The University of Manchester
Simon Building
Oxford Road
Manchester M13 9PL

Dear Editor,

Thank you for the opportunity to revise our manuscript, “Secondary ice production affects tropical convective clouds under different aerosol conditions.”

We have carefully considered Reviewer 2’s continuing concern that the manuscript does not establish a single, observationally constrained control simulation before interpreting the sensitivity experiments. We agree that the observations should be used more clearly to evaluate model performance, identify model limitations, and provide context for interpreting the effects of secondary ice production (SIP).

Two distinct questions. The reviewer’s concern brings together two questions that we believe should be distinguished: (1) how well the simulations reproduce the available observations, and (2) what effect SIP has when diagnosed from controlled comparisons. These serve different purposes. The observations are used to assess the realism and limitations of each simulation individually. The effect of SIP, by contrast, is diagnosed using matched pairs of simulations in which the prescribed droplet number, N_d, is held fixed. Thus, noSIP-200 is the matched control for allSIP-200, noSIP-400 for allSIP-400, and noSIP-800 for allSIP-800. In each comparison, only the SIP treatment changes.

Using noSIP-400 as the sole control for experiments run at other N_d values would change both the SIP treatment and the prescribed droplet number simultaneously. Any resulting difference could not therefore be attributed uniquely to SIP. The matched-N_d comparisons avoid this ambiguity and provide the appropriate controls for isolating the SIP response.

$N_d = 400$ cm^{-3} is not a uniquely constrained value. In the earlier study by Connolly et al., aerosol size and composition measured by the Dornier 228 in Hector’s inflow region were used to infer CCN activation and corresponding droplet concentrations as a function of updraught velocity. The inferred N_d therefore depends strongly on updraught, while entrainment and spatial variability introduce further uncertainty. The earlier analysis also used a different modelling and microphysical framework. We therefore regard 400 cm^{-3} as a plausible intermediate value, rather than as a uniquely known, a priori control condition for the present simulations. Accordingly, we retain 200, 400 and 800 cm^{-3} as plausible low, intermediate and high prescribed droplet-number conditions, each with its own matched noSIP control.

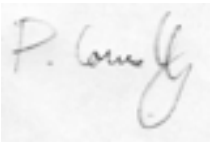
We are also cautious about defining a unique control by tuning the noSIP configuration to obtain the closest agreement with observations. If SIP represents a missing physical process, adjusting other aspects of a noSIP simulation to reproduce the observed state could compensate for its absence. Apparent agreement could then result from offsetting errors, for example in droplet number, primary ice production, aggregation, sedimentation or raindrop freezing. We therefore consider it more informative to retain the sensitivity framework and ask whether including SIP moves each matched simulation towards or away from the observed state, rather than selecting or tuning a single “best” noSIP control after the fact.

Strengthened observational evaluation. To address the reviewer's underlying request for clearer observational grounding, we have substantially expanded this evaluation. MTSAT-derived domain-averaged outgoing longwave radiation is now included directly in Figure 4a alongside all six allSIP and noSIP simulations. Egrett observations of ice particle number concentration and ice water content are now included in Figures 7c and 7e alongside the full allSIP/noSIP sensitivity range, with the more detailed aircraft evaluation retained in Figure S1. Together with the existing comparisons against CPOL reflectivity and precipitation and the observed thermodynamic profile, these additions give a broader and more transparent assessment of how each simulation compares with the observations. We have also expanded the discussion of model uncertainties and revised the Conclusions to emphasise the need for future in situ measurements within the mixed-phase region.

We agree that consistency with the upper-anvil observations does not exclude compensating errors in the mixed-phase microphysics, and we have therefore avoided describing any simulation as fully validated. The aircraft observations constrain the simulated upper-level ice field, but the absence of corresponding measurements within the mixed-phase region limits how directly the simulated SIP response can be evaluated. We now state this distinction explicitly in both the Results and Conclusions.

We therefore respectfully ask the Editor to accept the matched-N_d control framework as the appropriate basis for diagnosing the SIP response. We have strengthened the observational evaluation and clarified the experimental logic, but we do not believe that using noSIP-400 as the sole control for experiments at other N_d values would be scientifically justified. Such comparisons would confound changes in SIP treatment with changes in prescribed droplet number, making attribution of the SIP effect less clear rather than more clear.

Yours sincerely,



Professor Paul Connolly
on behalf of all authors
Department of Earth and Environmental Sciences
The University of Manchester

Responses to reviews
"Secondary ice production affects tropical convective clouds under different aerosol conditions"
by Sun et al.

We thank both referees for their constructive comments. In response, we have further revised the manuscript to make the observational evaluation more explicit, clarify the experimental design, incorporate the aircraft observations into the main figures, expand the comparison with MTSAT-derived OLR, and strengthen the discussion of observational constraints and model uncertainties. The corresponding changes are highlighted in the revised manuscript.

Review #1

My comments have been addressed. However, I think that a very important point is that the validation of the modelling results against in situ aircraft data should be explicitly stated in the Conclusions section as a perspective of the study before the final publication of the paper. In my view, the last sentence of the Conclusion should be further developed with an explicit reference to in situ aircraft observations, in order to highlight the need to validate the ice particle number concentration produced by the model.

Response:

We thank the referee for this helpful suggestion. We agree that the need for further validation of the simulated ice particle number concentration should be stated explicitly in the Conclusions. We have therefore expanded the final paragraph of the Conclusions to acknowledge both the value and the limitations of the currently available aircraft observations and to identify case-matched in situ measurements in the mixed-phase region as an important priority for future work.

The revised text reads (lines 761-765):

"The available aircraft observations provide a useful constraint on upper anvil ice but do not directly constrain ice multiplication within the mixed-phase region. Future work should combine targeted modelling with co-located in situ aircraft measurements in the mixed-phase region, particularly of ice particle number concentration and particle size distributions, to evaluate the modelled ice-phase properties and constrain SIP responses across aerosol regimes."

Review #2

The authors have responded in detail to the referee comments, but I do not believe that the underlying scientific concern has been adequately addressed in the revised manuscript. My principal concern remains that the manuscript does not establish a clearly defined and observationally evaluated baseline ("control") simulation before interpreting the sensitivity experiments.

At present, comparisons with observations are interwoven with the sensitivity simulations. This makes it difficult to distinguish evaluation of the model from evaluation of the sensitivity tests. A more rigorous approach would be to identify a single control simulation (the Nd = 400

cm^{-3} case appears to be the most appropriate based on the available observations), evaluate

that simulation comprehensively against the available measurements, and only then present the sensitivity experiments as perturbations about this validated baseline. Comparisons with observations should not be interwoven with the sensitivity experiments.

This distinction is important because, if the model contains structural biases (for example in aggregation or other aspects of the ice microphysics), one of the sensitivity experiments may appear to agree better with observations simply through compensating errors rather than because it provides the most physically realistic representation of the cloud system. In the present manuscript, the sensitivity experiments span a very wide range of prescribed droplet concentrations, so they effectively provide multiple opportunities for individual experiments to compensate for an underlying model bias and consequently appear to agree better with observations. Without a clearly validated control simulation, the sensitivity experiments themselves begin to act as an implicit calibration, making it more difficult to interpret the physical significance of the simulated SIP responses.

I also remain unconvinced by several of the reasons given for not making greater use of the available observations.

The comparison of upper-tropospheric ice number concentration and ice water content (Figure S1) is, in fact, a useful and convincing validation of the simulated ice field and deserves to be presented in the main manuscript rather than in the Supplement. Although these aircraft observations do not sample the mixed-phase region directly, this is not a major limitation. The upper-tropospheric ice population is the cumulative result of the preceding convective ascent, and the amount of supercooled liquid transported aloft for subsequent homogeneous freezing is itself influenced by SIP processes operating lower in the cloud (e.g. Phillips and Patade, 2022, JAS). Consequently, these observations remain an important validation of the simulated ice microphysics.

Similarly, I do not find the arguments regarding the cloud droplet concentration measurements persuasive. Although the number of available Dornier aircraft samples is small and confined to levels below the freezing level, even a limited number of observations provides a useful constraint on the prescribed cloud droplet concentrations, which are central to the experimental design. The associated uncertainty can readily be quantified using confidence intervals or standard errors. The purpose is not to perform a statistically rigorous

validation, but to demonstrate that the selected control simulation is physically consistent with the observed environment.

Likewise, I am not persuaded by the justification for omitting cloud-fraction profiles. The fact that previous studies analysed cloud fraction for different periods does not preclude performing the same diagnostic for the present case. Cloud fraction is a standard model diagnostic and would provide valuable additional information on the realism of the simulated cloud structure.

Overall, I believe the manuscript would be substantially strengthened by introducing a dedicated model evaluation section centred on a single control simulation. This section should include the comparison presently shown in Figure S1 together with the existing radar and satellite evaluations and any additional available observational constraints. The observational evaluation should be clearly separated from the subsequent sensitivity experiments so that the SIP responses are interpreted relative to a validated baseline state rather than using the sensitivity experiments themselves as part of the model evaluation.

Finally, I note that the tracked-changes version indicates that only a very small number of sentences have been modified in the main manuscript during this revision. Consequently, despite the detailed point-by-point response, I do not consider that the manuscript itself has been revised sufficiently to address the substantive concerns raised in the previous review.

For these reasons, I recommend that the manuscript be reconsidered after major revision.

General response:

We thank the referee for restating the underlying scientific concern. We agree that the role of the observational comparisons should be clearly distinguished from the calculation of the SIP-induced differences. We have therefore substantially strengthened the observational evaluation in the main manuscript and clarified how the sensitivity experiments are interpreted.

In particular:

(1) MTSAT-derived domain-averaged OLR observations are now included directly in Figure 4a together with all six allSIP and noSIP simulations.

(2) Upper-level aircraft observations of ice particle number concentration and IWC are now included in Figures 7c and 7e together with the simulations at all three prescribed cloud droplet number concentrations (Nd).

(3) A new paragraph has been added to Section 3.3.1 to discuss the aircraft comparison and its observational scope.

(4) The detailed comparison between the aircraft observations and the allSIP-400 and noSIP-400 simulations, including model interquartile ranges and observational sample counts, is retained in Figure S1.

(5) The Conclusions now explicitly distinguish the constraint provided by the upper-anvil

aircraft observations from the need for future measurements in the mixed-phase region.

(6) The experimental design in Section 2.2 has been clarified to state explicitly that the effect of SIP is quantified using matched allSIP–noSIP comparisons at each prescribed Nd, rather than through comparison with a single control simulation.

However, we have not designated the Nd = 400 cm−3 simulation as a single control. The reasons for this decision, and our responses to the specific points, are given below.

1. Definition of a single control simulation

Response:

We appreciate the referee's thoughtful suggestion regarding a single baseline. However, the primary objective of our experimental design is to examine how the effect of SIP changes across different prescribed droplet concentrations representing different aerosol conditions. The combined SIP effect is therefore quantified at each Nd by comparing allSIP with noSIP while holding Nd fixed.

In the earlier study by Connolly et al., aerosol size and composition measured by the Dornier 228 in Hector's inflow region were used to estimate activated droplet concentrations over a range of updraught velocities. The inferred N_d therefore depends strongly on updraught, while entrainment and spatial variability introduce further uncertainty. The earlier analysis also used a different modelling and microphysical framework. We therefore regard $N_d = 400\ \mathrm{cm}^{-3}$ as a plausible intermediate value rather than a uniquely constrained, a priori control condition for the present simulations.

We also recognise that our previous wording may not have made the control structure sufficiently explicit. The study contains a matched control at each prescribed droplet number concentration. Thus, noSIP-200 is the matched control for allSIP-200, noSIP-400 for allSIP-400, and noSIP-800 for allSIP-800. In each comparison, only the SIP treatment changes.

Using noSIP-400 as the sole control for simulations conducted at $N_d = 200$ or $800\ \mathrm{cm}^{-3}$ would change both the SIP treatment and the prescribed droplet number simultaneously. Any resulting difference would therefore combine the effects of SIP and N_d, making attribution of the SIP response ambiguous. The matched-N_d controls avoid this problem and provide the appropriate basis for isolating SIP effects.

Furthermore, the better overall agreement of the Nd = 400 cm^{-3} simulations with the available radiation and precipitation observations is itself a result of the observational comparison. It was not an assumption used to define the experimental design. We therefore retain this agreement as a result rather than using it retrospectively to define Nd = 400 cm^{-3} as the control.

We are also cautious about defining a unique control by selecting or tuning the noSIP configuration that gives the closest agreement with observations. If SIP represents a missing physical process, apparent agreement in a noSIP simulation may result from compensating errors in other uncertain aspects of the microphysics, including droplet number concentration,

primary ice production, aggregation, sedimentation, or raindrop freezing. In that situation, introducing SIP could worsen the tuned agreement, not because SIP is physically inappropriate, but because its effects had already been compensated for elsewhere in the model. We therefore consider it more informative to ask whether enabling SIP moves each matched simulation towards or away from the observed state than to retrospectively select a single best-performing noSIP control.

We have clarified this experimental design in Section 2.2. The revised text (lines 194-203):

"For each Nd value, six SIP scenarios were tested: all SIP mechanisms active (allSIP), no SIP process enabled (noSIP), rime-splintering only (RS), and three combined schemes involving RS with one additional mechanism: RS with Mode 1 (RS+M1), RS with Mode 2 (RS+M2), and RS with ice–ice collisional breakup (RS+BR). For each prescribed N_d, the noSIP simulation provides the matched control for the corresponding SIP experiments. The matched comparison between allSIP and noSIP is used to quantify the combined effect of SIP, while the RS, RS+M1, RS+M2, and RS+BR simulations are used to examine the contributions and aerosol dependence of individual mechanisms. RS was retained as the reference SIP configuration because it is the default SIP process in CASIM. This design enables a systematic assessment of how aerosol conditions influence the impacts of SIP on cloud microphysics, radiation, and precipitation."

We have also clarified the basis for selecting the three prescribed Nd values (lines 189-193):

"These values span the range reported during ACTIVE, where cloud droplet number concentrations of approximately 150, 500, and 900 cm^{-3} were associated with the monsoon-break, pre-monsoon, and burning-period environments, respectively (Connolly et al., 2013). The three prescribed values therefore provide a consistent framework for examining how the SIP response varies across plausible aerosol conditions."

The observational comparisons are used to evaluate the model performance and remaining differences from the observations, whereas the SIP effect is calculated from matched differences between simulations at the same Nd. To make this distinction explicit in the manuscript, we have added the following statement at the beginning of Section 3 (lines 231-233):

"The observational comparisons presented below are used to evaluate the realism and limitations of each simulation individually. SIP effects are subsequently diagnosed from matched SIP–noSIP differences at fixed Nd."

This explicitly separates observational evaluation from the calculation of SIP-induced differences, while retaining the matched-pair experimental design required to examine the aerosol dependence of the SIP response.

2. Model uncertainties and possible compensating effects

Response:

We agree that model uncertainties can affect the agreement between individual simulations and observations. For this reason, we do not use the better agreement of allSIP-400 with

some observational diagnostics either to define a unique control simulation or to determine the SIP effect. The effect of SIP is quantified from the matched allSIP–noSIP differences at each Nd and is interpreted together with the diagnosed changes in cloud droplets, raindrops, ice crystals, graupel, IWC, and vertical velocity.

To make the issue of compensating errors explicit in the manuscript, we have added the following paragraph to Section 4.1 (lines 606-609):

"Selecting a noSIP simulation on the basis of observational agreement could introduce compensating-error tuning if SIP represents a physically important missing process. The present study therefore evaluates whether enabling SIP improves or worsens agreement across a plausible range of prescribed droplet-number conditions rather than calibrating a noSIP configuration before assessing SIP."

We have also retained and strengthened several explicit qualifications in the manuscript:

(1) The discussion of Figure 2 states that the simulated minimum OLR values remain higher than observed, indicating a remaining difference between the simulations and observations.

(2) The revised Figure 4a discussion states that all simulations underestimate the observed domain-averaged OLR before and after the mature stage, indicating remaining differences in the timing and evolution of the simulated convection.

(3) The Conclusions state that, although allSIP-400 shows the best agreement among the simulations, differences from the observations remain in both radiation and precipitation.

(4) Section 4.2 discusses uncertainties associated with the SIP parameterisations, heterogeneous ice nucleation, raindrop freezing, aggregation, sedimentation, and environmental conditions, and states that the magnitude of the simulated SIP contribution should be interpreted with caution.

(5) The aircraft comparison is explicitly described as a constraint on the upper-level anvil ice field rather than a direct observation of ice multiplication in the mixed-phase region.

Thus, the better agreement of allSIP-400 is presented as one result from the observational comparisons. It is not used to select a control simulation or as the basis for calculating the SIP response.

3. Upper-tropospheric aircraft observations and Figure S1

Response:

We agree with the referee that the upper-tropospheric aircraft observations provide a useful constraint on the simulated ice field, even though they do not directly sample the mixed-phase region. We have therefore incorporated these observations into the main manuscript.

Specifically, the Egrett observations of ice particle number concentration and IWC are now included in Figures 7c and 7e, respectively. These panels retain the allSIP and noSIP simulations at all three prescribed Nd values, allowing the observations to be compared with

the complete paired sensitivity experiments. Observational samples are grouped into 0.5 km height bins; black circles show the bin means, black horizontal bars show the first and third quartiles, and numbers indicate the sample count in each bin.

We have also added the following paragraph to Section 3.3.1 (lines 478-484):

"Upper-level aircraft observations are included in Figures 7c and 7e to evaluate the simulated ice field. Within the sampled anvil layer at approximately 12–14 km, the simulated ice particle number concentration and IWC profiles are broadly consistent with the available measurements. This comparison provides an observational constraint on the simulated upper-level ice properties, although the aircraft observations do not sample the mixed-phase region between approximately 5 and 9 km, where the largest SIP-induced enhancements occur."

Figure S1 has been retained to provide the more detailed comparison for Nd = 400 cm−3, including the model interquartile ranges and observational sample counts. Thus, the aircraft observations are now presented in the main manuscript, as requested, while Figure S1 provides the additional distributional information that cannot be shown clearly in the multi-simulation panels of Figure 7.

The final paragraph of the Conclusions has also been revised, as described in our response to Reviewer 1, to state explicitly that future work should include co-located in situ measurements in the mixed-phase region.

4. Satellite and radar evaluation

Response:

We agree that the roles of the observational comparisons and the sensitivity differences should be clearly distinguished. However, because the SIP effect at each Nd requires a matched allSIP–noSIP comparison, we have not created a separate evaluation section centred on one selected Nd simulation. Instead, the relevant observations are included directly in the corresponding figures and results sections, while the SIP response is calculated only from matched simulation pairs at fixed Nd.

The revised manuscript now includes the following complementary observational comparisons:

(1) Figure 2 compares the spatial evolution of MTSAT-derived equivalent OLR with the simulated allSIP cases, with the corresponding noSIP fields provided in Figure S2.

(2) Figure 3 compares the vertical distribution of radar reflectivity with the CPOL observations.

(3) Figure 4a compares the temporal evolution of MTSAT-derived domain-averaged equivalent OLR with all six allSIP and noSIP simulations.

(4) Figure 5 compares accumulated precipitation patterns with CPOL observations.

(5) Figure 6 compares precipitation rate and accumulated precipitation with CPOL-derived estimates.

(6) Figures 7c and 7e compare the simulated upper-level ice fields with Egrett aircraft observations.

In particular, the former supplementary OLR time-series comparison has been incorporated into Figure 4a and expanded to include all six simulations. The corresponding text now states (lines 279-289):

"The simulations broadly reproduce the observed decrease and subsequent recovery of OLR during this period. In all simulations, OLR decreases after approximately 02:00 UTC as convection develops, reaches a minimum during the mature stage, and recovers towards the end of the analysis period. During 06:00–08:00 UTC, the observed OLR minimum lies within the range of the simulations. At Nd = 400 cm−3, noSIP-400 remains above the observed minimum, whereas allSIP-400 falls below it, demonstrating that SIP produces a substantial shift in OLR under this condition. The SIP-induced change is smaller at Nd = 800 cm−3 and is small and of opposite sign at Nd = 200 cm−3. All simulations underestimate the observed domain-averaged OLR before and after the mature stage, indicating remaining differences in the timing and evolution of the simulated convection."

Because the former Supplement Figure S3 is now incorporated into Figure 4a, that supplementary figure has been removed to avoid duplication. The remaining supplementary figures have been renumbered accordingly. Figure S1, containing the detailed aircraft comparison, has been retained.

This approach shows the agreement and remaining differences between the simulations and observations across the experiment set, while the SIP effects themselves continue to be quantified only from matched simulations at the same Nd.

5. Dornier cloud droplet measurements

Response:

We agree that case-matched cloud droplet observations would be valuable. As described in our previous response, we examined the available Dornier FSSP data for 6 February 2006 and matched them with the corresponding aircraft position and altitude information. After applying a liquid water content threshold of LWC > 0.001 g m−3, only four valid samples remained. When grouped by height, these samples form two narrowly separated bins at approximately 0.38 and 0.41 km, with only two samples in each bin, as shown in Figure R1.

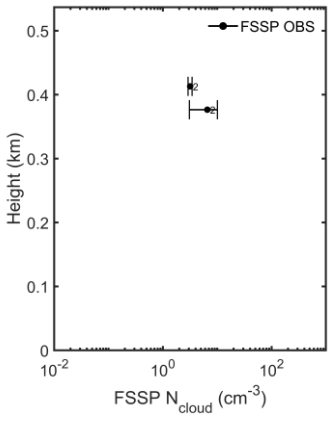


Figure R1. *FSSP cloud droplet number concentrations from the available Dornier aircraft measurements on 6 February 2006 after applying a liquid water content threshold of LWC > 0.001 g m^{-3}. Black circles show the mean values in each height bin, horizontal bars show the first and third quartiles, and numbers indicate the sample count in each bin. Only four valid measurements were available.*

The limitation is therefore not only the statistical uncertainty of the mean. Confidence intervals or standard errors could quantify the spread among these four samples, but they would not address their limited spatial, vertical, temporal, and cloud-regime representativeness. Moreover, with only two samples in each height bin, the resulting uncertainty estimates would carry substantial statistical uncertainty. In particular, these low-level samples cannot provide a representative constraint on the cloud droplet number concentration within the deep convective and mixed-phase regions examined in this study.

We therefore do not consider these four samples sufficient to identify one of the three prescribed N_d values as a uniquely constrained control condition. More broadly, the inflow aerosol observations and activation calculations support a plausible range of activated droplet concentrations, while the four valid FSSP samples provide only very limited direct information about low-level cloud droplets. Neither dataset uniquely determines the effective droplet concentration throughout the deep convective cloud.

We provide Figure R1 here to show transparently the amount and vertical coverage of the available case-specific Dornier data. However, we have not included this comparison in the manuscript or Supplement because four low-level samples cannot provide a representative constraint on the droplet number concentrations within the simulated deep convective system. Including them as a case-wide constraint could therefore give more weight to these observations than their limited sampling supports.

6. Cloud-fraction profiles

Response:

We agree that cloud fraction can be a useful diagnostic when model and observational estimates are available for the same case and are calculated using comparable definitions and

sampling criteria. However, the cloud-fraction estimates used in previous TWP-ICE/ACTIVE analyses represent broader campaign periods or different sampling periods rather than a case-matched observational profile for the 6 February 2006 storm. Differences in temporal sampling, spatial coverage, hydrometeor thresholds, and the definition of cloud occurrence would prevent a direct quantitative comparison with the simulated cloud fraction for this case.

A simulated cloud-fraction profile without a corresponding case-matched observational estimate would provide an additional model diagnostic, but it would not provide an independent observational constraint for selecting one of the Nd conditions as a control. We have therefore retained the case-matched diagnostics that evaluate different aspects of the simulated storm: CPOL reflectivity for the vertical convective structure, MTSAT-derived OLR for cloud-top and anvil evolution, CPOL precipitation for the surface rainfall response, and Egrett measurements for the upper-level ice field. For these reasons, we have not added a separate cloud-fraction profile.

7. Extent of the manuscript revision

Response:

We understand the referee's concern regarding the extent of the revisions. In this revision, we have made substantial modifications to the main manuscript to address these points. The principal revisions are as follows:

(1) Section 2.2 now explicitly defines the matched allSIP–noSIP comparison at each prescribed Nd and explains why the three Nd values form a controlled sensitivity framework.

(2) Section 2.3 now identifies where the MTSAT and upper-level aircraft observations are used in the main manuscript.

(3) Figure 4a has been revised to include the MTSAT-derived equivalent OLR together with all six allSIP and noSIP simulations. Its caption and the associated results paragraph have been revised accordingly.

(4) Figures 7c and 7e have been revised to include the Egrett observations of ice particle number concentration and IWC together with the simulations at all three prescribed Nd values.

(5) The Figure 7 caption now describes the observational binning, quartiles, sample counts, sampling limitations, and the relationship to the detailed comparison in Figure S1.

(6) A new paragraph has been added to Section 3.3.1 to compare the simulated upper-level ice field with the aircraft observations and to define the scope of this comparison.

(7) The former supplementary OLR time-series figure has been removed because the comparison is now incorporated into Figure 4a. The remaining supplementary figures and all associated cross-references have been renumbered.

(8) Figure S1 has been retained as a detailed comparison of the aircraft observations with allSIP-400 and noSIP-400, including model interquartile ranges and observational sample

counts.

(9) The final paragraph of the Conclusions has been expanded to identify the need for future case-matched in situ aircraft measurements of ice particle number concentration and particle size distributions in the mixed-phase region.

(10) The noSIP-200, noSIP-400 and noSIP-800 simulations are now explicitly identified as the matched controls for the corresponding SIP experiments.

(11) We have clarified in our response that the ACTIVE measurements distinguish between measured inflow aerosol properties and the updraught-dependent inference of activated droplet concentrations.

(12) The manuscript now explicitly distinguishes observational evaluation from SIP attribution and explains why a single noSIP simulation is not selected retrospectively on the basis of best observational agreement.

We believe that these revisions provide a clearer and more comprehensive comparison with the available observations, while preserving the matched-pair design required to address the central scientific question of how SIP responses vary across aerosol conditions.