Response to Referee #1

Review for "Aircraft Sampling Representativeness for LASSO Domain Averages over the ARM Southern Great Plains (SGP) site" (egusphere-2026-2786) submitted by Liang et al.

Summary:
The manuscript takes up an important challenge – how well do models represent in-situ aircraft sampling of cloud water content (chosen as a variable to represent cloud properties) if the aircraft sampling strategies are represented in selecting model data as opposed to using domain averages.

This is an important topic, and the manuscript tackles this problem with a thoughtful methodology. Nevertheless, I felt the manuscript was missing many pieces and the text could benefit from better organization. Often I found key information was missing, or the authors made statements/arguments before introducing the details that could potentially support those statements/arguments (some of my comments below get answered partially by the text but much after they arise). Specifically, improvements can be made to the description of the flight tracks, a better link can be established with real-world aircraft observations, and issues associated with the "flat profile" type should be addressed, etc. Finally, the authors have a wealth of information given the work emulate "flight tracks" within model output - many details lost within aircraft observations can be inferred here to really do a more comprehensive evaluation of not just the model output relative to domain averages, but also the aircraft sampling itself.

I recommend major revisions before the manuscript is reconsidered with my detailed comments below.

Detailed comments:

Title: The manuscript title suggests the work is broader than it is, especially given the length of the results section which has essentially three figures, two of which examine CWC. The abstract represents this directly in Line 10, and I suggest the title be reworded to mention "CWC" specifically, or similar.

Author response: We thank the referee for this suggestion. We agree that the original title was broader than the scope of the manuscript, as the analysis primarily focuses on the representativeness of cloud water content (CWC) sampling. We have therefore revised the title to explicitly highlight CWC and better reflect the main focus of the study.

The revised title is:
"Evaluating the Representativeness of Aircraft Sampling Strategies for Cloud Water Content in LASSO Simulations over the ARM Southern Great Plains (SGP) Site"

Line 29: I believe the authors can do a better job motivating the need for observations and simulations to be combined to answer cloud-related questions in ESMs. This sentence comes a bit abruptly.

Author response: We thank the referee for this suggestion. We agree that the original transition

from the importance of cloud representation to the need for combining observations and simulations was too abrupt. We have revised this paragraph to better explain the complementary roles of observations and high-resolution simulations in evaluating cloud representation in Earth system models. Specifically, the revised text clarifies that observations provide constraints on cloud properties and their variability, while high-resolution simulations provide process-level information that can help interpret these observations and their representation in models. The revised text has been added to Lines 33–40.

Line 35: This should be re-phrased. Active remote sensing does penetrate clouds and can often get comparable vertical resolution.

Author response: We thank the referee for pointing this out. We agree that the original wording was too broad and could incorrectly imply that remote-sensing instruments cannot penetrate clouds or resolve their vertical structure. We have therefore revised the sentence to more accurately describe the complementary role of aircraft observations. The revised text emphasizes that aircraft provide direct in situ measurements along targeted three-dimensional trajectories, while ground-based and remote-sensing observations, including active sensors, can profile detailed information on cloud vertical structure with different sampling geometries and sensitivities.

Line 43: I recommend citing this recent article that summarizes the latest AAF campaigns and a dataset from ARM AAF - https://essd.copernicus.org/articles/16/5429/2024/.

Author response: We thank the referee for recommending this recent overview of the ARM Aerial Facility (AAF) campaigns and airborne datasets. We agree that Mei et al. (2024) provides a useful and up-to-date context for the AAF within the ARM user facility and for the airborne measurements available from recent campaigns. We have therefore added Mei et al. (2024) as a reference and included information on the recently integrated AAF airborne dataset to provide additional context for the observational framework of this study.

The revised text has been added to Lines 56–58.

Line 60: “These aircraft measurements” is confusing, the AAF deployment during HI-SCALE should be introduced first?

Author response: We thank the referee for pointing this out. We agree that the original wording referred to the aircraft measurements before the relevant AAF deployment had been clearly introduced (Line 51–60). We have therefore reorganized the paragraph to first introduce the role of the ARM Aerial Facility, followed by the deployment of the G-1 aircraft during the HI-SCALE campaign over the Southern Great Plains (SGP) site. We then describe the resulting airborne measurements and their complementary role alongside ARM ground-based observations. This revision provides a clearer introduction to the aircraft observations used to motivate the sampling framework in this study.

Line 84: The definition of WRF should be moved to line 78.

Author response: We thank the referee for this helpful suggestion. We agree that WRF should be

defined at its first occurrence. We have therefore reorganized the paragraph so that the Weather Research and Forecasting (WRF) model is defined when the LASSO framework is first introduced, before the abbreviation is used subsequently. The paragraph now introduces the LASSO framework and its WRF-based simulations first, followed by previous applications of LASSO and the two releases analyzed in this study.

Line 120: This citation refers to ERA-Interim – is that the version of the reanalysis that was used here? This should be mentioned here in the text explicitly and the variable(s) used should be provided or the methodology, if novel, should be addressed here directly.

Author response: We thank the referee for pointing this out. The ECMWF forcing used in these simulations is not derived from ERA-Interim. Instead, it is derived from the ECMWF Integrated Forecast System (IFS) using the Diagnostics in the Horizontal Domains (DDH) approach, which provides the physical and dynamical tendencies used to construct the large-scale forcing. We have clarified this in the revised manuscript, including the 114 km forcing scale used in this study, and removed the potentially misleading reference to ERA-Interim.

Line 128: typo for "hydrometeor".

Author response: We thank the referee for pointing this out. The typo has been corrected from "hydrometer" to "hydrometeor".

Line 129: add references?

Author response: We thank the referee for this suggestion. We have added the reference for the Thompson microphysics scheme and its aerosol-aware formulation and revised the text to describe the scheme more precisely. The revised manuscript now cites Thompson et al. (2008) and Thompson and Eidhammer (2014), with the latter describing the aerosol-aware formulation used in the simulations.

Line 140: I'm a bit confused here. How can there be consistency between the datasets when the LASSO cases used are not overlapping with HI-SCALE aircraft observations according to lines 67-69: "The LASSO simulations analyzed in this study correspond to different cases aside from those sampled during HI-SCALE and are not directly constrained by the HI-SCALE aircraft observations."

Author response: We thank the referee for pointing this out. We agree that the original sentence could be misleading because the HI-SCALE observational data were not used for a direct comparison with the LASSO simulations. Only the latitude and longitude coordinates of the aircraft flight tracks were used as a reference for constructing the aircraft-like sampling trajectories within the LASSO domains. We have therefore removed the sentence that implied consistency between the simulated and observational datasets and clarified the role of the HI-SCALE aircraft data in the revised manuscript.

Section 2.2: Some of this information belongs in the introduction section when the campaign is discussed (or at least this section should be referred to over there).

Author response: We thank the reviewer for this suggestion. We have reorganized Section 2.2 to avoid repeating HI-SCALE campaign information that is now presented in the Introduction. The general campaign background, including the aircraft deployment, sampling location, horizontal coverage, and altitude range, is now described in the Introduction, while Section 2.2 focuses on the AAFMERGED dataset and its use in this study.

We clarified that only the aircraft latitude and longitude coordinates are used from the AAFMERGED dataset. These position data are used to reconstruct the HI-SCALE flight trajectories and characterize the aircraft sampling patterns used in the LASSO sampling analysis.

Line 150: Are these the only parameters used from the AAF dataset? If so, this should be specified much sooner. If not, this sentence should be edited.

Author response: We thank the referee for pointing this out. We confirm that the aircraft latitude and longitude are the only variables used from the AAFMERGED dataset in this study. The other measurements contained in the dataset are not used in the analyses represented here. We have clarified this in Section 2.2, where the AAFMERGED dataset is introduced, to make the scope of the aircraft data used in this study explicit.

Line 157: I'm not sure how CWC directly provides an estimate of cloud thickness, perhaps the initial part of the sentence can be updated to "CWC and its vertical profile/integral…"?

Author response: We thank the referee for pointing this out. We agree that CWC alone does not directly provide an estimate of cloud thickness. In this study, CWC is calculated from the cloud water mixing ratio and air density as CWC = ρq_c, while its vertical profile and vertical integral are used to characterize the vertical distribution and total amount of liquid water within the simulated clouds. We have revised the text accordingly and removed the statement that CWC directly diagnoses cloud thickness.

Line 162: I like this figure and the corresponding discussion. However, it is missing any reference to the height of the aircraft during these legs. Are these constant altitude legs or is the aircraft altitude changing? If so, how much? If only the lat-lon are used from these patterns and CWC is examined at multiple model levels, that should be mentioned.

Author response: We thank the referee for pointing this out. We agree that the original description did not explicitly clarify the role of aircraft altitude in Figure 1. The patterns shown in Figure 1 represent only the horizontal (latitude–longitude) component of the aircraft tracks and are used to characterize the horizontal sampling geometry; aircraft altitude is not represented in this figure. For the sampling analysis, these horizontal patterns are combined with the prescribed vertical sampling strategies to construct the three-dimensional trajectories through the LASSO domain. We have clarified this point in the Methods section.

Fig 1 and 2: The font sizes for the labels and legends need to be increased for accessibility.

Author response: We thank the referee for the suggestion. The font sizes of the labels and legends

in Figures 1 and 2 have been increased to improve readability and accessibility.

Line 200: I'm not sure if "flat" refers to the altitude being constant or something else? The term "flat profile" on Line 214 seems confusing.

Author response: We thank the referee for pointing this out. The two uses of "flat" referred to different concepts. At Line 200, "flat" refers to the model lower boundary, where terrain variability is not explicitly represented. We have revised the text accordingly. At the former Line 214, "flat profile" refers to a prescribed vertical sampling strategy consisting of horizontal flight legs at selected cloud-relative altitudes. To avoid ambiguity, we have revised the profile to "level-leg" and revised its description accordingly.

Line 227: Should it be "horizontal tracks"?

Author response: We thank the referee for pointing this out. The typo has been corrected.

Line 238: This detail needs to be described much sooner in the manuscript.

Author response: We thank the referee for pointing this out. We have moved this clarification to the beginning of Section 3, where the sampling framework is first introduced. The revised text now indicates that the aircraft position data are used as a reference for constructing the idealized sampling trajectories. We have also clarified at the beginning of Section 3 that the prescribed sampling profiles are defined in model time and applied continuously throughout the 6 h analysis period, rather than following the duration or timing of individual aircraft flights. The subsequent description retains the details of how the 10-min LASSO output is sampled at 1-s intervals along these trajectories.

Line 240: "As the LASSO simulations archive output at 10 min intervals, model fields were temporally expanded to 1 s resolution by repeating the corresponding 10 min output (mean fields) across each interval." I apologize but I'm struggling to understand how this works. A better description or schematic may be needed here as this is critical to the study. The subsequent discussion of sampling density as a function of aircraft speed then follows this confusion.

Author response: We thank the referee for pointing this out. We agree that the original description of the temporal sampling procedure was unclear and that this information is important for interpreting the subsequent analysis of spatial coverage and aircraft speed. We have moved the description of the sampling framework to the beginning of the Methods section and revised it to explain the procedure more explicitly. The LASSO model fields are available at 10 min intervals and are not temporally interpolated to 1 s resolution. Instead, each model field is held fixed over its corresponding 10 min interval while the prescribed aircraft-like trajectory advances through the domain at 1 s intervals. Model values are then sampled at each successive trajectory position within that interval. This clarification is provided before the discussion of the prescribed flight patterns and speed categories.

Line 245: This ties back to the need for using only the 6-12h model outputs as this sentence suggests that temporal constraint wouldn't be necessary.

Author response: We thank the referee for pointing this out. We agree that the original wording could imply that the sampling procedure was independent of the temporal constraints applied to the simulations. The sampling trajectories are applied only during the 6–12 h analysis window to exclude the initial 6 h spin-up period and to restrict the analysis to the same simulation day. We have revised the text to make this temporal constraint explicit.

Section 3.1 and 3.2: A lot of the text describes the sampling strategies – however one key element is missing – how much time does the "aircraft" spend in-cloud during these legs? Or what does the model cloud field need to look like along the "aircraft track" for the case to be used in this study? Is that considered (or should it be considered) during case selection or during flight strategy determination? I'm assuming this is accounted for when the authors remove instances of cloud edge, but it should be properly discussed here in the context of the flight tracks.

Author response: We thank the reviewer for pointing out this important aspect of the aircraft-like sampling design. In our framework, in-cloud sampling time was not used as an additional criterion for selecting LASSO cases. Specifically, cloud layers were diagnosed from the model CWC, using a threshold of 0.2 g m^{-3}, and the prescribed vertical sampling profiles were positioned relative to the diagnosed cloud base and cloud top. When the diagnosed cloud depth exceeded 200 m, 100 m offsets from cloud base and cloud top were applied to reduce sampling directly at cloud boundaries. For shallower cloud layers, these offsets were not applied. The actual in-cloud sampling time therefore depends on the prescribed trajectory and the evolving three-dimensional cloud field during the 6 h analysis period, rather than on a predefined minimum in-cloud duration or an additional case-selection filter.

Line 248: The first type of "profile" has major issues. Semantically speaking, this is not "flat" and the terminology is misleading as "flat profile" is an oxymoron. Additionally, this is an impossible flight track in the real world, judging by Fig. 3a, this implies the "aircraft" instantaneously rises or drops up to 500 m? While ascending/descending, an aircraft will still propagate horizontally, and any "flat" profile will look closer to profile types shown in panels (b) and (c) depending on the pitch angle. I understand the authors' intent here, but if they want "aircraft sampling representativeness", the profile in panel (a) is extremely unrealistic.

Author response: We thank the referee for this important comment. We agree that the original "flat profile" was misleading and that the instantaneous transitions between sampling levels would not represent a physically realistic aircraft trajectory. We have therefore redesigned this sampling strategy and renamed it the segmented level-leg profile.

In the revised design, the trajectory consists of three level legs located at fixed cloud-relative altitudes derived from the case-mean cloud base, cloud-middle, and cloud-top heights, respectively. The lower and upper sampling levels are offset by +100 m from the case-mean cloud base and −100 m from the case-mean cloud top, while the middle level is defined from the case-mean cloud-layer midpoint. These three altitudes remain fixed throughout the 6 h analysis period for each case rather than varying with the instantaneous cloud boundaries along the trajectory. Vertical transitions between the level legs are explicitly represented rather than being treated as

instantaneous changes in altitude. This design preserves the original intent of sampling representative vertical levels within the cloud layer while providing a more physically plausible trajectory that accounts for the finite time required for vertical transitions. Because the present framework samples model CWC directly, it does not account for instrument-specific response or equilibration times during altitude transitions. Such effects could further reduce the amount of usable observational data during ascent and descent and should be considered when applying this framework to individual aircraft instruments.

The revised level-leg trajectory has been used throughout the analysis, and the corresponding figures, calculations, and statistical results have been updated accordingly. The term "flat profile" has also been replaced throughout the manuscript.

Line 258: I would assume this is the case in most clouds even with extreme entrainment. I'm not sure if the authors meant to state this as a result derived from their figure?

Author response: We thank the referee for pointing this out. The statement was not intended to relate to entrainment. The June 9, 2015 case is used only as an illustrative example because the cloud distribution provides a clear view of the prescribed sampling trajectories. The case is used for visualization only and is not intended to represent a particular entrainment or cloud-process regime.

Line 262: "As the aircraft ascends or descends through a cloud, it samples a coherent cloud column over a short horizontal distance", I don't believe think this statement is accurate. The horizontal air speed of the aircraft would typically be greater than the ascent/descent rate. I would suggest the authors follow such arguments (throughout the manuscript) with an example from actual or representative horizontal and vertical speeds of the G1 aircraft from HI-SCALE.

Author response: We thank the referee for pointing this out. We agree that the original statement was misleading and we did not intend to imply that the aircraft remains within a coherent cloud column. The aircraft continues to travel horizontally while changing altitude, with horizontal motion occurring much more rapidly than vertical motion. We have removed the reference to sampling a "coherent cloud column" and revised the text to attribute the temporal clustering primarily to the sampling of different trajectory positions within the same 10-min LASSO model field.

Line 264: While the authors don't intend to compare exact profiles, this argument deserves the use of actual aircraft observations since those data are readily available.

Author response: We thank the referee for this suggestion. We agree that aircraft observations would be appropriate for evaluating the realism of simulated cloud properties for the model–observation comparison. However, the LASSO simulations analyzed in this study are not directly constrained by observations. A direct comparison of observed and simulated CWC would address a different question from the sampling representativeness analysis considered here. The HI-SCALE aircraft position data are used only to inform the design of realistic sampling trajectories, while the subsequent analysis evaluates how different prescribed sampling strategies represent the LASSO model fields. We have revised the corresponding discussion to avoid implying validation

of the simulated cloud structure against aircraft observations.

Figure 3: The figure is discussed already but even by Line 275, there is no description of the June 9, 2015 case that is being plotted. I'm not sure why or how that day is chosen or what it represents. I'm not entirely sure if this is an actual flight or model simulation and the tracks being based on a flight?

Author response: We thank the referee for pointing this out. The June 9, 2015 case is a LASSO simulation selected as an illustrative example to show the prescribed sampling trajectories. It does not correspond to a specific HI-SCALE aircraft flight. The trajectories shown in Figure 3 are idealized aircraft-like sampling paths, with the HI-SCALE aircraft position data used only as a reference for the spatial sampling geometry. The corresponding text has been revised to clarify the case selection.

Section 3.3: This section represents the type of considerations to be made, and it is a nice addition. However, it feels incomplete because it ties partly to the temporal resolution and somewhat to the model domain size but does not relate the potential (mis)match with the model grid spacing. If the model grid spacing is 100 m and the aircraft speed is 100 m s-1, a temporal resolution of 1 s directly matches the aircraft measurements. But otherwise, there is a mismatch, and a proper, direct comparison becomes unlikely.

Author response: We thank the referee for this helpful point. We agree that the relationship between aircraft sampling distance and the 100-m LASSO horizontal grid spacing should be clarified. In the revised Section 3.3, we now explicitly note that at 1 s sampling interval, aircraft speed determines the horizontal distance between consecutive sampling positions. A speed of 100 m s^{-1} corresponds to approximately one LASSO grid spacing per sampling interval, while slower or faster speeds result in sampling distances smaller or larger than one grid spacing, respectively. This does not imply that aircraft samples must correspond one-to-one with individual model grid cells; rather, it determines the spatial separation and coverage relative to the model resolution. We have added this discussion to clarify how aircraft speed relates to the spatial sampling characteristics evaluated in this study.

Line 301: While it is fine to construct a flight track that crosses regions of enhanced CF, this would be improbably to replicate in real life during an aircraft campaign. How would flight planning work to ensure enhanced CF along the track unless the forecasts were highly accurate or CF was high across the domain? "Aircraft sampling representativeness" needs to be thought of as a real-time scenario. Otherwise, the authors are creating an artificial selection bias by selecting tracks based on known model output. Would it not be more strictly representative if the flight tracks were randomly assigned to model case days? I'm not sure how one answers this question but I'm curious to see the authors comments.

Author response: We thank the referee for raising this point. We agree that the original wording could imply that the sampling trajectories were designed using prior knowledge of the simulated CF distribution. The horizontal sampling patterns were prescribed independently of the CF field and were not adjusted to target regions of enhanced cloud fraction. When applied to each LASSO case, the trajectories encounter regions of higher or lower CF according to the spatial distribution

of the simulated cloud field rather than through preferential track selection.

We have revised the discussion of Figure 4. We appreciate the referee's suggestion regarding random assignment of flight tracks to model cases. Our objective here is to compare a common set of prescribed sampling strategies across the analyzed LASSO cases rather than to reproduce case-specific real-time flight-planning. Applying the same sampling framework across cases without adjusting the trajectory based on the simulated CF field provides a consistent basis for this comparison.

Line 315: Is the comparison with domain means conducted because those are the values typically used for model-observational comparisons? The discussion would benefit from some references or examples here to illustrate why this specific comparison is conducted.

Author response: We thank the referee for this helpful suggestion. We agree that the rationale for comparing the trajectory-derived CWC with the domain mean was not sufficiently explained. In our analysis, the domain mean is not intended to represent an observational truth or a standard model–observation reference. Rather, it provides a domain-scale reference for quantifying how representative the spatially limited aircraft-like sampling is of the simulated cloud field. Because the prescribed trajectories sample only specific portions of the three-dimensional domain, the corresponding domain mean is calculated over the same vertical layers sampled by each trajectory. This provides a consistent comparison between the trajectory-derived CWC and the modeled domain-scale CWC within the sampled vertical range.

We have revised this section to clarify the motivation for this comparison and the calculation of the vertically matched domain mean. We have also added references discussing sampling representativeness and the effects of spatial sampling differences in model–observation comparisons.

Figure 5: This is an interesting figure.
On visual inspection, it seems the height increases generally as CWC over output domain increases, but this isn't necessarily true for CWC over aircraft track? (i.e., colors get lighter going toward the right but not going toward the top?). Do the authors have any thoughts or quantitative assessment of this discrepancy (if it even is so?)
I'm not sure why the units of kg m-3 were chosen? It would be better to show CWC in g m-3.

Author response: We thank the referee for this insightful observation. An additional quantitative assessment using Spearman's rank correlation coefficient was performed. The domain-mean CWC shows a moderate positive correlation with sampling height (r = 0.609). In contrast, the correlations between sampling height and trajectory-averaged CWC vary substantially across sampling strategies and aircraft speeds, ranging from weakly positive to near-zero or slight negative values. This indicates that the vertical dependence in the domain-mean cloud field is not consistently represented by individual aircraft-like trajectories, which are also influenced by horizontal cloud heterogeneity and the spatial sampling imposed by different flight trajectories. We have revised the discussion of Figure 5 accordingly.

We have also revised the CWC units in Figure 5 from kg m^{-3} to g m^{-3}, which provides a more

interpretable scale for the CWC values presented.

Figure 6: This is a very nice way of succinctly providing essentially a sensitivity test for different sampling profiles and aircraft speeds!

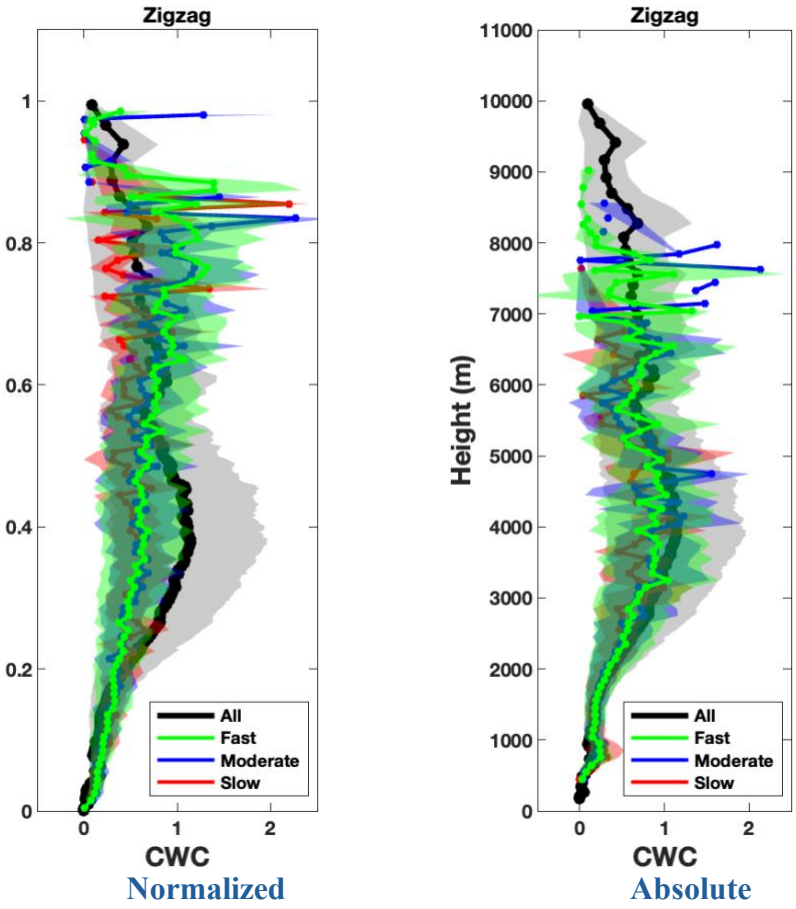
Table 1: The average and std. dev. Values for the model and aircraft sampled CWC should be provided for each scenario.

Author response: We thank the referee for the suggestion. We have revised Table 1 to include the mean and standard deviation of CWC for both the model-domain and the aircraft-sampled values for each sampling scenario.

Figure 8 – with the amount of noise in the CWC values as a function of altitude, I'm wondering if the authors tried looking at CWC as a function of height above cloud base or normalized height above cloud base? Where the latter is defined as Z – CTH / CTH – CBH. Surely the advantage of using model data is that they can retrieve cloud base heights for all "aircraft" transects even if the aircraft did not sample the cloud base. This is one major limiting factor in examining CWC profiles when looking at real-world aircraft observations. I suspect their correlations will drastically improve.

Author response: We thank the referee for this suggestion. We have performed an additional analysis in which CWC is expressed as a function of normalized cloud height, defined as $z^* = (z - CBH) / (CTH - CBH)$. The normalization replaces clouds of different depths and absolute altitudes on a common relative vertical coordinate.

We use the zigzag profile as an example. The correlations between the aircraft-sampled and domain-mean CWC increase from approximately 0.32 – 0.50 using absolute altitude to 0.42 – 0.57 using normalized cloud height across the three speed categories. However, the normalized profiles do not show a consistent improvement in bias or RMSE. These results indicate that normalization improves the correspondence in vertical structure, while differences associated with horizontal cloud heterogeneity and trajectory sampling remain. Given that the normalized-height analysis does not change the main conclusions of the study, we have retained the absolute-height presentation in Figure 8.



Normalized Absolute

Line 436: I don't mean to sound harsh or trivialize the authors' work, but it seems a lot more could have been done here with the model and "flight track" setup. Did the authors look to examine other cloud properties? One gets the feeling there's just not enough analysis here yet to draw meaningful conclusions. I'm hesitant to recommend publication of a revised manuscript without more detailed assessment of the dataset. I recommend investigating liquid water path and adding process rates that aircraft observations can lack. The authors have an opportunity to answer questions such as "what are aircraft observations missing" and "how can models inform better aircraft sampling strategies", etc. The lack of any real aircraft comparisons or such insights essentially boils the work down to comparisons between model domain average CWC and CWC simulated along a few different categories of "line tracks" within the model.

Author response: We thank the reviewer for this thoughtful comment. We agree that the LASSO framework provides opportunities to examine additional cloud properties and model-derived process rates. The objective of the present study, however, is more narrowly focused on quantifying how aircraft-like sampling geometry and spatial coverage affect the representativeness of CWC relative to the LES domain.

CWC was selected as the primary variable because it provides a direct measure of cloud condensate that can be consistently sampled along the prescribed aircraft-like trajectories and compared with the corresponding model-domain mean. The revised manuscript extends the analysis beyond the original trajectory–domain comparison by examining cloud-base and cloud-

top representativeness, vertical CWC profiles, sensitivity to aircraft speed and trajectory geometry, and differences among cloud-fraction regimes. These analyses are intended to identify which aspects of the simulated cloud field are more or less reliably represented by limited aircraft-like sampling.

Additional properties such as LWP and model-derived process rates could provide further insight into what aircraft sampling may miss and how models could inform future sampling design. Such extensions would require additional variable-specific definitions and sampling diagnostics and would broaden the study beyond the CWC-based representativeness framework developed here. We therefore identify these as important directions for future work rather than incorporating them into the present analysis. We have revised the discussion to clarify the scope and to emphasize the implications of the current results for aircraft sampling design.

Line 463: "Additional factors such as trajectory scale and orientation relative to the mean wind may also modulate sampling performance through their interaction with cloud morphology." Similar to my previous comment – the authors should have these data from the model simulation and this can be examined instead of having a speculative statement.

Author response: We thank the referee for pointing this out. We agree that trajectory scale and orientation relative to the mean wind should not be presented as established controls without a dedicated sensitivity analysis. These factors were not independently varied in the present experimental design, which was designed to examine horizontal trajectory geometry, vertical sampling strategy, and aircraft speed. The current set of trajectories therefore does not allow the effects of trajectory scale or wind-relative orientation to be isolated quantitatively. We have revised the text to identify these factors explicitly as limitations of the present experimental design rather than as demonstrated effects.

Response to Referee #2

This manuscript uses LASSO large-eddy simulations over the ARM SGP site to ask how well aircraft-like sampling reproduces domain-average cloud water content (CWC). Synthetic trajectories built from five horizontal patterns (four derived from HI-SCALE flight tracks, plus a zigzag), four vertical profiles, and three speed classes are flown through the model fields, and the sampled CWC is compared with layer-restricted domain means through regression slopes, correlations, and normalized error metrics. The authors conclude that sampling representativeness is governed mainly by vertical trajectory design and flight speed, with faster flight generally improving the agreement with the layer-restricted domain mean and with horizontal geometry playing a secondary role.

My questions center on interpretation and reporting rather than on the design itself. Most of them trace to one difficulty: as written, it is hard to tell how much of the headline findings reflects the physics of aircraft sampling and how much follows from choices made in constructing the emulation. I recommend minor revision. Reviewer 1 has already covered the realism of the flight profiles, the organization of the text, and the case for deeper analysis; I have tried not to repeat those points and note below where I second them.

Main Comments

Sect. 2.1.3 says 'a subset of the simulation library has been used in this study to make the processing manageable' (L118). Table S1 in the Supplement does list the analyzed case days, but the main text never states this number, and I could not find the criteria used to select the subset or any note on whether the selection could influence the results; the only rationale given concerns the choice among forcing options rather than the choice of cases. Could the authors state the case and ensemble-member counts in Sect. 2.1.3, explain how the subset was chosen, and comment on whether the results are expected to be sensitive to that choice?

Author response: We thank the referee for pointing this out. The analysis includes 91 LASSO case days available in the dataset used for this study. Although the documented SGP shallow-convection library contains 95 case dates, four case dates were not available in the dataset accessed for the present analysis. No case days were excluded based on cloud properties or sampling performance. For each case, one ensemble member with the WRF–Thompson configuration, IFS-DDH forcing at the 114 km scale, and VARANAL surface fluxes was analyzed.

We have revised Section 2.1.3 to state the number of analyzed case days and to distinguish the availability of case days from the selection of the simulation configuration. The 91 available case days were all retained, and no additional case-selection criterion based on cloud conditions or sampling performance was applied.

Two related details would also help: are the 2015–2016 cases taken from Alpha2 or from Version 1, and could the Version 1 domain size (25.5 km, currently mentioned only in passing at L181–182) be stated in Sect. 2.1.2? On the sampling parameters, Sect. 3.3 gives characteristic speed ranges for the three aircraft classes (L276–283), but unless I missed it, the specific airspeed values used to construct the slow, moderate, and fast trajectories are never stated; please

provide them.

Author response: We thank the referee for identifying these points and agree that these details should be stated more explicitly in the Methods. We have revised Sections 2.1.1, 2.1.2, and 3.3 accordingly. The 2015–2016 cases were obtained from Alpha2, whereas the 2017–2019 cases were obtained from Version 1. We have also added the Version 1 horizontal output domain size and 100 m horizontal grid spacing to section 2.1.2.

We also clarified the specific speeds used to construct the trajectories. The slow, moderate, and fast trajectories use nominal speeds of approximately 33, 100, and 200 m s^{-1}, respectively. The moderate speed is based on the characteristic G-1 research speed of approximately 100 m s^{-1} (Schmid et al., 2014), while the slow and fast trajectories correspond to approximately one-third and twice this reference speed, respectively. These values and their relationship to the 100 m LASSO horizontal grid spacing have now been clarified in Section 3.3.

Please also define the threshold on qc (or CWC) that identifies cloud base and cloud top along the track, state how the domain-mean cloud top and base in Fig. 7 are computed (over cloudy columns only?), and clarify what happens to the ±100 m offsets when the local cloud depth is under 200 m, in which case base +100 m would lie above top −100 m. Extending Table S1 with the ensemble members and the release for each case would cover much of this.

Author response: We thank the referee for pointing out these details. The cloud base and cloud top are identified using a CWC threshold of 0.2 g m^{-3}. For each cloudy model column, cloud base and cloud top are defined as the lowest and highest model levels exceeding this threshold. Along the aircraft-like trajectories, the corresponding sampled cloud-base and cloud-top heights are diagnosed from the cloudy columns intersected by the trajectory.

For Figure 7, the cloud-base and cloud-top heights of the domain are calculated from all cloud columns using the same CWC threshold, and the resulting column values are averaged to obtain the domain-mean cloud-base and cloud-top heights. Cloud-free columns are excluded from this.

The ±100 m offsets are applied only when the diagnosed cloud depth exceeds 200 m. For cloud depths of 200 m or less, these offsets are not applied as the two bounds would overlap. We have clarified these points in the revised Methods section.

The description at L240–243, where each 10-min output is repeated at 1-s resolution, together with the fixed 6-h window sampled at 1 Hz (L290–292), left me unsure what actually differs between the speed categories. If every trajectory collects one sample per second for six hours, the sample count is the same at every speed; what increases with speed is the distance covered, i.e. the fraction of the domain seen. So when the results attribute the improvement to 'sampling density' (L341, L389–390), is it not more accurate to say that a faster aircraft covers more of a snapshot that is frozen for 10 minutes at a time, which makes convergence toward that snapshot's domain mean almost guaranteed? I second Reviewer 1's request for a clearer description of this expansion step. Beyond the description, I would suggest tuning down the stated benefit of the higher speed categories. What limits fast platforms in reality is that the cloud field evolves while the aircraft transits, with shallow cumulus turning over on timescales of tens of minutes, and that

effect cannot operate in this emulation.

Author response: We thank the referee for this important clarification. We agree that the original use of "sampling density" was imprecise. All three speed categories are sampled at the same 1 s interval over the same 6 h analysis period and therefore contain the same nominal number of trajectory positions. Increasing the prescribed speed does not increase the temporal sampling frequency or sample count. Instead, it increases the distance traveled between consecutive samples and the spatial extent of the model domain. We have revised the manuscript accordingly and avoid attributing the improved agreement at higher speeds to increased "sampling density".

We have also expanded the description of the temporal matching procedure. The LASSO fields are available at 10-min intervals and are not temporally interpolated to 1 s resolution. Instead, each 10 min model field is held fixed for the corresponding 600 1-s trajectory positions. Thus, within each 10 min interval, faster prescribed speeds allow the trajectory to traverse a larger portion of the same model field.

We agree that this treatment can favor increased agreement with the domain mean at higher prescribed speeds because a larger spatial extent of each fixed model state is sampled. It does not represent cloud evolution within the 10 min interval and therefore does not capture the competing effects that would occur during real aircraft sampling of an evolving shallow-cumulus field. We have revised the interpretation of the speed dependence throughout the manuscript to clarify that the observed improvement applies only within the present idealized sampling framework and should not be interpreted as a general advantage of fast aircraft platforms under real atmospheric conditions.

One more question here: L361–362 gives zigzag sample counts rising from 19410 (slow) to 27617 (fast). At 1 Hz over a fixed 6 h that cannot be the count for a single trajectory. Could the authors state what is being counted (in-cloud samples only? Or pooled over cases?)?

Author response: We thank the referee for pointing this out. The values of 19410 and 27617 refer to the total number of valid CWC samples pooled across all analyzed cases for the corresponding zigzag speed category. Each individual trajectory contains the same number of 1 s sampling positions over the 6 h analysis period. Only valid CWC samples are included in the data number counts. We have revised the text to clarify this distinction.

In Table 1, every nBias entry is positive: 0.49–0.67 for the flat, sine wave, and staircase patterns, up to 1.27 for zigzag, and 1.38–1.51 for the flat pattern in the profile comparison of Sect. 4.3. The text calls this 'limited systematic deviation' (L354), but a sampled average running 50–150% above the reference would be a substantial systematic offset. Furthermore, the vertical profiles are selected to stay between cloud base +100 m and cloud top −100 m, the example track is 'constructed to deliberately cross regions of enhanced CF' (L301–302), and the reference is a layer mean over the full domain including clear air, so the trajectories conditionally sample cloudy air while the reference does not. Therefore, I echo Reviewer 1's question about why the all-sky domain mean is the appropriate comparison target, and would take it one step further. Could the authors compute the same metrics against an in-cloud conditional domain mean over the same layers? All required fields appear to be in hand, and the comparison would separate the

cloud-conditional offset from genuine trajectory effects. It would also show whether the bias mainly comes from comparing in-cloud samples with all-sky averages. If it does, no flight strategy can remove it, and the remedy lies in how the comparison statistic is constructed rather than in how the aircraft flies.

Author response: We thank the referee for raising this point and apologize for the confusion caused by the original presentation. The nBias and nRMSE row labels were inadvertently interchanged in the original Table 1 and the underlying calculations were not affected. In the corrected table, nBias for the level-leg, sine wave, and staircase profiles ranges from approximately -0.17 to -0.01, whereas nRMSE ranges from approximately 0.45 to 0.66. The zigzag profile shows nBias values from –0.13 to 0.20 and nRMSE values from 0.76 to 1.27. These corrected values do not indicate the large systematic positive offsets implied by the original table. Table 1 and the corresponding discussion have been revised accordingly.

We also clarify that the trajectory was not constructed to deliberately cross regions of enhanced CF. The horizontal trajectory was defined independently of the simulated CF distribution and could encounter regions of either relatively high or low CF as it traverses the domain.

How is the quantity labeled nRMSE in Table 1 calculated? Most entries are negative (e.g. −0.11), which no standard RMSE definition allows, yet the same metric is positive in Sect. 4.3 (1.32–2.05). Could you please give the formulas for nBias and nRMSE, including the normalization, and apply them consistently between the two sections? Is the identical zigzag moderate and fast column (0.37, 0.20, 1.27 in all three metrics, to two decimals) a copy error, or do the two speeds genuinely produce the same statistics despite different sample counts?

Author response: We thank the referee for identifying this inconsistency. The negative values labeled as nRMSE in the original Table 1 resulted from an inadvertent interchange of the nBias and nRMSE row labels during preparation of the table. The underlying calculations were not affected.

The nBias and nRMSE are calculated as

$$\mathrm{nBias} = \frac{\overline{CWC_{\mathrm{sampled}}} - \overline{CWC_{\mathrm{domain}}}}{\overline{CWC_{\mathrm{domian}}}}$$

and

$$\mathrm{nRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(CWC_{\mathrm{samped},i} - CWC_{domain,i}\right)^2}}{\overline{CWC_{\mathrm{domain}}}}$$

The identical moderate- and fast-speed values originally reported for the zigzag profile were also a table transcription error. The corrected moderate-speed values are r = 0.74, nBias = 0.02, and nRMSE = 0.76, whereas the fast-speed values are r = 0.37, nBias = 0.20, and nRMSE = 1.27. Table 1 and the corresponding discussion have been corrected accordingly.

Could the correlation values quoted in Sect. 4.1 also be checked against Table 1? The text says r

ranges 'from about 0.69 to 0.88' (L348–349), while Table 1 contains several values of 0.60–0.66, and 'the lowest correlation occurs for the diagonal pattern at slow speed ($r \approx 0.72$–0.73)' (L353) does not match the table either (diagonal-slow is 0.74 for flat and 0.60 for sine wave and staircase). Last, L408–409 attaches a single p-value ($p = 0.032$) to three different correlation coefficients. Can the authors confirm what this p-value refers to, report the three values separately, and note how the effective sample size was handled given that the profile points are adjacent 30-m model levels with strong vertical autocorrelation?

Author response: We thank the referee for identifying these inconsistencies. We have rechecked all correlation values reported in Section 4.1 against the corrected Table 1 and revised the text accordingly. For the level-leg, sine wave, and staircase profiles, the correlations range from 0.57 to 0.83 across the horizontal patterns and the speed categories. The previous statements were corrected.

We have also rechecked the p-values associated with the vertical-profile correlations. The values differ among the three speed categories and should not have been represented by a single p-value. More importantly, the vertical-profile points correspond to adjacent model levels and are therefore not statistically independent because of vertical autocorrelation. Because no correction for effective sample size was applied, these nominal p-values cannot be interpreted as formal tests of statistical significance. We have therefore removed the p-values and the associated significance statements from the revised manuscript and now use the correlation coefficients as descriptive measures of agreement between the sampled and domain-mean vertical profiles.

Minor Comments

L73–74 states the study aims to quantify sampling of 'key cloud and aerosol properties', but only CWC is analyzed; other cloud and aerosol properties are deferred to future work (L481–483). I echo Reviewer 1's comment on the title here: the aims statement should match the delivered scope.

Author response: We agree with the referee that the aims statement should reflect the scope of the analysis presented in this study. We have therefore revised the corresponding text to specify that the study evaluates the representativeness of aircraft sampling strategies for CWC in LASSO simulations, rather than broadly referring to cloud and aerosol properties. Consistent with this change and Reviewer 1's suggestion, we have also revised the title.

Could an explicit formula for CWC be given in Sect. 2.3 (CWC = ρqc?). And does the CWC sampled from the model include only cloud water, or also mass that the Thompson scheme classifies as rain? Since in-situ probes on a real aircraft would count drizzle-sized drops as part of the measured liquid water, whether qr is included affects how the model-sampled CWC maps onto what an aircraft would report (e.g., King Probe), and one clarifying sentence would help.

Author response: We thank the referee for this helpful suggestion. We have added the explicit definition $CWC = \rho q_c$ in Section 2.3, where q_c is the model cloud water mixing ratio and ρ is air density. The CWC used in this study includes only the cloud-water category of the Thompson microphysics scheme and does not include the separate rainwater mixing ratio. The model-derived

CWC represents cloud-water mass concentration rather than total liquid water that may be detected by an in situ aircraft probe across both cloud and drizzle-sized droplets.

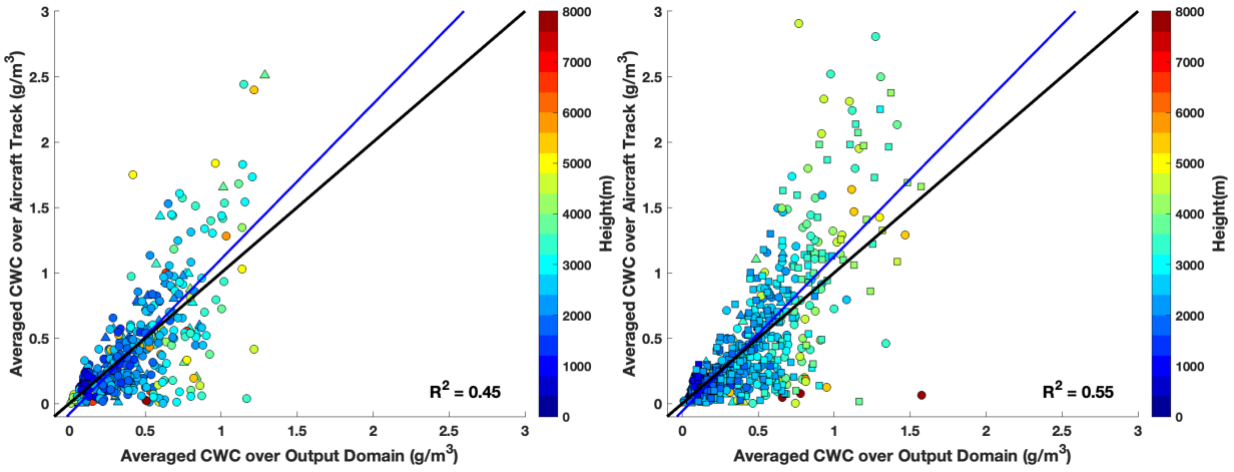
Sect. 2.2 gives the horizontal coverage of the aircraft sampling as around 100 km (L143–144), yet the LES domains are only 14.4 and 25.5 km across. Since the sampling paths are idealized, how were they derived from the observed latitude–longitude tracks, if at all: truncated to the domain, rescaled to fit it, or redrawn from scratch? This affects how literally the patterns in Fig. 1b–f can be read as HI-SCALE geometries. Please clarify.

Author response: We thank the referee for pointing out that the spatial transformation of the observed flight tracks was not sufficiently described. The observed HI-SCALE tracks extend over a substantially larger horizontal area than the LASSO LES domains and were therefore not applied at their original spatial scale. Instead, the latitude–longitude patterns were spatially scaled to the corresponding model-domain dimensions while preserving their characteristic horizontal geometries. Although the doubly periodic lateral boundaries of the LASSO domains would allow the original-scale trajectories to wrap repeatedly through the domain, this approach was not used in the present analysis. The tracks were neither truncated to the model domain nor reconstructed from scratch. We have clarified this in Section 2.2 and stated that the patterns retain the characteristic HI-SCALE geometries but not their original horizontal spatial scale.

L181–182 notes that horizontal sampling is performed separately for the two domain sizes. Are the statistics in Table 1 and Fig. 6 then pooled across both domains (and both LASSO releases)? Representing a 25.5 km domain from a localized track is presumably harder than a 14.4 km one, so I wonder whether domain size could confound the pattern and speed comparisons; a sentence on this, or numbers stratified by domain, would help.

Author response: We thank the referee for raising this point. The sampling trajectories were constructed and applied separately for the two LASSO domain sizes, with the horizontal patterns scaled to the corresponding model domain before sampling. The statistics presented in Table 1 and Figure 6 are subsequently pooled across the analyzed cases from both LASSO releases.

To examine whether domain size could influence the results, we additionally stratified the analysis by domain size, using the 2015–2016 cases with the 14.4 × 14.4 km domain and the 2017 cases with the 25 × 25 km domain. For example, for the moderate-speed zigzag configuration, the 2015–2016 cases (18 cases) yield r = 0.67, nBias = -0.01, and nRMSE = 0.79, compared with r = 0.74, nBias = 0.04, and nRMSE = 0.76 for the 2017 cases (30 cases). The corresponding domain-mean CWC values are also very similar, at 0.412 g m^{-3} and 0.411 g m^{-3}, respectively. These results do not indicate a systematic reduction in sampling representativeness for the larger domain in this comparison. Because domain size is also associated with different years and LASSO releases, however, its influence cannot be isolated independently from other case-to-case differences.



2015&2016 2017

Could the regression intercept be reported alongside the slope for Fig. 5 and the configurations in Fig. 6? Unless I missed it, neither the text nor the figures give the intercept, and with slopes near unity and uniformly positive nBias the offset must reside there, so reporting it would connect Figs. 5–6 with Table 1. Please also state whether the scatter points come from all cases and ensemble members.

Author response: We thank the referee for this suggestion. The positive nBias values in the original Table 1 resulted from the inadvertent interchange of the nBias and nRMSE row labels; the corrected nBias values are generally small and are not uniformly positive. The large systematic offset inferred from the original table therefore does not occur in the corrected results.

We calculated and examined the regression intercepts for all configurations. The intercepts are generally small and do not indicate a consistent systematic offset across the sampling configurations. For the level-leg configurations, the intercepts are approximately 0.03–0.05 g m^{-3}, while those for the sine wave and staircase configurations are generally close to zero, approximately –0.04 to 0.01 and –0.03 to 0.02, g m^{-3}, respectively. For the zigzag, the intercepts are approximately 0, –0.05, and 0.14 g m^{-3} for the slow, moderate, and fast speed categories, respectively. The nBias values already quantify the systematic differences between sampled and domain-mean CWC, and the intercept analysis does not reveal a consistent additional offset across configurations.

For each trajectory configuration, paired sampled and domain-mean CWC values are calculated separately for each 10-min model-output interval within the 6 h analysis period. Each analyzed case therefore contributes up to 36 paired values to the scatterplot. The regression is calculated using these paired values across all analyzed cases. For each case, only the selected LASSO ensemble member with the consistent model configuration is used; the remaining ensemble members are not included. We have clarified this in the revised manuscript.

In the paragraph on Fig. 6 (L342–346), it is not always clear which vertical profile the horizontal-pattern statements refer to; for example, 'for the circle horizontal pattern, the slope value is greater than 0.94' presumably applies to the staircase row only. Could each quoted number be anchored to its row in the figure?

Author response: We thank the referee for pointing this out. We agree that the original discussion did not consistently identify the vertical sampling profile associated with each quoted slope value. We have revised the paragraph describing Figure 6 so that each numerical value and comparison is explicitly associated with the corresponding vertical profile and horizontal sampling pattern.

The case grouping by cloud fraction (L457–464) appears for the first time in the conclusion, as a finding ('no systematic dependence of sampling performance on cloud amount'), with the supporting material only in Fig. S2. Since Fig. S2 shows scatter plots without per-category metrics, could this analysis be presented in Sect. 4, with slope or r for each CF bin? The statement that differences among CF categories 'primarily reflect variations in the number of available cases' is also hard to assess without those case counts.

Author response: We thank the referee for this suggestion. We agree that the cloud-fraction-stratified analysis should be presented in the Results. We have therefore added this analysis to Section 4 and report regression statistics separately for the four CF categories (0–25%, 25–50%, 50–75%, and 75–100%). The stratified results show variability among CF categories, but no consistent monotonic dependence of sampling performance on CF across the different trajectory configurations.

For example, for the level-leg and circle configuration, r ranges from approximately 0.76–0.86, 0.43–0.68, 0.38–0.74, and 0.61–0.82 for the four successive CF categories, respectively, depending on speed. The corresponding slopes likewise vary among CF categories and configurations rather than changing consistently with increasing CF. We have added the number of cases in each CF category: 39 cases for 0–25%, 42 cases for 25–50%, 7 cases for 50–75%, and 3 cases for 75–100%. The substantially smaller sample sizes in the two highest-CF categories are now noted when interpreting the stratified statistics.

Technical Comments

At L279–281 the C-130 appears twice in the moderate-speed category ('C-130/LC-130' and 'Lockheed C-130'). And is the BAe-146 correctly placed in the 200–250 m s−1 class? Please correct me if I am wrong: from what I can find online, its indicated airspeed during sampling is about 100 m s−1, corresponding to a true airspeed of roughly 100–120 m s−1 (Sanchez-Marroquin et al., 2019, https://doi.org/10.5194/amt-12-5741-2019), which would place it with the moderate-speed platforms.

Author response: We thank the referee for identifying these issues. The duplicate reference to the Lockheed C-130 has been removed. We also rechecked the classification of the BAe-146 and agree that it is more appropriately placed in the moderate-speed category for atmospheric sampling. We have therefore moved the BAe-146 from the high-speed to the moderate-speed category and added the suggested reference. The high-speed category now includes the NASA DC-8 and WB-57 as representative examples.

The text at L403–404 says Fig. 8 shows the 'flat, sine wave, and zigzag' vertical patterns; the caption says 'flat, sine wave, and staircase'. The following paragraphs then discuss all four.

Please reconcile which patterns are plotted.

Author response: We thank the referee for identifying this inconsistency. Figure 8 includes all four vertical sampling patterns: level-leg, sine wave, staircase, and zigzag. The original text inadvertently omitted the staircase pattern, while the figure caption omitted the zigzag pattern. We have corrected both the text and caption so that all four vertical sampling patterns are identified consistently.