This document includes the reviewer comments (red) and author responses (black).

Review of "Precipitation Susceptibility of Marine Stratocumulus with Variable Above and Below-Cloud Aerosol Concentrations over the Southeast Atlantic" by Gupta et al.

This paper presents airborne observations from the ORACLES project that examine how cloud and precipitation characteristics vary with perturbations driven largely from the entrainment of free-tropospheric biomass burning aerosols into the southeast Atlantic marine boundary layer. The authors extend their previous study by incorporating a much larger observational dataset from additional flight years and they extend their prior work to look at precipitation susceptibility. I commend the authors for synthesising such a large dataset and I found the paper to be generally well written. The topic area is certainly suitable for publication in ACP. However, I do have a major concern about the use of the CAS probe to measure liquid water content and cloud drop size from the 2016 campaign (see below), that I feel the authors need to address before this paper can be published.

The authors thank the reviewer for their thorough review. These suggestions will improve the quality of the submitted manuscript. Each of the reviewer's comments has been addressed in this document.

Main concern

1. Use of CAS probe for 2016 flights. I have concerns about the use of the CAS data for calculating microphysical properties on the 2016 campaign. Although the cloud drop number concentration from the CAS looks reasonable when compared against the PDI (Fig S1a), the LWC looks to be underestimated when compared against both the PDI and King probes (Fig S1b, Fig S2). The authors show that the PDI can give higher LWC values when compared to the adiabatic value and so choose not to use that instrument. But given that the bulk LWC estimate from the King probe is also much higher that the CAS (Fig S2), I think the authors need to provide some additional justification as to why the CAS probe is thought to be reliable (for measurements of LWC and effective radius). A possible approach would be to look at cases where the cloud was expected to be more adiabatic (well mixed boundary layer, non-drizzling etc) and examine if the difference with the adiabatic LWC value shown in Fig S2 is still apparent. Alternatively, in precipitating clouds, can the overlap with the 2DS probe be looked at to at least check for consistency at the larger cloud drop sizes that contribute significantly to LWC. I also note however that a similar low bias in LWC from the CAS is shown in the 2017 and 2018 campaigns when compared against a CDP (Fig S4 and S6), which suggests that it could be a general measurement issue with the CAS measurements. Related to this point, if the authors removed the 2016 data from their analysis, do any of the conclusions of the paper change?

The main concern regarding the use of CAS data for the 2016 campaign is justified given the differences between the CAS and PDI datasets for 2016 and the CAS and CDP datasets for 2017/2018. Specific portions of the main concern are addressed:

A possible approach would be to look at cases where the cloud was expected to be more adiabatic (well mixed boundary layer, non-drizzling etc) and examine if the difference with the adiabatic LWC value shown in Fig S2 is still apparent.

Contact profiles underwent precipitation suppression and were not drizzling as heavily as separated profiles. However, the differences between CAS LWC and the adiabatic LWC were observed during both separated and contact profiles. It is unlikely the differences between CAS LWC and the adiabatic LWC were dependent on the adiabaticity of the clouds.

in precipitating clouds, can the overlap with the 2DS probe be looked at to at least check for consistency at the larger cloud drop sizes that contribute significantly to LWC

During frequent flight legs through regions with high aerosol concentration (N_a), soot deposition occurred on the optical lenses of the 2D-S probe which led to data artifacts. These artifacts were removed using additional constraints during data processing (Gupta et al., 2021). Due to these constraints, 2D-S measurements for droplets with diameter (D) below 50 μm were unusable. Since the CAS measurement range does not exceed 50 μm, the CAS and 2D-S datasets were not compared. Comparisons between the closest CAS and 2D-S size bins are shown below for cloud profiles from 6 September 2016 (within the response to minor comment number 13).

if the authors removed the 2016 data from their analysis, do any of the conclusions of the paper change?

If 2016 data were removed, there were minor changes in the trends of S_o versus cloud thickness (H) and negligible changes in the S_o differences between contact and separated profiles (Fig. 1). Most noticeably, there was a decrease in S_o for thick clouds ($H > 256$ m) with a negligible decrease in the overall S_o (Table 1). This was accompanied by a decrease in the number of thick clouds when the 2016 data were removed. These changes highlight that the discussion of S_o in the study was robust as it relates to the exclusion of the 2016 data. In the absence of CDP data, the N_c and LWC from the CAS and PDI datasets were compared to choose which dataset should be used to represent the 2016 campaign. These comparisons were quantified in the supplement for transparency.

S_o was calculated using PDI data from the 2016 campaign to determine the impact of choosing CAS data over PDI data. When the PDI data were used, there were minor changes in the trends of S_o versus H (Fig. 1) and the magnitude of S_o for different aerosol regimes (Table 1). This was because S_o depends on N_c and precipitation rate (R_p). The CAS and PDI datasets had small differences in the average N_c over the 2016 campaign (95 % confidence intervals of 9 to 12 cm^{-3}) and R_p was calculated using droplets with $D > 50$ μm which did not include contributions from either the CAS or the PDI. Since the 2016 campaign contributed about a third of the ORACLES measurements, data from the 2016 campaign were included in the study so as not to reduce the size of the dataset.

I also note however that a similar low bias in LWC from the CAS is shown in the 2017 and 2018 campaigns when compared against a CDP (Fig S4 and S6), which suggests that it could be a general measurement issue with the CAS measurements.

There could be a sizing bias in the CAS given the small differences in N_c and large differences in *LWC* from the CAS and PDI datasets (see supplement). The sizing bias could impact the quantitative results presented in the study. Since R_p does not depend on CAS LWC, the impact of the sizing bias would be limited to effective radius (R_e) and LWC for droplets with $D < 50$ µm (cloud water content or CWC). Assuming the King hot-wire provided an accurate estimate of LWC for the sampled clouds, the CAS droplet size distribution could be adjusted using the King LWC following the methodology by Painemal and Zuidema (2011). For the 2016 research flights, the King LWC and the CAS LWC had a best fit slope (a) between 0.46 and 0.63. The CAS $n(D)$ was scaled by adjusting the CAS size bins as

$$CAS\ LWC = a\ x\ King\ LWC,\ \ D_i^* = a^{-1/3}\ D_i\ , \tag{AC1}$$

where D_i is the bin midpoint for the i^th^ bin and D_i* is the scaled bin midpoint. CAS bin midpoints increased by up to 30 % since $R_i^* > R_i$ for $a < 1$ and each flight had $a < 1$. This led to higher R_e and CWC for both contact and separated profiles (Fig. 2). The difference between the average R_e for contact and separated profiles increased from 1.5 µm to 1.7 µm when CAS data were scaled (Table 2). The difference in the average LWC decreased from 0.02 g m^{-3} to 0 when CAS data were scaled. Scaling the CAS data thus would not have a large impact on the results.

To avoid confusion with the use of two datasets from the CAS (original or scaled using King LWC), the authors used the original CAS dataset in the study. Given the minor impact of using the PDI/CAS/no data on S_o (Table 1), the original CAS dataset could be used. The differences between the datasets from different probes were quantified in the supplement. We believe these responses addressed any concern regarding the use of CAS data from the 2016 campaign. We believe the documentation of the differences between the ORACLES cloud probes in the supplement provides appropriate information about the uncertainties associated with any analyses using the ORACLES in situ cloud probes.

Minor comments

1. Line 34: Suggest changing to "changes in microphysical properties" in this sentence as the proceeding sentence states that LWP and cloud thickness are similar. Also it is not clear what the reference to "existing relationships" means. Which relations are you referring to? Are these based on previous observations or parameterized/simulated in models for example?

The sentence was updated by adding the underlined words: "These results suggest the changes in cloud microphysical properties were driven by ACIs rather than meteorological effects, and the existing relationships between R_p and N_c in model parameterizations must be adjusted to account for the role of ACIs".

2. Line 68: assuming constant LWP in what?

The sentence was updated by adding the underlined words: "Since τ has greater sensitivity to LWP compared to N_c, assuming constant LWP under different aerosol conditions can lead to underestimation of the cloud albedo susceptibility to aerosol perturbations (Platnick and Twomey, 1994; McComiskey and Feingold, 2012)."

3. Line 78: sink of liquid water rather than LWP?

"LWP" was changed to "LWC".

4. Line 82: Suggest changing to "relates the change in Rp…." as the actual definition is Eq 9.

The sentence was updated.

5. Line 87: Suggest changing to "parameterized in models"

The sentence was updated.

6. Line 93: Suggest changing to something like "A focus of recent field experiments in the southeast Atlantic Ocean has been to study ACI in this unique meteorological ……..".

The sentence was updated to: "Recent field campaigns focused on studying ACIs over the southeast Atlantic Ocean because unique meteorological conditions are present in the region"

7. Line 98: I assume the values of single scattering albedo and above cloud optical depth are from ORACLES. Please make that clear. Much higher optical depths can occur in the region e.g. Peers at al. (2021).

Lines 110-115 were moved ahead of the sentence and the sentence was updated to: "During ORACLES, the aerosol layer was comprised of shortwave-absorbing aerosols (500 nm single-scattering albedo of about 0.83) with above-cloud aerosol optical depth up to 0.42."

8. Line 102: Suggest changing "positive forcing" to "aerosol absorption of SW radiation". Also please expand on why this decreases entrainment e.g. strengthening inversion.

The sentence was updated to: "Warming aloft due to aerosol absorption of solar radiation strengthens the temperature inversion which decreases dry air entrainment into clouds, increases LWP and cloud albedo, and decreases the shortwave CRF".

9. Line 149: Please include references for the different probes.

The references were added.

10. Line 159: Change to “Hawaii”

The sentence was updated.

11. Line 165: What thresholds on the PCASP data were used to screen cloud?

This was moved after Line 209 where in-cloud measurements thresholds were defined.

12. Line 168: Change to “with the CDP N(D) for 50 < …..”

We have left this unchanged given that CAS n(D) was used for ORACLES 2016 and CDP n(D) was used for ORACLES 2017 and 2018.

13. Line 170: Can the authors comment on how well the different probes compared for drop sizes where they overlap?

The probes generally had good agreement in the overlap regions. This was determined by comparing the number distribution function, $N(D)$, from different probes. Figure 3 shows the sawtooth flight patterns flown to sample clouds on 6 September 2016. Figure 4 shows the average $N(D)$ from the CAS, the 2D-S, and the HVPS-3 during these flight patterns. The largest difference between $N(D)$ from the largest CAS size bin and the smallest 2D-S size bin was sampled during the 3rd sawtooth when the average CAS $N(D)$ was 3.1×10^{-3} cm^{-3} $\mu\mathrm{m}^{-1}$ and the average 2D-S $N(D)$ was 6.2×10^{-3} cm^{-3} $\mu\mathrm{m}^{-1}$. It is difficult to determine the differences between the 2D-S and the HVPS-3 because large droplets with sizes where these probes have good overlap (800 to 1200 μm) were very rarely sampled during the ORACLES research flights. For the 3rd sawtooth, the 2D-S $N(D)$ from the size bin centered near 450 μm was 22.9×10^{-9} cm^{-3} $\mu\mathrm{m}^{-1}$ and the corresponding HVPS-3 $N(D)$ was 8.8×10^{-9} cm^{-3} $\mu\mathrm{m}^{-1}$.

14. Line 176: refer the reader to the supplement.

The sentence was updated.

15. Line 215: Please briefly outline what you mean by shallow boundary layers can provide an underestimate of LWP adjustments. And I don’t understand what you mean in the sentence on line 217. Please clarify.

This text (Line 214-218) was removed for brevity.

16. Line 219: Have you done any analysis of the thermodynamic data to ascertain the frequency of well-mixed vs decoupled boundary layers from the vertical profiles used in this study? And are the clouds studied typically a single layer of stratocumulus, rather than cumulus rising into stratocumulus. If the latter form a significant number of profiles,

can the authors comment on how that may impact the results e.g. cumulus could transport aerosol from the surface mixed layer up into the overlying cloud.

Such analysis was not conducted due to the cloud sampling strategy during ORACLES. A complete sounding of the boundary layer or the free troposphere was not conducted near most profiles. Instead, cloud sampling was conducted in a sawtooth pattern where the aircraft ascended or descended through the cloud layer in quick succession. During the sawtooth patterns, the aircraft only flew up to about 100 m above or below cloud to maximize cloud sampling within a short distance (Fig. 3). Therefore, if a cumulus layer was present, it was not always sampled. When multiple cloud layers were sampled during a sawtooth pattern, a single cloud layer that capped the boundary layer was selected to limit data analysis to stratocumulus clouds. Thermodynamic analyses based on in situ measurements were also hindered by concerns about aircraft measurements of dew point and mixing ratio during descents (Gupta et al., 2021).

17. Equation 1: Is effective radius calculated as a function of height or is it a cloud top value.

Effective radius (R_e) was calculated for each 1 Hz data sample within cloud. Equation 1 was updated to avoid confusion:

$$“R_e\,(h) = \int_3^\infty D^3\, N(D,h)\, dD \Big/ \int_3^\infty 2\, D^2\, N(D,h)\, dD\,. \qquad (1)”$$

The following text was added: “Based on the aircraft speed, 1 Hz data samples corresponded to roughly 5 m intervals in the vertical direction.”

18. Equation 4: What does LWC(zB) mean? Is it your fixed value of 0.05 g m-3 to define cloud base?

This could be any value greater than 0.05 g m^{-3}. Equation 4 was removed for brevity.

19. Line 283/Table 4: Is the effective radius calculated at cloud top, or is it an average value through the depth of the cloud?

R_e was calculated for each 1 Hz data sample within cloud. During a cloud profile, this meant a value for effective radius was calculated at 5 m intervals. These values were then averaged over the entire depth of the cloud.

20. Line 377: Suggest changing “ascent” to “updraft”

The sentence was updated by adding the underlined words: “Alternatively, there may not have been sufficient time for the updraft to produce the few large droplets needed to broaden the size distribution and initiate collision-coalescence.

21. Line 385: Suggest changing “explained” to “parameterized”

The sentence was updated.

22. Figure 5: y-axis labels should be SAUTO95 and SACC95

The labels were updated.

23. Line 420: Add comment as to why you use CO as a proxy for the airmass from biomass burning aerosol source regions.

The following text was added: “Carbon monoxide (CO) concentrations were examined since CO acts as a biomass burning tracer that is unaffected by precipitation scavenging (Pennypacker et al., 2020).”

24. Line 422: Suggest the authors may want to comment on how the different regimes occur. For example, does the S-H regime have high boundary layer aerosol loadings because of previous entrainment events prior to the aircraft sampling and does the C-L regime have low aerosol loadings because there hasn’t been sufficient time for aerosol to be mixed down into the boundary layer from the overlying aerosol plume.

Pennypacker et al. (2020) found the boundary layer CO concentrations during the biomass burning season (July to October) were elevated compared to CO concentrations from the non-burning season (December to April). They argued the boundary layers were previously polluted by biomass-burning aerosols with low aerosol concentration (N_a) due to precipitation scavenging. The C-L regime could be explained by precipitation scavenging since these boundary layers also had elevated CO concentrations. Aerosol-induced precipitation suppression would explain why the number of C-L cases was much less than C-H cases. The S-H regime could be hypothesized to be associated with prior entrainment events (Diamond et al., 2018). We have not added these statements to the text because of their speculative nature given that the studies referenced examined boundary layers at a different location (Ascension island) and flight days.

25. Line 433: Figure 6 does not show comparisons of Re as stated in the text.

The sentence was updated by adding a reference to Table 5 for the comparisons of R_e.

26. Line 443: The authors test boundary layer aerosol concentration thresholds of 300 to 400 cm-3 to split “low” and “high” aerosol conditions. Even the value of 300 cm-3 seems like a moderately polluted boundary layer though compared to pristine marine conditions, where I would expect values < 100 cm-3 to be typical. Did ORACLES measure cleaner boundary layer conditions and if yes, how do the cloud properties in these cleaner clouds compare to the broader “low” aerosol regime used in this study? Or do future studies need to compare/contrast with more offshore airborne measurements from the CLARIFY campaign for example?

Ground-based observations at Ascension Island showed monthly accumulation mode N_a around 200 cm^{-3} in 2016 and around 400 cm^{-3} in 2017 during the biomass burning season (Pennypacker et al., 2020). These values suggest our definition of low N_a was reasonable since these measurements were collected during the peak of the burning season. Only 2 out of the 329 profiles used in this study (having 40 1-Hz in-cloud samples) were flown within a boundary layer with $N_a < 100$ cm^{-3}. These profiles had average below-cloud $N_a = 53$ cm^{-3} and average $N_c = 23$ cm^{-3}. Given the small number of data samples, measurements outside the burning season may be needed to compare the data used in this study with pristine boundary layers.

27. Line 448: Is quartiles the correct term, given that the bins don't have an equal number of profiles (82,80,85,82 in Table 6)?

The sentence was updated to: "The populations were divided at $H = 129$, 175, and 256 m to ensure similar sample sizes (Table)."

28. Figure 9: It looks like there are data points with Nc = 0 cm-3. Is that correct?

There were no data points with N_c below 10 cm^{-3} even though it may seem like they have values close to zero. The definition of in-cloud measurements ($N_c > 10$ cm^{-3} and King LWC > 0.05 g m^{-3}) was used to screen each data sample used in the study.

29. Line 489: Suggest adding "decreased with H from …..include your numbers for the lowest H bin data….to 0.53"

The sentence was updated by adding the underlined text: "For separated profiles, S_o decreased with H <u>from 1.47 ± 0.10 for $H < 129$ m</u> to 0.53 ± 0.09 for $129 < H < 175$ m and to 0.34 ± 0.07 for $175 < H < 256$ m".

30. I found some of the text hard to follow in section 6.3, that was at least in part due to the number of figures that were referred to in short succession. For example, the short paragraph beginning on line 489 refers to four different figures. I would suggest the authors consider rewriting some of the text and highlighting key points, to make it easier for the reader.

The text in Section 6.3 has been edited for clarity based on this comment.

31. Line 491: I struggle to see how the reader can see this change by looking at Fig 9 b,c.

The reference to Figure 9 is removed.

32. Line 513: The paragraph starts by mentioning appendix B, but does not then summarize the key point of that sensitivity study. Further, it states that the appendix investigates the inclusion of precipitating clouds, but I think it instead looks at the impact of the removal of non-precipitating clouds.

The text was updated to: “The sensitivity of S_o to removal of clouds based on R_p was examined in Appendix B. The removal of clouds with low R_p and high N_c or with high R_p and low N_c resulted in lower average S_o consistent with previous work (Duong et al., 2011).”

33. Line 600: What is the mechanism that results in higher RWP for these contact profiles?

This is because a subset of the contact and separated profiles was compared. A physical basis for increased RWP in polluted conditions for model low-cloud cover > 0.95 is not expected.

34. Line 625: Suggest rephrasing this statement. The separated polluted boundary layers have presumably also experienced entrainment events prior to the aircraft sampling, even though there is no contact at the time of the measurements. The timescales of entrainment and history of the airmass do also need to be considered (Diamond et al., 2018), rather than just an instantaneous measure of “contact” vs “separated”.

Conclusion #2 was changed to: “Aerosol-induced cloud microphysical changes in both clean and polluted boundary layers.” The following line was updated: “Contact profiles were more often located in polluted boundary layers and had higher below-cloud CO concentration (27 to 29 ppb higher) which suggests more frequent entrainment of biomass-burning aerosols into the boundary layer compared to separated profiles.”

35. Line 630: Make it clear that this is when compared to separated profiles.

The sentence was updated by adding the underlined words: “Contact profiles had 25 to 31 cm^{-3} higher N_c and 0.2 to 0.5 μm lower R_e in clean and 98 to 108 cm^{-3} higher N_c and 1.6 to 1.8 μm lower R_e in polluted boundary layers compared to separated profiles.”

References

Diamond, M. S., et al., 2018. Atmos. Chem. Phys., https://doi.org/10.5194/acp-18- 14623-2018, 2018.

Peers, F., et al., 2021. Atmos. Chem. Phys., 21, 3235–3254, https://doi.org/10.5194/acp-21-3235-2021.

REFERENCES:

Gupta, S., McFarquhar, G. M., O'Brien, J. R., Delene, D. J., Poellot, M. R., Dobracki, A., Podolske, J. R., Redemann, J., LeBlanc, S. E., Segal-Rozenhaimer, M., and Pistone, K.: Impact of the variability in vertical separation between biomass burning aerosols and marine stratocumulus on cloud microphysical properties over the Southeast Atlantic, Atmos. Chem. Phys., 21, 4615–4635, https://doi.org/10.5194/acp-21-4615-2021, 2021.

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., 116, D24206, https://doi.org/10.1029/2011jd016155, 2011.

TABLES AND FIGURES:

Table 1: S_o ± standard error for all profiles, with sample size and R in parentheses.

H	CAS data from 2016	No data from 2016	PDI data from 2016
All	0.88 ± 0.03 (329, 0.33)	0.83 ± 0.03 (258, 0.33)	0.90 ± 0.02 (329, 0.35)
28 to 129 m	0.67 ± 0.07 (82, 0.28)	0.58 ± 0.07 (80, 0.26)	0.68 ± 0.07 (84, 0.29)
129 to 175 m	0.68 ± 0.05 (80, 0.32)	0.73 ± 0.05 (63, 0.35)	0.73 ± 0.05 (79, 0.35)
175 to 256 m	0.54 ± 0.05 (85, 0.20)	0.84 ± 0.06 (58, 0.31)	0.71 ± 0.05 (86, 0.26)
256 to 700 m	1.13 ± 0.04 (82, 0.40)	0.75 ± 0.04 (57, 0.30)	1.10 ± 0.04 (80, 0.41)

Table 2: Average and standard deviation for R_e and CWC over all three ORACLES deployments calculated with original CAS data from 2016 (underlined) and CAS data from 2016 scaled using King LWC (bold).

Parameter	Contact	Separated
R_e (μm)	7.5 ± 2.1	9.0 ± 3.0
R_e (μm)	**7.8 ± 2.1**	**9.5 ± 2.9**
CWC (g m^{-3})	0.23 ± 0.17	0.21 ± 0.14
CWC (g m^{-3})	**0.24 ± 0.17**	**0.24 ± 0.15**

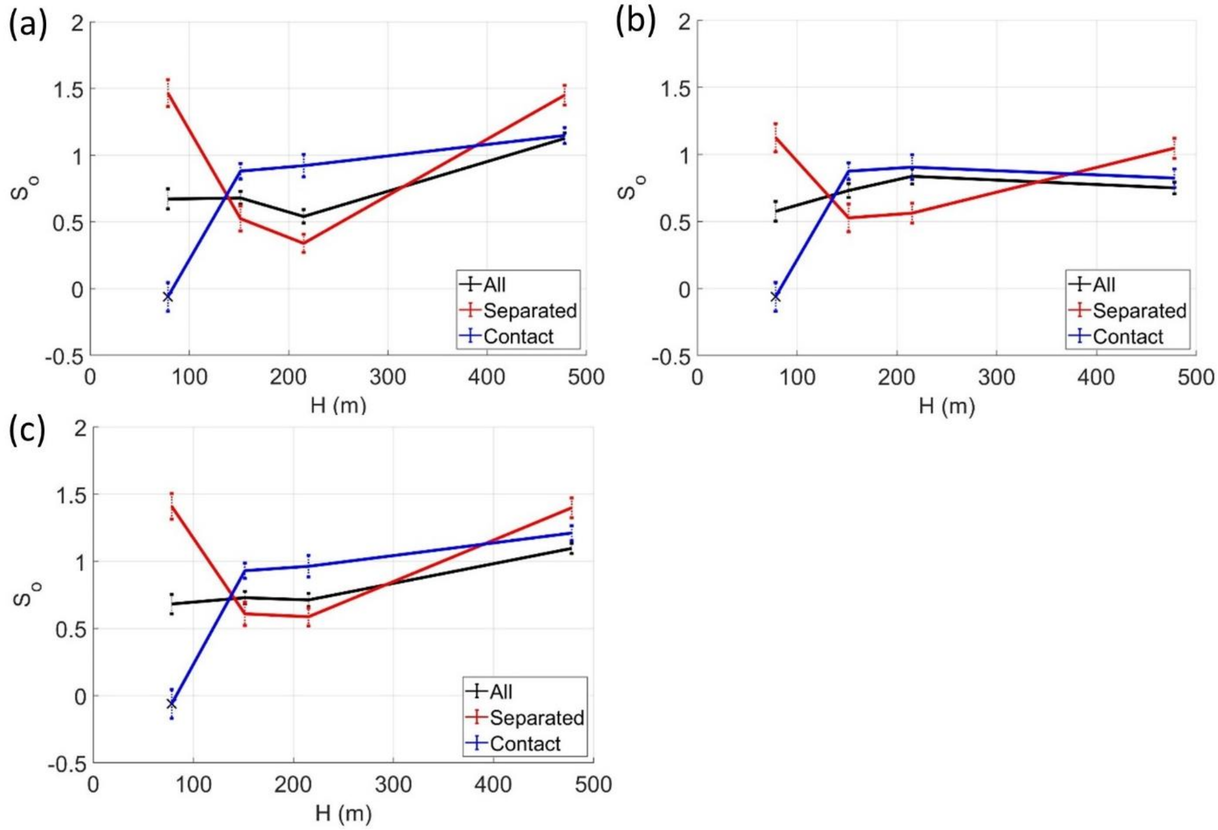

Figure 1: S_o as a function of H (error bars extend to standard error from regression model) using (a) CAS data from 2016, (b) no data from 2016, and (c) PDI data from 2016.

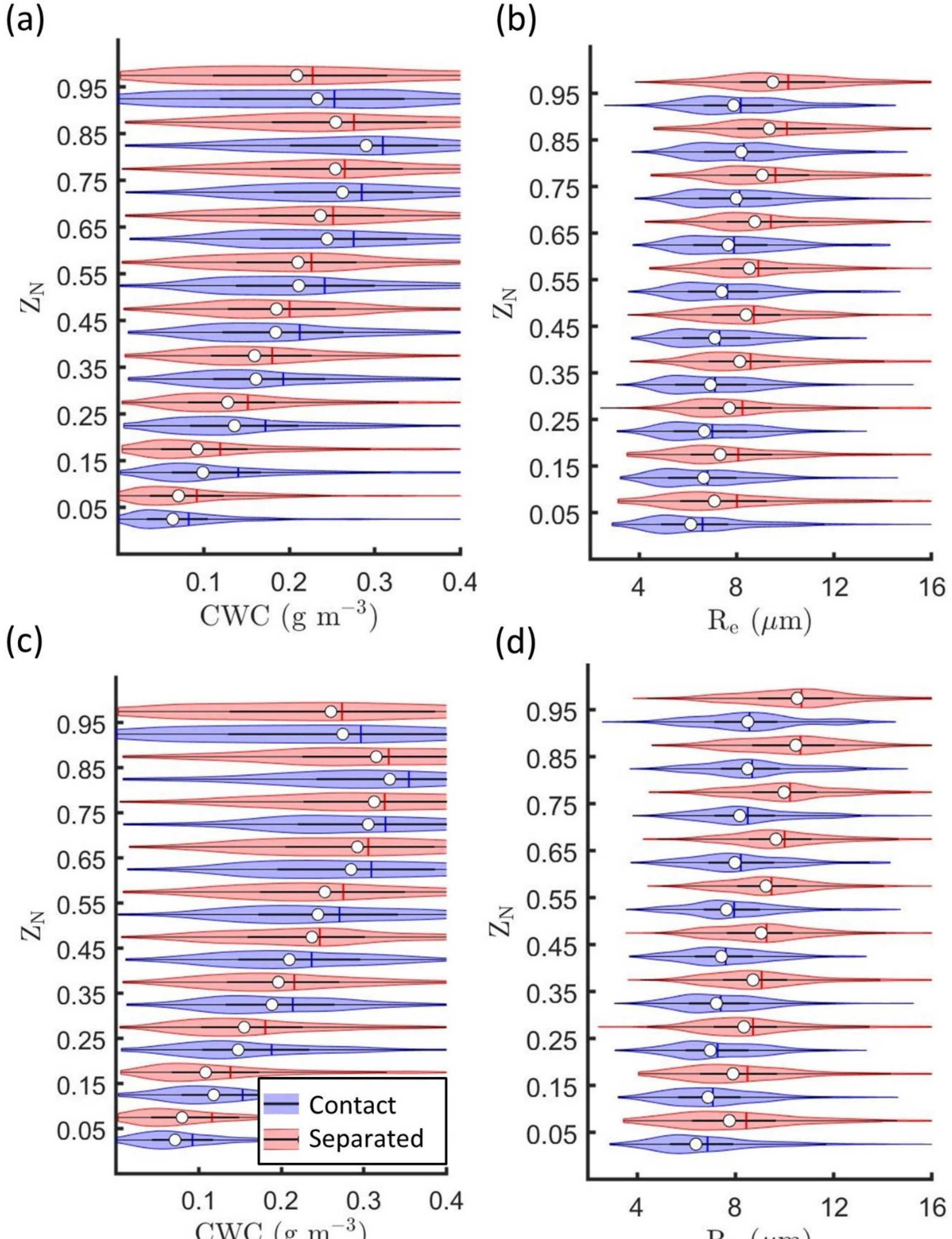

Figure 2: Kernel density estimates (indicated by the width of shaded area) and boxplots showing the 25th, 50th (white circle), and 75th percentiles for R_e and CWC over the three ORACLES deployments calculated using (a, b) original CAS data from 2016 and (c, d) CAS data from 2016 scaled using the King LWC.

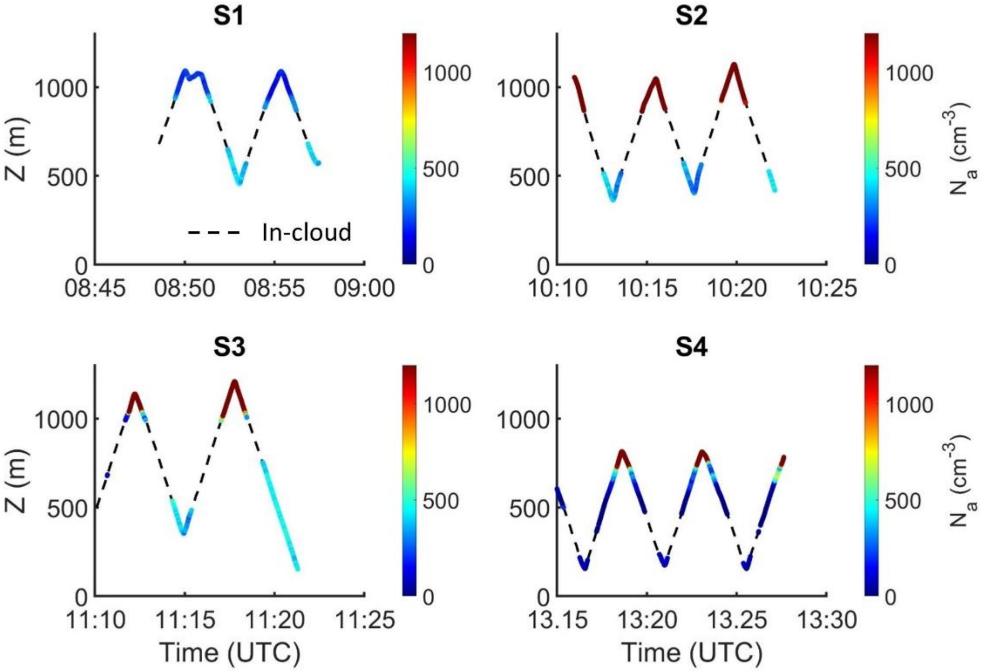

Figure 3: P-3 aircraft altitude as a function of time during sawtooth flight patterns. Data are colored by accumulation mode aerosol concentration (taken from Gupta et al., 2021).

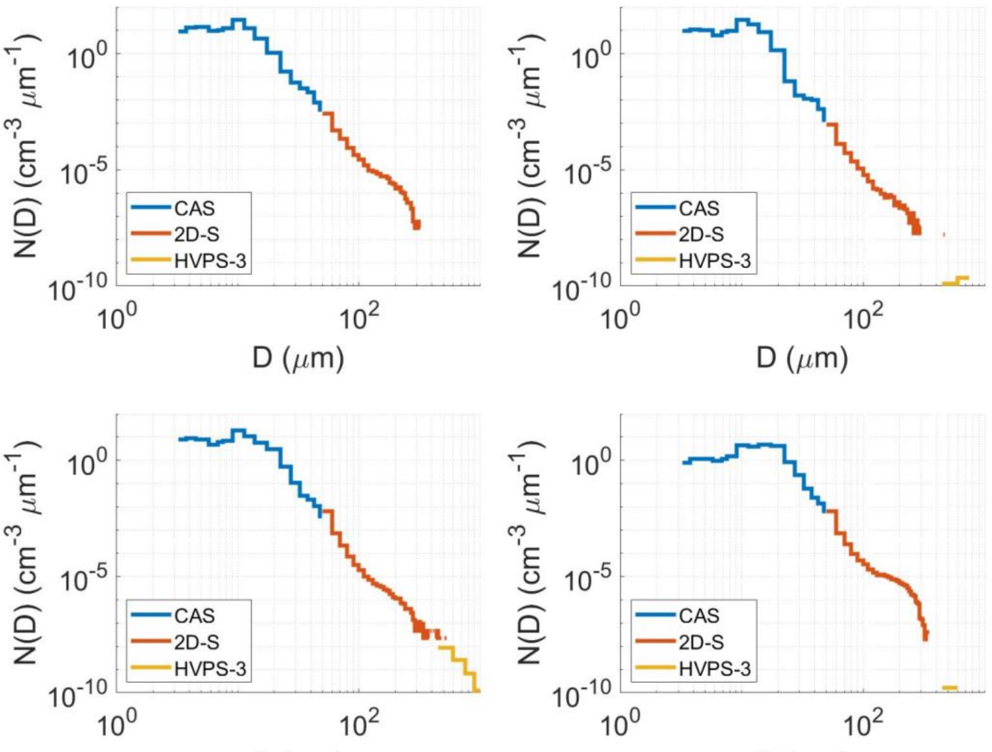

Figure 4: The average number distribution function *N(D)* from different probes for sawtooth patterns shown in Fig. 3.