

Dear editor and reviewers,

We thank the editor for handling our manuscript, and we are grateful to both reviewers for their careful reading of the manuscript and for the thoughtful comments. We have addressed all comments and revised the manuscript accordingly.

In this response document, reviewer comments are reproduced in black. Our replies are provided in green, and references to the manuscript are indicated by 'green italicized text in quotation marks'.

Reply to Reviewer 1

This is an important and relevant study that explores how contrails form and evolve in the absence of engine-emitted particles/nuclei. The focus and motivation is the case of H₂ fuel, which does not have the potential for carbonaceous particle emissions. The study specifically addresses how the lack of engine-emitted particles changes the nature of the contrail particle formation process and that it is significantly different when contrails are forming only on ambient particles. Existing tools that have been developed for carbon-based fuel use make some assumptions, which may be valid for the carbon-fuel case, but which are totally inappropriate for the case where only ambient nuclei are involved. This is an important distinction and calls into question previous modeling studies that used tools developed for carbon-based fuel. This is an important advance in understanding and needs to be recognized by the community.

There are several points that I feel should be further clarified and also some minor fixes that should be addressed.

Further clarification:

The paper focused on the distinction between modeling the entrained ambient particles with their continual entrainment from the "single nucleation pulse" that has been published previously. These are important distinctions and are the main point of the paper. But to put this comparison/distinction in context, a sentence or two should also make the comparison to carbon-based fuels (where pulse nucleation may be more appropriate). In particular, when discussing figure 9 c and d, those results beg the question of how those figures compare to a conventional fuel's N_{act} in m^{-1} . For K15-H₂ vs box model w/ mic. the factors range from ~4 to ~12 or so. What is the range when comparing to carbon-based fuels?

Thank you for this highly constructive comment. Providing a direct comparison to conventional fuels indeed helps to put the observed discrepancy between the single-pulse approach and the full microphysical simulation into a broader physical perspective.

For conventional kerosene combustion with a typical soot emission index of $El_{soot} \sim 10^{15} \text{ kg}^{-1}$, the number of activated particles N_{act} is on the order of 10^{12} m^{-1} . In that soot-rich regime, the single-pulse approximation utilized in K15 aligns well with time-resolved microphysical box-model simulations (i.e., a ratio close to 1) as shown in Figure 7b in Bier et al. (2022).

We have updated the manuscript at the end of the discussion of Figures 9c–d to explicitly provide this context to the reader:

'To put the depicted numbers into context, for conventional kerosene combustion with a typical soot emission index of $El_{soot} \sim 10^{15} \text{ kg}^{-1}$, N_{act} is on the order of 10^{12} m^{-1} . In that soot-rich regime, the single-pulse approximation of K15 aligns well with time-resolved box-model simulations (Bier et al., 2022).'

On lines 427-428, there is discussion about " . . . slower dilution yields fewer ice crystals in the box model because early-activated aerosols have more time to consume the available water vapor and

therefore suppress later activation more efficiently." This is an important additional distinction connected with the continual entrainment vs pulse nucleation. I think this needs to be discussed in more detail. The differences between the curves "with mic" and "no mic" in Figures 9 a and b vs c and d are, perhaps, somewhat related to this, since the supersaturation continues even after the pulse step and as some water is being continually removed by entrained ambient particles. Is the competition between newly entrained particles and the already growing "early-activated" aerosols clearly described and properly captured by the kinetics? The "early-activated" particles have more surface area, which may already be ice, so the water uptake is presumably significantly favored on those particles relative to newly entrained particles. This should be discussed and explained in more detail.

Thank you for this excellent comment. The reviewer describes the underlying microphysical competition very accurately. The ambient aerosols that are entrained and activated early have more time to grow and freeze into ice crystals, thereby developing a larger surface area. Water vapor deposition is favored on these pre-existing, larger ice surfaces rather than on newly entrained ambient particles.

We want to emphasize that this competitive kinetic process is fully captured by the LCM, which utilizes a particle-based approach. Instead of relying on simplified bulk assumptions, the model explicitly tracks the time-resolved diffusional growth equations for individual simulation particles (SIPs), continuously accounting for their evolving surface areas and phase states (liquid vs. ice).

Connection to Figure 9:

- In the "**no mic**" case, where condensation is neglected, the evolution of relative humidity is solely driven by dilution.
- In the "**with mic**" case, the growing, early-activated SIPs continuously consume water vapor, which gradually lowers the relative humidity peak compared to the "no mic" baseline.

Crucially, however, our results show that this vapor depletion by early-activated particles does not occur abruptly as aerosols are continuously entrained and activated. Because the available ice surface area grows continuously rather than instantly, the supersaturation is not immediately depleted. It remains elevated long enough to allow newly and continuously entrained ambient SIPs to overcome their activation thresholds. This explains why the final number of activated aerosols in Figures 9c–d continues to rise long after the initial K15 pulse step.

The Role of Dilution: This kinetic behavior directly explains why a **slower dilution speed** results in fewer ice crystals: if the plume dilutes slowly, new ambient particles are introduced at a lower rate. This gives the pre-existing, early-activated ice SIPs more time to consume the available water vapor before the next batch of ambient aerosols is entrained. Consequently, under slow dilution, the early-activated particles suppress the later activation stage much more efficiently.

To reflect this nuanced competition and the role of the particle-based kinetics more clearly, we have expanded the discussion in the manuscript:

‘Early-entrained and activated aerosols have already grown and may already be frozen, such that both their larger surface area and ice surfaces favor water vapor uptake relative to newly entrained particles. Slower dilution gives these early-activated aerosols more time to consume the available water vapor and therefore suppress later activation more efficiently.’

In addition, wouldn't these arguments suggest that the behavior of the two modeling approaches should also differ prior to the pulse nucleation in Figure 9? The subsequent paragraph explains why the K15-H2 droplets are larger, but shouldn't similar arguments suggest that there should be differences between K15-H2 and the box model prior to the step change?

The reviewer is conceptually correct that one might look for earlier differences. The near-perfect alignment prior to the step change can be explained as follows:

K15 tracks droplet growth during this initial phase but explicitly *'assumes negligible water vapor depletion so that the supersaturation evolves as it would without microphysical sink terms'* (Sect. 4.1.1).

In the **box model**, droplet growth and plume humidity evolution are strictly coupled. However, at the very beginning, the newly activated droplets are still **few in number and small**, meaning that their water vapor consumption is basically negligible. Consequently, *'the box model results support the assumption of zero water vapor depletion in K15-H2 until $t = t_{peak,K15}$ '* (Sect. 4.1.2).

As time progresses toward the peak, the models naturally begin to diverge. The fundamental split occurs at $t = t_{peak,K15}$, where K15 compresses the relaxation into a single pulse step (which is valid in the case of kerosene combustion), while the box model continuously resolves the prolonged activation kinetics.

Line 440: "This continuous process is especially critical for hydrogen combustion," This seems very true and is the motivation and focus of the paper. But this study, and this particular phrase, also raises the question of what would happen if all particles/nuclei could be removed from carbon-based fuel exhaust? Recent studies have highlighted that stage lean-burn combustors using JetA have essentially no soot emissions. If oil emissions were also eliminated (thought experiment), the same situation of needing ambient aerosol to nucleate contrail particles would exist. It seems the authors should mention that this modeling approach would apply to that situation, also. Of course, the water EIs and CO2 EIs would be different, but the modeling approach of continual entrainment of the primary nuclei would still be needed.

Thank you for this very interesting thought experiment. We completely agree with the reviewer's conceptual point: methodologically, if a future carbon-based aviation scenario were to eliminate not only soot but also all volatile particles (such as those formed by fuel sulfur content and organic compounds), solely ambient aerosols would remain as condensation nuclei. In such a hypothetical particle-free regime, our modeling approach and the focus on continuous entrainment would indeed provide the appropriate framework, even though the water vapor and combustion heat emission indices would differ from the hydrogen case.

However, recent modeling and measurements highlight the importance of volatile particles in soot-poor regimes (e.g., Yu et al., 2024; Dischl et al., 2025; Voigt et al., 2026). We therefore consider this scenario to be purely theoretical at this stage. To avoid adding highly speculative discussions to the manuscript and to keep the text strictly focused on the physical application of hydrogen combustion, we have chosen to keep the concluding, universally applicable formulation:

'Our analysis demonstrates that the K15 parameterization should not be applied in regimes where droplet formation on entrained particles is dominant.'

This statement inherently covers any regime where ambient aerosols become the primary driver of ice nucleation, whether in hydrogen exhaust or under hypothetical particle-free conventional fuel conditions, without shifting the focus away from the primary scope of this paper.

Minor comments:

1) Line 18, in the introduction, the authors appropriately quote references on the importance of contrails to aviation's radiative forcing. But they should also mention Petzold et al., 2025 Nature Communications, which raise questions that those earlier papers may be over estimates. Clearly more work is needed, but that study should be acknowledged.

We added general comments on the uncertainties associated with radiative forcing estimates to the manuscript:

'Major sources of uncertainty in the estimated radiative forcing include uncertainties in engine emissions (e.g., Moore et al., 2017; Märkl et al., 2024; Dischl et al., 2025), contrail formation (e.g., Lewellen, 2020; Yu et al., 2024; Ponsonby et al., 2025; Zink et al., 2025), the subsequent life cycle and properties (e.g., Bock and Burkhardt, 2016b; Unterstrasser et al., 2017; Akhtar Martínez et al., 2025) and radiative impact of contrail cirrus (e.g., Myhre et al., 2009; Forster et al., 2012; Lee et al., 2021), as well as the frequency of air traffic within pre-existing cirrus and its impact on cirrus properties and radiative effects (e.g., Seelig et al., 2025; Petzold et al., 2025; Verma and Burkhardt, 2026).'

Line 30 "Since thermodynamics and microphysics are substantially different for . . ." There should be no comma after "Since"

Corrected.

Line 434 ". . . it does not check, whether an aerosol particle . . ." There should be no comma after "check"

Corrected.

Reply to Reviewer 2

This manuscript presents a neural-network-based parameterization for predicting the number of ice crystals formed in contrails from hydrogen-powered aircraft. The work builds on the authors' previous studies and represents the final step toward developing a parameterization suitable for implementation in large-scale climate models.

The topic is timely and important given the increasing interest in hydrogen aviation and the need to better understand its non-CO₂ climate impacts. The manuscript is generally well written, scientifically sound, and based on an extensive database of LCM simulations. I particularly appreciated the detailed comparison with the K15-H2 framework, which helps clarify the physical differences between soot-dominated and ambient-aerosol-dominated contrail formation.

Overall, I believe the study represents a valuable contribution to the field and is suitable for publication after revision. However, I have several comments that I encourage the authors to address.

Major Comments 1. Generalization and applicability of the parameterization

The proposed parameterization is trained within a carefully defined range of environmental and aerosol conditions (Table 1). While the authors clearly state that extrapolation should be avoided, it remains unclear how the parameterization would behave when implemented in a GCM where atmospheric conditions may occasionally fall outside the training range.

I encourage the authors to discuss this issue in more detail. In particular:

- How should users handle situations where one or more predictors exceed the limits of the training dataset?
- Is clipping of input variables recommended?
- Are there regions near the parameter-space boundaries where prediction errors become systematically larger?

Thank you for raising these questions regarding the behavior of the parameterization at and beyond the training boundaries.

The training dataset was constructed including rather extreme values, meaning that atmospheric conditions that allow persistent contrail formation should rarely exceed these boundaries.

Due to the gridded nature of our training dataset with clearly defined boundaries, prediction errors do not systematically increase near the edges of the parameter space. Furthermore, testing shows that the networks do not exhibit spurious behavior or numerical instability even slightly outside the training domain.

Nevertheless, we recommend clipping predictors to the respective training boundaries if they fall outside the prescribed ranges. Although the networks behave smoothly near the boundaries, their behavior further outside the training domain has not been systematically evaluated and therefore cannot be reliably assessed. Clipping provides a conservative approach that avoids relying on extrapolation beyond the range covered by the training data.

The parametrization has already been integrated and is in use in the GCM ECHAM-CCMmod. In such GCM simulations, formation of persistent contrails outside of the parameter-space boundaries are very rare and, hence, do not pose an issue.

In the main text we added:

'Given that the training dataset includes a large range of values, conditions exceeding the parameter range are expected to occur rarely in situations where persistent contrails can form. Furthermore, because the networks were trained on a structured grid with sharp boundaries, performance remains stable near the edges of the parameter space without a systematic increase in prediction errors or spurious behavior. Nevertheless, we recommend clipping predictors to the respective training boundaries if they fall outside the prescribed ranges, as the behavior of the parameterization beyond the training domain cannot be reliably assessed. This provides a conservative approach that avoids relying on extrapolation beyond the range covered by the training data.'

2. Uncertainty associated with the neural-network approach

The manuscript reports relatively low MdAPE values and demonstrates good agreement with the simulation database. However, no estimate of predictive uncertainty is provided.

Since the parameterization is intended for climate applications, some discussion of uncertainty would strengthen the paper considerably. For example:

- How sensitive are predictions to uncertainties in aerosol number concentration or hygroscopicity? Since the neural networks emulate the underlying box-model simulations smoothly (as illustrated by the vertical profiles in Fig. 6 in the main manuscript), the sensitivity of the network predictions to changes in input variables—such as aerosol concentration or hygroscopicity—directly mirrors the physical response of the box model itself. The network does not introduce artificial sensitivity, discontinuities, or neural network artifacts. This ensures that any uncertainty in the input variables propagates predictably. Consequently, the total predictive uncertainty in practical applications is fundamentally tied to two factors (apart from the fitting error): the quality/accuracy of the provided input data (e.g., background aerosol fields) and the inherent

physical limitations of the underlying box model itself (such as the homogeneous freezing uncertainties discussed in Section 4.2).

In the main text we added:

‘As the neural networks emulate the physical sensitivities smoothly without introducing artificial discontinuities, the propagation of input uncertainties (e.g., in aerosol number concentration) into uncertainty in ice crystal number directly mirrors the response of the underlying box model without introducing neural network artifacts. The total predictive uncertainty is therefore mainly a combination of the uncertainty in the provided input data and the inherent physical limitations of the underlying box model discussed in Sect. 4.2.’

- How much variability exists among different network initializations?

To explicitly test the sensitivity of the parameterization to different network initializations, we trained **10 additional networks** with identical architecture as the FullNet but varying seed values. All 10 networks demonstrated highly stable behavior, with their predicted vertical profiles deviating by **less than 15%** from our original reference model FullNet (see Figure 1 in this review document).

- Would an ensemble of neural networks produce similar results?

While an ensemble of neural networks could theoretically be used, it would introduce a substantial and unnecessary **computational overhead** when implemented in a GCM. Given the high stability and low variability (<15%) among different seeds, employing a single, well-tuned neural network is fully sufficient and ensures maximum computational efficiency for climate applications.

I am not requesting a comprehensive uncertainty study, but a discussion of these aspects would be valuable.

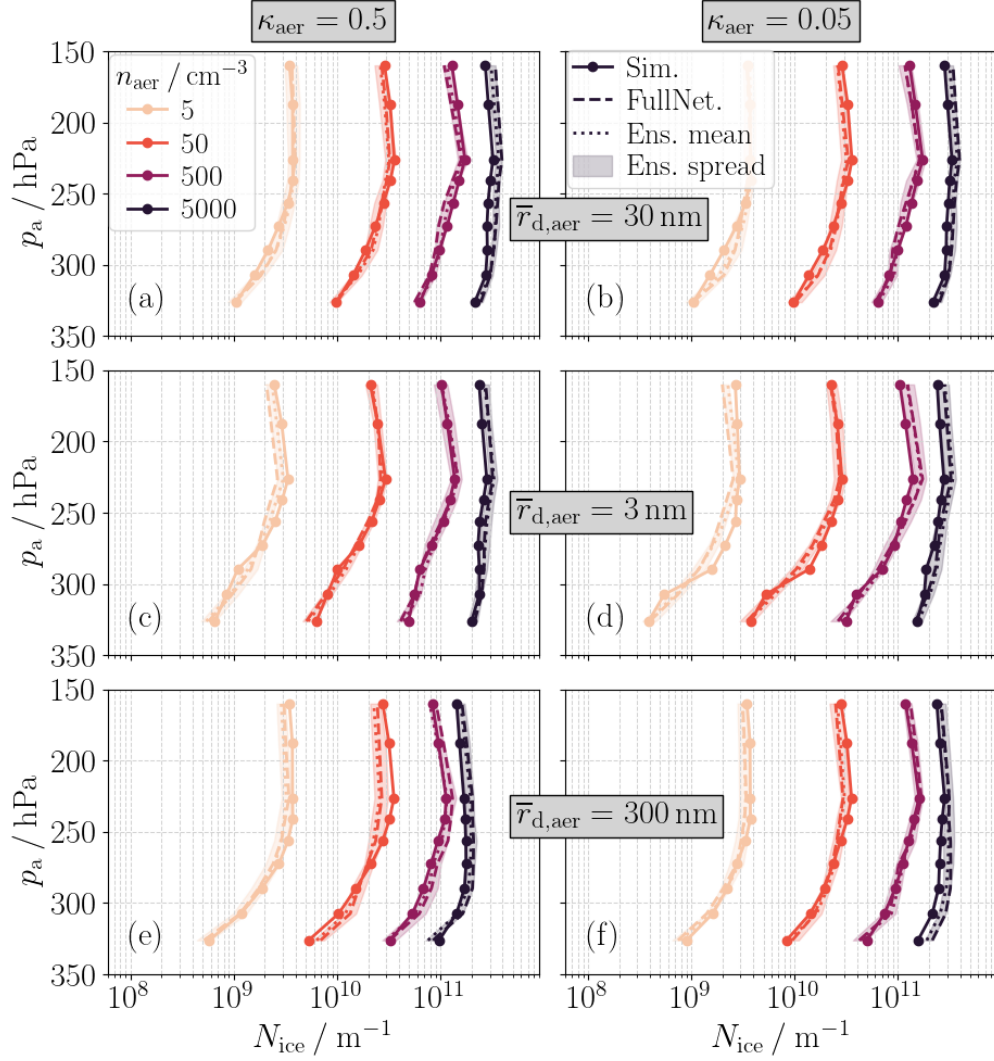


Figure 1: Vertical profiles of ice crystal number, formatted as in Figure 6 of the main manuscript. The plot compares the underlying box-model simulations (solid lines with dots) against the predictions of the original FullNet reference model (dashed lines). Additionally, the ensemble mean of ten additional network initializations (varying seed values) is shown (dotted lines), with the shaded area indicating the full ensemble spread between the minimum and maximum predicted values.

3. Dependence on the underlying box-model assumptions

The neural networks effectively reproduce the behavior of the LCM simulations. Consequently, the validity of the parameterization is tied directly to the validity of the underlying box model.

The authors acknowledge this point in Section 4.2, but I think it deserves stronger emphasis. In particular, uncertainties associated with homogeneous freezing parameterizations at high cooling rates and elevated contrail formation temperatures remain significant. Since these processes are central to hydrogen contrail formation, a more detailed discussion would be beneficial.

Thank you for this valuable comment. We have revised the discussion in Section 4.2 to better emphasize the uncertainties related to homogeneous freezing.

We added:

‘For instance, the formation of ice crystals in hydrogen exhaust is primarily limited by the

homogeneous freezing temperature of supercooled droplets. In contrast to contrail formation when using kerosene, this homogeneous freezing threshold is typically lower than the classical Schmidt–Appleman threshold temperature in hydrogen combustion scenarios (Bier et al., 2024). The employed homogeneous freezing parameterization (described in Appendix B of Bier et al., 2024) uses a combined approach that accounts for both the cooling rate (i.e., the change of supersaturation) (Riechers et al., 2013) and the solution effect of aqueous droplets (O and Wood, 2016). This formulation captures the strong, highly non-linear decrease in freezing fraction of supercooled droplets toward $\sim 235\text{K}$ observed in laboratory experiments (e.g., Murray et al., 2010; O and Wood, 2016; Tarn et al., 2021; Ponsonby et al., 2024). However, these laboratory setups are typically more representative of natural cirrus cloud formation with significantly lower turbulence and slower cooling rates than in a highly turbulent jet. Consequently, translating these non-linear thresholds to actual contrail conditions remains a major source of uncertainty. Therefore, measurements at contrail formation conditions for which our simulations indicate a strong influence of homogeneous freezing (ambient temperatures $\sim 230\text{K}$) would be particularly valuable for validation.'

4. Potential role of lubrication-oil particles

I found the discussion of lubrication-oil particles particularly interesting. The manuscript suggests that even very small oil emissions may substantially alter ice crystal formation.

This raises an important question regarding the practical applicability of the current parameterization. If oil-derived particles become dominant under realistic engine conditions, then ambient aerosol particles may no longer represent the primary source of ice nuclei.

I recommend that the authors clarify:

Thank you for this discussion regarding the role of lubrication-oil particles. We have addressed your points as follows:

- Under which conditions the current parameterization is expected to remain valid.
As stated in the conclusions, our parameterization is specifically developed for the formation of contrail ice crystals on **entrained ambient aerosols** in a particle-free exhaust. As shown by Zink et al. (2025) (specifically their Figs. 5 and 6), the regime where ambient aerosols dominate versus where oil particles dominate depends heavily on assumptions regarding oil particle size and number in the exhaust. Since these parameters are currently unconstrained by real-world data, the parameterization remains valid as a baseline for the ambient-aerosol-dominated scenarios identified in their work.
- Whether oil-particle effects are expected to be important for all hydrogen engine concepts or only specific designs.
We agree that the state of research does not allow for a definitive generalization across all engine concepts. As already discussed in the manuscript, the impact depends strongly on the specific engine configuration—primarily where the lubrication system vents (e.g., core exit vs. far from hot regions): *'The impact depends strongly on where the lubrication system vents. If venting occurs close to the hot core exit, entrainment into high-temperature regions promotes evaporation and re-nucleation. If the oil breather flow is released far away from these hot regions, the oil might be entrained into the plume when it has already cooled down such that it does not evaporate. These oil particles are then expected to be larger (a few hundred nanometers) and correspondingly fewer in number, thus contributing only marginally to the total ice crystal number. Thus, lubrication oil poses a potential concern for certain engine configurations, though technical measures to reduce or mitigate this effect may be feasible.'*

- How future parameterizations might incorporate this mechanism.
Currently, the state of research does not provide further clarity on this matter. The hypothetical scenarios discussed in Zink et al. (2025) remain unconstrained due to the lack of measurements, preventing a more definitive classification at this stage. We have added the following sentence to the end of this paragraph in the manuscript:
'To incorporate these effects into future parameterizations, representative in-flight or test-rig measurements are required to constrain currently hypothetical modeling scenarios with empirical data. Methodologically, the potential effect of lubrication-oil particles, including its dependence on engine architecture, could be incorporated into the parameterization by training the neural network on representative scenarios in which properties of the lubrication-oil particles, such as number concentration and size distribution, are included as additional predictors.'

Minor Comments

1. Section 1.2 explicitly states that the text is repeated from Part 1. While understandable, the repeated material could be shortened to improve readability.
We understand the reviewer's concern regarding the repetition. However, since this manuscript is part of a series of three distinct papers with different focuses, we believe keeping this brief summary in Section 1.2 provides essential context and guidance for the reader. To balance readability, we have ensured this section remains as concise as possible.
2. The distinction between FullNet, ReducedNet, and ClassificationNet is eventually clear, but a schematic workflow illustrating how the three networks interact would help readers.
Thank you for this suggestion. The schematic workflow and the interactions between these three networks are already illustrated in the flowchart in Figure A1.
3. Figure 2 could be improved by explicitly indicating the number of inputs used by the different network configurations.
Thank you for this note. While the number of inputs was already indicated by the open and closed circles, we have updated the caption of Figure 2 to explicitly mention the exact number of inputs for each network configuration.
'For both the FullNet and ClassificationNet, all depicted input variables are used (in total six). The ReducedNet uses only the input variables marked by dashed circles (in total three).'
4. The rationale for choosing MdAPE as the primary evaluation metric deserves a brief explanation.
Thank you for this comment. Since the ice crystal number spans several orders of magnitude, a relative metric is required to ensure accurate performance across the entire range. Therefore, we chose the absolute percentage error as metric. As shown by the cumulative density function in Figure 5, the median provides a fair representation of the overall error behavior, justifying its use as a single metric. We have added a brief explanation of this rationale to the manuscript:
'As the APE is a relative error metric, it provides a consistent measure of performance across the full range of ice crystal numbers. Moreover, as demonstrated in Sect. 3.1, the MdAPE provides a representative summary of the APE distribution and is therefore suitable as a single evaluation metric. We therefore use Eq. (10) as the criterion for selecting the best-performing model.'
5. Appendix A appears central for practical implementation. Consider highlighting the workflow more prominently in the main text.
We deliberately placed these details in Appendix A to maintain a clear separation between the scientific evaluation in the main text and the technical steps required for practical

implementation. This structure ensures the primary manuscript remains focused on the physical analysis and network architecture, while readers interested in integrating and applying the parameterization can find all necessary steps consolidated in the appendix.

6. The manuscript would benefit from a statement regarding availability of trained weights, source code, and implementation examples.

Thank you for highlighting this. We completely agree that the availability of these resources is essential for the utility of the parameterization. As stated in the "Code and data availability" section, the *'python code following the steps described in Appendix A, along with the trained weight matrices, will be made available once the manuscript has been accepted'*. As the parametrization may have changed during the review process, we decided to publish the parametrization at the very end of this process (as also indicated in the cover letter to the journal). This procedure should ensure that only a single version of the parametrization is eventually published and in use.

The manuscript addresses an important problem in contrail climate modeling and presents a technically sound and well-executed solution. The neural-network parameterization appears robust, computationally efficient, and physically well motivated. I recommend publication after minor revision addressing the points raised above.

Additional changes to the manuscript

On-going collaborations with users of the parameterization revealed the need to expand the section on scaling properties. This led to changes in the middle part of section 3.3. (see document with track changes). Furthermore, a few citations of papers that appeared recently were added or updated.