured' the right word here? I understand you're referring to physical adjustments that offset the Twomey effect, not factors that interfere with its reliable quantification (e.g., swelling of aerosol near cloud edges).*

AR: We take the reviewer's point. Changed to "offset".

`Line 33 (Introduction)`

However, the initial Twomey cooling is often obscured by rapid cloud adjustments. → However, the initial Twomey cooling is often offset by rapid cloud adjustments.

Line 28 (Khadri et al., 2022 citation):

RC: *Khadri et al. (2022) is a surprising citation here; are the authors confident that paper supports this claim?*

AR: On review we agree it does not clearly support the specific claim about entrainment-driven LWP decrease, and have removed it.

`Line 34 (Introduction)`

...which may enhance evaporation and lead to to a subsequent decrease in LWP (Wood, 2007; Khatri et al., 2022; Ackerman et al., 2004; Bretherton et al., 2007). → ...which may enhance evaporation and lead to a subsequent decrease in LWP (Wood, 2007; Ackerman et al., 2004; Bretherton et al., 2007).

Line 53 (modern support for the N_d–k relationship beyond Martin et al., 1994):

RC: *Do you have support for the tight relationship between Nd and k from more modern sources than Martin et al. (1994)? I'm aware of some good work out of the cloud chamber field, but in my experience this relationship has not turned up so cleanly in recent aircraft observations. That said, Lebsock & Witte (2023) are still able to derive a physically plausible relationship, as discussed above.*

AR: Citing Lebsock and Witte (2023) as modern support.

`Line 60`

...reducing the v_{eff} to as little as 0.05; this is further supported by the more recent aircraft-derived k–N_d relationship of Lebsock and Witte (2023), discussed further in Sect. 2.3.

Lines 77–78 (citations for v_{eff} varying with aerosol loading, entrainment, precipitation onset):

RC: *Citations would again be useful here.*

AR: Resolved. Citations added: Martin et al. (1994) for aerosol loading, Lim and Hoffmann (2026) for entrainment-driven broadening, and Miles et al. (2000) for precipitation/drizzle-driven broadening.

`Line 90 (Sect. 2.1)`

This fixed assumption introduces a systematic bias, as the v_{eff} of a cloud varies significantly with aerosol loading, entrainment, and precipitation onset. → ... of a cloud varies significantly with aerosol loading (Martin et al., 1994), entrainment (Lim and Hoffmann, 2026), and precipitation onset (Miles et al., 2000)

Equations 5 and 6 (LWC–height inconsistency):

RC: *There are inconsistent physical assumptions between these equations in terms of how LWC and re vary with height above cloud base. I'd recommend sticking with the adiabatic assumptions in Grosvenor et al. (2018) given the paper's focus.*

AR: Clarified as follows.

`Lines 99-104 (Sect. 2.1)`

Here $H = 500\,\text{m}$ is the fixed geometric cloud thickness, used consistently in both cloud generation and N_d derivation to convert LWP to mean liquid water content (LWC = LWP/H)... Under the adiabatic assumption, Eq. (5) and Eq. (6) are mutually consistent: substituting Eq. (6) into Eq. (5) recovers $N_d = \tau/(2\pi k r_e^2 H)$.

Line 145 (Fu et al., 2022 does not support the statement):

RC: *I'm not sure how the results in Fu et al. (2022) support this statement.*

AR: Agreed; replaced with sources that support the respective claims.

`Line 129 (Sect. 2.3)`

...typical non-drizzling marine stratocumulus clouds (Fu et al., 2022)... → ...typical non-drizzling marine stratocumulus clouds (Wood, 2012)...

Lines 223–224 (IPCC AR6 ERFaci convergence):

RC: *In the latest IPCC report, the ERFaci estimate from observations and models are very similar.*

AR: Agreed; the original framing of a "persistent discrepancy" was outdated.

`Line 310 (Discussion, opening)`

The results of this study identify a systematic retrieval bias that likely contributes to the persistent discrepancy between satellite-derived and model-simulated estimates of aerosol-cloud radiative forcing... → The results of this study identify a systematic retrieval bias relevant to the ongoing challenge of constraining aerosol-cloud radiative forcing. While Forster et al. (2021) report improved convergence between observation-based and model-based ERFaci estimates, the uncertainty range remains large...

Lines 285–288 (would the recommendation change to simply adopting the Lebsock and Witte correction?):

RC: *Given the existence of Lebsock & Witte (2023), would this recommendation change to simply adopting their correction? Or is further work required?*

AR: Our answer is a qualified no. Blanket adoption removes the polluted-regime bias but introduces a comparable overcorrection of the clean background (mean clean N_d bias worsening from -7.65% to -11.26% under the central baseline); the correction performs best precisely where the distribution is genuinely narrow, arguing for conditional application to identified polluted pixels rather than scene-wide adoption. This recommendation itself assumes a tight coupling between droplet dispersion and concentration. As shown in Sect. 3.4, if the true coupling is weaker, the

correction systematically overcorrects the polluted regime instead of removing the bias, so its benefit is smaller and less certain than the deterministic case would suggest.

`Line 430 (Conclusion)`

...future work must focus on computationally efficient correction strategies for these standard operational algorithms. This could include developing machine learning models capable of dynamically estimating v_{eff} from existing satellite radiances, or deploying adaptive algorithms that conditionally apply lower dispersion assumptions when specific features, such as ship tracks, are detected. → ...future work should focus on computationally efficient correction strategies... or empirical k–N_d corrections of the kind proposed by Lebsock and Witte (2023), when specific features, such as ship tracks, are detected. The latter approach is supported by the finding here that the correction nearly eliminates the polluted-regime N_d bias when applied to identified ship-track pixels, but degrades the retrieval of the clean background when applied indiscriminately. These findings indicate that the fixed effective-variance assumption is a relevant source of bias for ship-track-based estimates of aerosol-cloud interactions and climate intervention efficacy, and should be evaluated alongside other known retrieval uncertainties in future work.

References (Stevenson et al., 2012 → Latham et al., 2012): AR: Corrected throughout.

`Line 365 (Discussion)`

...where the injection of sea salt aerosols will yield a significant increase in albedo (Stevenson et al., 2012). → ...where the injection of sea salt aerosols will yield a significant increase in albedo (Latham et al., 2012).

We thank both reviewers again for reviews that have materially improved the paper, in particular, the direction to Lebsock and Witte (2023) prompted a new quantitative analysis (Sect. 3.4) that we believe substantially strengthens the paper's central claims by testing their own robustness.