Comment "Predicting Ice Supersaturation for Contrail Avoidance: Ensemble Forecasting using ICON with Two-Moment Ice Microphysics" by Hanst et al.

General Comments

The study presents an improved Two-Moment microphysical scheme and then evaluates the performance of the ICON model equipped with this scheme in predicting ice supersaturation up to 48 hours in advance. The evaluation is carried out using observation data measured by radiosonde and aircraft as references. The performance of the model with this new parameterization is also compared to that of the operational One-Moment version of the scheme.

The study is part of efforts to improve flight routing in order to avoid areas of ice supersaturation. Assessment scores that are particularly relevant and well suited to this objective were used. The results show that the Two-Moment scheme provides better performance than One-Moment one for ISSR forecasts. The authors also explored a machine learning approach, which proved promising. This work is of high quality and importance, and it represents a significant contribution to ISSR forecasting. The paper fits the scope of ACP. I highly recommend its publication, but some improvements are necessary.

Overall, the manuscript could benefit from a more concise presentation. Some statements are repeated throughout the text, and certain elements currently included in the Introduction would be more appropriately placed in the Model setup Section. In addition, a significant issue remains concerning the evaluation of the model against fine-scale observations used as a reference, as well as the treatment of the uncertainty associated with these observations.

Specific Major Comments

1. The introduction could be presented more concisely:

    a. Lines 69 to 100 should be limited to presenting the ensemble forecasting system as described, and include a bit more details on the one-moment scheme, mentioning its limitations, and then explaining the motivation for transitioning to a two-mode scheme.

    b. The sentence spanning lines 69–77 should be moved to the end of the Introduction; otherwise, the aims of the work are introduced too early.

    c. The paragraph from lines 80 to 82 should be moved to the Model *Setup* section.

    d. The paragraph from lines 101 to 104 should be merged with the last paragraph of the introduction, for example as follows: "This work consists of presenting a new version of the scheme … and assessing its performance against observations and the old version. It is structured as follows: …".

2. Description of the parameterization: The appendix (lines 609–610) indicates that this parameterization is a simplified version of that of Köhler and Seifert (2015), but it does not specify how it differs, and this is also not clarified in Section 2.1. I suggest summarizing Sections A1, A2, and A3 – within the main body of the article in Section 2.1, to improve readability, understanding, and reproducibility, especially since this parameterization does not appear to have been published elsewhere. This could be achieved without significantly lengthening the manuscript.

3. The presentation of the evaluation metrics is currently scattered throughout the *Results* section, which somewhat affects the readability. It may be clearer and more convenient for the reader if the authors dedicate a specific section to these metrics and concisely.

4. Line 154: You mention that "the system includes stochastic perturbations of selected parameterisations". Could you specify which parameterisations are perturbed in your simulations? Furthermore, given that the model may become unrealistic with certain parameter values, could you clarify how the stochastic perturbations are applied? Is it limited to an area of the space of the parameter values, relevant to ISSR forecasting, in order to ensure that the model produces realistic values for these forecasts?

5. Section 4.1.2 – from line 230 to line 235: You mentioned that, in order to perform the spatio-temporal comparison, the observations (radiosonde and IAGOS data) were made comparable to the model grid by vertically interpolating the observations to the model levels. Unless I am mistaken, no horizontal spatialization was performed, only colocation. Consequently, these vertically interpolated observation values can be considered as relatively local.

a. Could you discuss the relevance of their approach of comparing these vertically interpolated values with those of the model, which has a horizontal resolution of 26 km?

b. Communities working on precipitation generally address resolution differences using spatial kriging, but this requires closer stations in order to calculate spatial covariance. I imagine that this type of approach is not usable in your study because measurements are not made everywhere at the same time and the distance between measurement points is very large. Consequently, it is important to discuss the implications of the differences in resolution on the results.

6. Ligne 258: In connection with the previous comment, is it appropriate to systematically use the threshold 100 % to define ice supersaturation in the model? Given the model's horizontal resolution (26 km), local areas of supersaturation may exist even when the grid-cell average remains below the saturation threshold. Have you tested the impact of using a slightly lower threshold, for example between 90 % and 100 %, to define ISSRs in the model, on its performance?

7. In the description of the observations, an uncertainty is mentioned, but it is not discussed further. Could the authors consider taking this uncertainty into account in the evaluation, as they did in Section 4.2.3 for the model by incorporating its ensemble spread? In other words, have you examined whether adjusting the threshold according to observations' uncertainties range, could impact the model performance score?

8. In lines 324-325: the authors state: “We also observe a more pronounced negative bias within the rank histogram, indicating that the model tends to underestimate RHice more often than it overestimates RHice.” Could this observation not be explained, at least in part, by the difference in spatial resolution between the observations and the model?

9. Ligne 616 : The mean diameter of ice crystals detrained from deep convection has been set to 200 µm. Is this choice based on observations? If not, would it not be relevant to consider it as a tuning parameter by defining a plausible range of values?

10. Lignes 629-630 : The authors assume that the concentration of desert aerosols (Ndus) is constant at 200 hPa and set to 1000 m^{-3}. Is this assumption realistic in all regions, particularly in tropical deep convection areas, where aerosol vertical distributions may vary due to convective transport processes? Wouldn't it be appropriate to set Ndus based on latitude and/or treat it as a parameter to be explored within a plausible range of values?

11. Section 5.3: The authors discuss the model resolution and neighborhood considerations. In connection with Comment #2, could the authors comment on the uncertainties introduced in the results by the difference in resolution?

Specific Minor Comments

1. Line 2. ISSR should be defined earlier as RHi>100/%, preferably in the first sentence rather than in line 6.

2. Line 23: Please specify the phenomena to which you are referring.

3. Line 59: The part of the sentence from “evaluated...” to “were found” is not very easy to read, as it presents metrics without explaining their implications and also uses acronyms without definitions. Since these details are not essential for understanding the rest of the paragraph,, I suggest that the authors replace this part with a more general formulation such as: *“evaluated using accuracy assessment metrics.”*

4. In the caption of Fig. 1, please specify the altitude or pressure level corresponding to what you define as the near tropopause.

5. The results presented for the machine learning approach are very interesting. However, they could be better highlighted in a separate article where the method would be described in more detail, making it accessible and useful to a wider audience.