Reply to Reviewer #1

We thank the Reviewer for the time she/he spent on the manuscript and for the useful comments. We have addressed all the comments from the Reviewer, which has improved the manuscript.

For better legibility, the Reviewer's comments are highlighted in **bold** and changes in the manuscript are in *italic*.

General Comments:

The paper evaluates the ability of ERA5 to correctly represent temperature and relative humidity in the upper troposphere using data from IAGOS as reference. Biases in both fields are detected and characterised and a correction method is applied that corrects both T and then RHi using the so-called quantile mapping method. It is found that this method indeed is able to reduce the biases. Unfortunately it turns out that in spite of this improvement the prediction of contrail formation or contrail persistence is not improved. But at least, for other quantities that are relevant for the assessment of aviation climate impacts, e.g. the optical thickness of contrails, improved estimation seems possible (not individually, but in a statistical sense).

While the method leads certainly to a significant improvement of the RHi-statistics of ERA5, the investigations and analyses in the paper are somtimes a bit lengthy and, at least to my view, not always necessary. The differences between the original ERA5, and after the application of the QM method of the correction of Teoh et al are often minor and it is not sure whether they are always real or simply caused by statistical noise. But even if they are real in certain cases, it is not always clear to me why the reader needs to know the potential causes of these differences. I have the feeling that the reader can easily get lost in these details and that the straight way from the analysis to the results and conclusions becomes unclear. So, I think, there is potential to make the paper more concise and to clearly convey a message.

Multiple sections and paragraphs of the manuscript have been shortened and revised to make the manuscript more concise. For specific modifications to the manuscript we would like to point the Reviewer to the revised manuscript with track-changes and to the specific answers to major comment number 2 as well as to minor comments numbers 14 and 16.

This paper needs at some places more elaborateness in its formulations. For

instance, the sentence (Page 1) "IAGOS flight trajectories are used to extract co-located meteorological conditions - temperature, relative humidity, and wind speed - and are compared with the IAGOS measurements" is obviously a faulty. In short, it says that IAGOS trajectories are compared with IAGOS measurements, which is evidently nonsense.

Following the Reviewer's comment the sentence has been corrected.

"IAGOS flight trajectories are used to extract co-located meteorological conditions from ERA5, namely temperature and relative humidity, which are compared with the IAGOS measurements."

Also "representation of contrail occurrence in ERA5" (P 1) is misleading, since contrails are not represented in the ECMWF model and reanalysis.

The sentence has been rephrased to be clearer.

"To estimate potential contrail formation on the basis of ERA5, data points from IAGOS as well as corrected and uncorrected data points from ERA5, temperature and relative humidity, are flagged for contrail formation using the Schmidt-Appleman criterion."

Therefore I recommend a major revision of this paper to make its messages stronger and clearer.

Major issues:

1.) Reading the abstract it seems that the evaluation merely uses bulk statistics, which I deem insufficient. An good indication of this is figure 5, where, apart from MD, all other measures are very insensitive to the applied corrections. The r² hardly change (and these tiny changes may be insignificant on a p=0.05 level). The r², like the other insensitive statistics, are bulk measures, that say nothing to a point-by-point comparison. The authors acknowledge this in the first sentence of section 3.4.

The Reviewer is right that we primarily use bulk statistics. However, these bulk statistics do rely on a point-by-point comparison after careful sampling of the ERA5 data. The proposed quantile mapping technique is commonly applied to remove biases in data sets, which is the goal of this study. It is not designed to improve the other statistics. Our study is nevertheless an improvement over the state-of-the-art. Earlier correction methods of relative humidity from ERA-interim and ERA5 estimated the contrail occurrence based on a simple scaling technique, which, in principle also aims to compensate a potential bias in the dataset. Our method is a more elaborate way of applying a bias correction as we not only use an established method but we also apply it as a function of temperature.

A more sophisticated method aiming at exploiting more of the ERA5 variables and/or their three-dimensional structure remains to be developed. This is the subject of ongoing work.

2.) Section 3.4: The selection of the score values is an unlucky choice to my view, since the TN cases dominate in this data set. Thus, if one would ignore everything and always predict "no ISS", one would already get an impressive accuracy and FA rate. The statement "which indicates that the overall performance of ERA5, even in the uncorrected form, is at least similar or even has improved" is thus misleading. It may be that the author made their choice to compare with Tompkins et al. (2007). If the goal of the paper is to provide better data on the occurrence of ISS, this might be ok, although I still think, that scores that downweight the default TN would be better. If the ultimate goal is, however, forecast of contrail persistence, I think, that the authors should better compare with Gierens et al. (2020) and also use the ETS score described there.

We followed the suggestion of the Reviewer and computed the equitable threat score (ETS) based on the four entries in the contingency table. Following this work, we replaced the calculation and discussion of hit rate, false positive rate, and accuracy with the ETS, which is indeed more informative. The calculated ETS is given in Table 5. Accordingly, the entire sections "Along-track contrail formation potential and the effect of applied corrections" and "Analysis of collocated contrail formation potential from ERA5 and IAGOS" were revised.

in the course of the revision of the metrics, we found a mistake in the contrail flagging algorithm, which has now been corrected. This leads to slightly different results in the contrail categorization and the contingency table in Figures 7 and 8 as well as Tables 3 and 5. The quantile mapping itself remains unchanged but the conversion from RHi w.r.t to ice and liquid water was unnecessarily applied in succession, leading to a cumulative error in the conversion of RH between ice and liquid water. The error was dominant for samples with low temperature, where the saturation curves of liquid water and ice are close to each other and small errors had a significant effect on the calculated RH (Ambaum, 2020).

The revised analysis and the evaluation of the impact of the bias correction on the contrail categorization with the ETS now shows an improvement of NPC, PC, and R detection from ERA5 data against the IAGOS reference observations. A similar improvement was identified for the scaling method in Teoh et al. (2022).

We refer the Reviewer to sections "Along-track contrail formation potential and the effect of applied corrections" and "Analysis of collocated contrail formation potential from ERA5 and IAGOS" in the revised manuscript and the manuscript with track-changes.

Minor issues:

1.) Lines 34-38: I suggest that you state that RF is a global (or at least regional) quantity, averaged over a long time period. On first reading, it was not so clear to me whether you refer to single contrails or contrails in general. Furthermore, aren't the quoted values ERF values (in Lee et al.) rather than RF values?

To be consistent within the text, we removed the citation from Lee et al. (2021) and only kept the references from Boucher et al. (2021) and Burkhardt and Kärcher (2011), who

give estimates for the radiative forcing.

"The influence of a perturbation, e.g., clouds, aerosols, or gases, on the Earth's atmosphere and its radiative transfer is quantified by the radiative forcing (RF). By definition, RF is defined as the difference in the net irradiance at the top of atmosphere under perturbed and unperturbed conditionsI(Ramanathan et al., (1989)). In the context of climate studies the RF is understood as the difference in the Earth energy budget due to a contributor to climate change (Bickel et al., 2020). For example, the aviation-induced global CO2-related RF is estimated to be around 30 mWm^{-2} (Boucher et al., 2021). Contrail RF is estimated to be stronger, at about 60 mWm^{-2}, but is subject to much larger uncertainties (Burkhardt and Kärcher, 2011)."

2.) L 48 and following: This text mixes up different things which is not good. In order to avoid contrails, one needs a precise PREDICTION of where they WILL occur. Knowledge of their occurrence is insufficient, it can only result in a kind of climatology. Schumann's use of the roof-camera is a bad example for predictive purposes, as it shows contrails that already exist and perhaps exist already quite some time. The same comment can be made for satellite or other observations of already present contrails. Then, contrail simulations in a climate model are not intended for contrail avoidance or prediction. CoCiP is the only example here where contrail prediction is the intention, but of course only, if it is fed with actual weather forecasts, not with ERA5.

We understand the objection of the Reviewer and rephrased the paragraph in the manuscript. We agree that there is only one way to avoid contrails, and this is by using numerical weather prediction models that predict regions that need to be avoided. However, it is useful to obtain a statistical database that documents the spatial and temporal distribution of regions where contrails are likely to form. The paragraph has been rephrased as follows:

"To lower the climate impact of aviation it is important to reduce CO2 as well as non-CO2 effects. An approach to minimize non-CO2 effects is active flight re-routing to avoid areas where contrails are likely to form and persist, which would require accurate numerical weather predictions. A useful prerequisite is to identify and document flight levels and regions of the Earth's atmosphere that are particularly prone to contrail formation due to meteorological and dynamical conditions that favor contrail formation. Such a statistical data base might be obtained in four different ways.

The first approach builds on ground-based observations. For example, [...]"

3.) L 75: It is a bit surprising that a "fourth approach" is now mentioned. I believe that the mixed-up list from above (L 48 ff) is now continued, but that is not sufficiently clear. Again, this is not an approach to prediction and should thus not be mentioned as something that has to do with contrail avoidance. I find also, that this section interrupts the logic of the argumentation. It would be better if the

paragraph that introduces various correction attempts would directly by followed with tha paragraph explaining the goal of the present paper.

We agree with the Reviewer and rearranged the paragraphs for a better logical flow. Please also see the answer to minor comment 2.
We now clearly state that the primary way to avoid contrails is active re-routing of flights by using numerical weather prediction and by circumventing contrail prone areas. However, creating a statistical database of contrail formation potential can be achieved by the four methods that we listed. (The list is not necessarily exhaustive.) Due to the length of the paragraphs we direct the Reviewer to the updated manuscript and the manuscript with track-changes.

4.) Fig. 2 and corresponding text: it seems that either the information in the figure is useless or something is wrong. For instance, why are there so many 175hPa data for EU, where I would expect a lot of landing and departures (i.e. low altitudes)? How do I have to read the figures? It seems, the interpratation is: on 300 hPa most flights are over EU etc..... and on 175 hPa again most flights are over EU. Is this, because most flights are over EU anyway? Then the information is useless. Should't it rather be the following: In EU most flights are on lower altitudes because of a lower fraction of cruise, similar in US but perhaps not as strong, and over the NA all flights are in cruise, therefore a predominance of high altitudes (or low pressures). It seems, the data should be organised the other way, i.e three panels "EU", "NA", "US" and then showing the fraction of pressure levels in the threee panels.

As described in section "Quantile mapping", the CDFs have been calculated on individual pressure levels that are determined by the ERA5 pressure levels. Individual levels have been considered because the temperature and humidity biases do not have to be consistent in the vertical (across pressure levels) nor in the horizontal (across sub-domains); see Fig. B1 in the appendix. Considering the spatial dependencies of the biases, it is most important to be aware of the distribution of the IAGOS data.

By separating into individual pressure levels (Fig. 2) we intended to clearly show the respective contributions of each region to measurements for a given pressure layer. It should be noted that the total number of samples is given for each sub-panel of the Figure. The majority of samples are obtained at the 200 hPa level, which corresponds to the cruise level.
The Reviewer pointed to the 175 hPa layer, which she / he assumes to be incorrect. It is true that the majority of IAGOS flights depart and arrive in the EU. However, the fraction of samples in the EU domain on the 175 hPa layer is high because aircraft reach their maximum altitude at the end of the flight (when the fuel tanks are almost empty).

5.) Section 2.1: Please indicate whether all data are used along a flight (it seems so) or a subset to avoid autocorrelation which might spoil the statistics. If autocorrelation has not been avoided, a check should be made whether this affects the results or at least good arguments for this should be given.

All IAGOS data was filtered for data quality using the provided quality flags provided by the IAGOS post-processing. After filtering, around 90 % of all the data was usable.

The analysis that is presented in the manuscript has been performed with all remaining 90% of the data but also for a second time, where only every fourth data point (one data point approx. every 4 km) was used (second analysis not shown in the manuscript). No significant differences in the results were found.

The data extraction method, i.e., selecting and sampling collocated data from ERA5 based on IAGOS flights in an identical manner, might lead to autocorrelation, which then exists in both extracted data sets and, therefore, should not influence the analysis. Autocorrelation might be a problem if trends within a single time series are analyzed. Autocorrelation is expected to be an issue for the temperature field, which is relatively homogeneous, but will not be an issue with the relative humidity field, which is subject to greater tempo-spatial variability.

To further elaborate on the question about the impact of spatial averaging of IAGOS data and potential autocorrelation we prepared the two plots below. The plot at the top shows IAGOS data at the original 4-s resolution and the lower plot shows temporally smoothed data (60 s). The obtained distributions change little between the original and the smoothed data and, thus, we argue that averaging IAGOS only has a second-order impact on the obtained statistics and subsequent analysis.
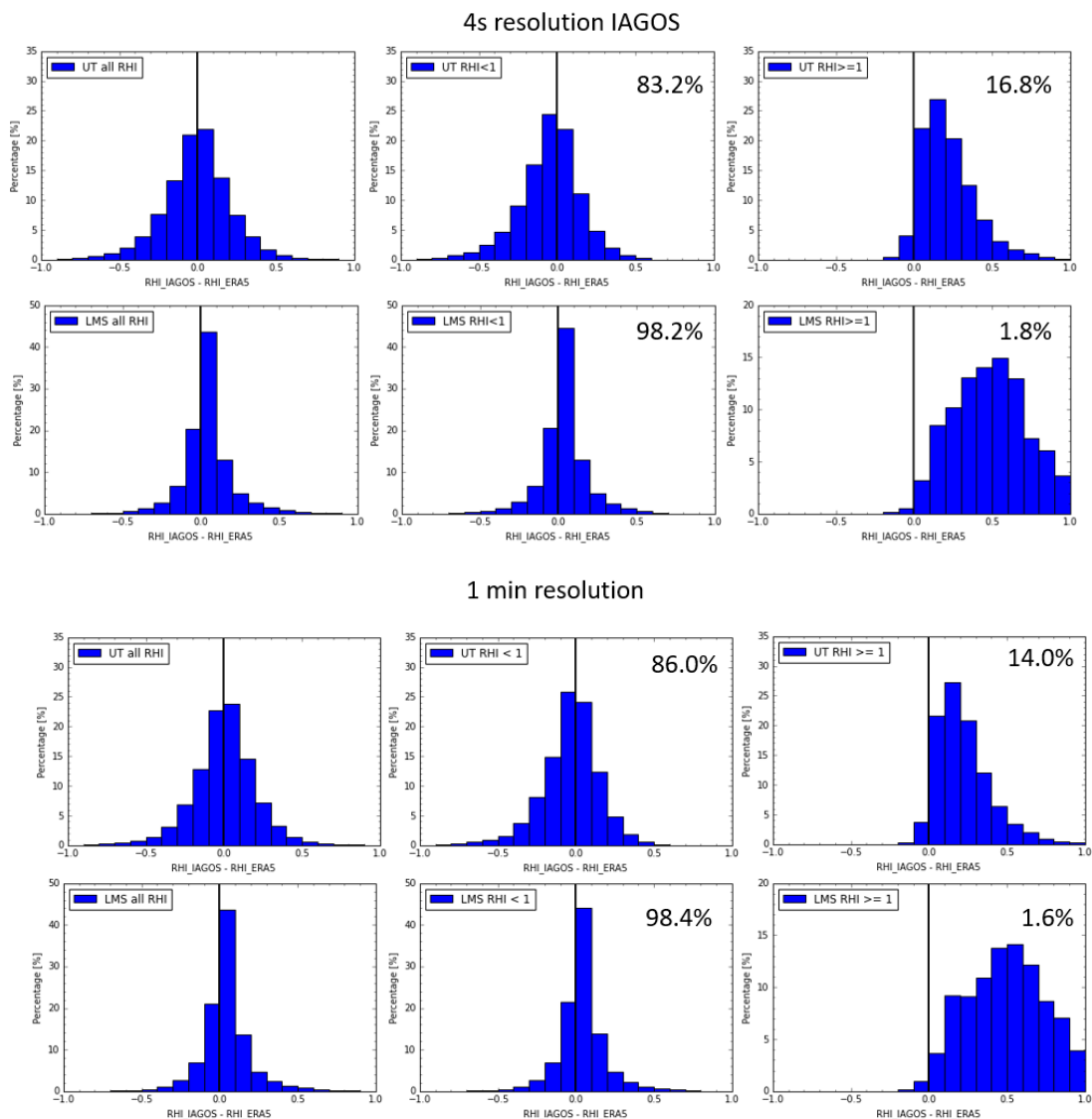
6.) L 153/154: I suggest to delete this statement since the procedure is better explained below from L 173 on. Anyway, I think the argument is weak (I don't remember Schumann's reasons for this) and refers rather to interpolation of specific humidity than relative humidity.

Interpolation of relative humidity is not straightforward as it depends on the underlying temperature and absolute humidity fields, and is determined on the basis of the exponential Clausius–Clapeyron relationship. Due to the nature of the Clausius–Clapeyron equation, linear interpolation, for example, leads to incorrect values of relative humidity. We rephrased the paragraph and removed the reference to Schumann et al. (2012). The end of the paragraph reads as follows:

"Spatial and temporal interpolation of relative humidity is not done because the relative humidity depends on the underlying temperature and absolute humidity field, which are both related through the Clausius–Clapeyron relationship.. Due to the exponential nature of the Clausius–Clapeyron equation, linear interpolation, for example, would lead to incorrect values of relative humidity."

7.) L 210/211: The statement "The CDFs describe the probability that a certain quantity, for example temperature or relative humidity, exists in the underlying data set" is wrong. I think, we know which quatities are in the data sets, therefore the probability is either zero or one. Please consult a textbook on probability and correct the sentence or leave it out. The CDFs describe the probability that a certain value of a quantity, for example temperature or relative humidity, exists in the underlying data set.

We agree with the Reviewer that the sentence was poorly phrased. The sentence has been rephrased to the following:

"The CDFs describe the probability that the value of a quantity (or random variable) X, for example temperature or relative humidity, has a value that is lesser or equal to x."

8.) L 239/240: You should delete the second part of the sentence. Contrail formation takes place a few tenths of a second after exhaust. It has nothing at all to do with the vortex phase which starts at, perhaps 20 seconds. I also dislike the next sentence. The SAC has been tested on many flights long ago, and it works excellently. There is a figure in the 1999 IPCC report that shows this (I think, the figure has been taken from a paper by Kärcher).

The Reviewer is right and we rephrased the sentence to be more precise.

"The SAc is based solely on thermodynamic principles and has been tested to be a valid approximation although it does not inform on the fate of the contrail, which is a more complicated function of the ambient conditions but also the interactions of the vortex phase with the environment."

9.) Section 2.4: I miss information on your choice of an overall propulsion efficiency.

This is correct and the propulsion / engine efficient is now mentioned in the text.

"Calculations are performed for kerosene with a fuel specific energy Q = 43.2 MJ kg^{-1} and an emission index of water EIH20 = 1.25. The overall engine efficiency η is set to a typical value of 0.3 (Rap et al., 2010)."

10.) L 272: is there really a general decrease of r_ice with p? Why then occurs ISS mostly directly below the tropopause?

In general, absolute humidity decreases with altitude (e.g. Kiemle et al., 2012; Kaufmann et al., 2018, Kruegger et al 2022). However, local maxima of relative humidity can occur just below the tropopause, e.g., in regions with strong vertical updraft or advection of humidity, which then leads to cloud formation. We followed the suggestion of the Reviewer and rephrased the sentence as following:

"With the general decrease in absolute humidity and possible intrusion of dry air from the stratosphere, the first mode becomes more and more pronounced with decreasing p, while the second mode flattens and almost vanishes."

11.) L 297/298: it is not clear to me why the T22 correction cannot modify the shape of the pdf. Please explain.

The sentence has been rephrased and is now less definitive. We intended to say that scaling all humidity values above a certain value by a constant increases the values as a whole and shifts the peak of the distribution to higher relative humidity values.

"Furthermore, differences in the second mode in relation to the IAGOS observations remain as the T22-correction only scales values above a certain threshold, which primarily shifts the bulk of data points from 100 % to higher r_{ice}."

12.) L 370 ff: To my view the discussion of the differences between the QM and T22 is not convincing. For instance, the quoted thresholds from CoCiP are very very low, so they are not really a constraint. To me, the first question is, whether these differences are statistically robust. If a subset of four out of the five years is used, how large is the change of the quoted values? If it is much smaller than the difference between QM and T22, then a real difference seems more plausible. Second, it might be that you compare apples with oranges, i.e. for instance contrail distance with relative frequency of occurrence.

To answer the first question of the Reviewer, we want to make clear that we gave the estimate from Teoh et al. (2020) to provide a reference for the reader and to set our results in the context of existing literature. The differences between the results using our approach and the approach by Teoh et al. (2020) are clearly stated in the manuscript. Differences in approaches can partly explain the differences. We do agree with the Reviewer that the

absolute numbers given in our manuscript and in Teoh et al. (2022) might be subject to spatial and temporal effects. At the same time the differences between our calculations of potential contrail occurrence and the calculations from Teoh et al. (2022) must not be overinterpreted. They only indicate the uncertainty due to different approaches.

Regarding the second question of the Reviewer, we think that the comparison is valid as the study by Teoh et al. (2020) entitled "Mitigating the Climate Forcing of Aircraft Contrails by Small-Scale Diversions and Technology Adoption" provides the percentage of flights that form contrails rather than the fraction of time it produces contrail. We direct the Reviewer to Table 1 in Teoh et al. (2020).

13.) L 408 ff: It is not clear to me how you interpret the SAc. Why does an rcrit occur? As far as I understand it, the SAc gives a threshold temperature which is the maximum T where a contrail could in principle form (at 100% RH). Is rcrit the RHi values along the tangential mixing line between the T threshold (rcrit=100% RH) and the lower temperature below which contrails are always formed (rcrit=0%). This is not completely clear to me.

The critical threshold r_{crit} is always 100 % (w.r.t liquid water). This was incorrectly formulated in the last version of the manuscript. In course of the manuscript revision the sentence in line 408 was also removed.

14.) Figure 7 and corresponding text: This text contains a lot of details and numbers that I find not necessary. What is the take-away message that the reader learns from these details? Again, how robust are these values, if you leave one your out, then another one, etc.?

Following the suggestion by the Reviewer the text has been shortened by removing unnecessary sections. Figure 7a is discussed in detail and the influence of the QM correction (Fig.7c) is presented. We focused on Fig.7 a and c as both are essential to see if the improved correction has an impact on the estimated contrail formation conditions. Furthermore, the discussion of the differences between the two methods is used to explain the rationale behind Fig. 7. Figure 7b and d are only briefly mentioned but not discussed in detail. We further highlight that Fig. 7 and the given numbers should be interpreted in a qualitative and not quantitative way, as they can change slightly between years. The section now ends with a summary (take-away message) that states the robustness of the ERA5 estimated PC occurrence with respect to the temperature and humidity correction. The minor impact of the corrections on the potential PC occurrence implies that ERA5 is suited to estimate regions that are prone for contrail formation. We would like to direct the Reviewer to the revised manuscript to see the changes.

15.) L 482: "Applying ... more correctly DETECTED...", please reformulate. The QM method does not change any detection.

The sentence has been rephrased and now reads as:

"The application of the QM-correction modifies the distributions of temperature and relative humidity in such a way that locations of PC are detected more often where they are supposed to occur and were missing before, resulting in an increase in ETS from 0.27 to 0.36."

16.) L 493-506: This is an interesting consideration. But it comes too late. Perhaps a lot of not-so-important numbers could be saved if this consideration would be presented earlier.

We agree with the Reviewer that this is an important point to mention. Therefore, the paragraph has been moved earlier in the section. The paragraph, which mentions the potential mismatches and that investigates the 3h time resolution, is now used to explain the baseline and sensitivity of the scores on the temporal-spatial misalignment. This provides a reference for the scores that have been calculated on the different correction methods. Differences that exceed the baseline are thus truly attributable to incorrect values of temperature and relative humidity in ERA5 or the corrections, and are therefore of relevance.
We also followed the Reviewer's major comment 2 and calculated the ETS. The ETS is provided in Table 5 and we also added a paragraph to the section, where we briefly mention the impact of the corrections on the ETS.
We would like to direct the Reviewer to the revised manuscript to see the changes in Section "Analysis of collocated contrail formation potential from ERA5 and IAGOS".

17.) Section 3.5 and Figure 9: please state whether you use for the analysis original or QM corrected ERA5 data.

The section and the figure are based on QM-corrected ERA5 values as we intended to evaluate the remaining errors after the correction that are associated with a certain classification. The following sentences have been added at the beginning of the section. The caption of the Figure has also been modified.

"Subsequently, we aim to quantify the mean differences in temperature and relative humidity that remained after the QM correction and that contribute to the misclassification of potential contrail formation. Within the following section all ERA5 values are QM-corrected."

18.) L 536-539: If that happens only on 225 hPa, is it then really likely that small scale variations cause these differences? Don't such variations occur on the neighbouring pressure levels as well? In which sense are T and rice "more homogeneous" on 200 hPa? What does that mean and from where do you derive such a conclusion?

Thank you for pointing this out. The text has been revised and rephrased as follows:

"Least frequent are the misclassifications 'NoC–PC' (light blue, 0.3 %) and PC–NoC (dark blue, 0.5 %). These two groups are subject to the largest ΔT and Δrice. Samples in these categories were only found at the 250 and 225 hPa p-level, while the PC–NoC (dark blue) is not found at the 200 hPa level. It is likely that data points in the two categories result from small scale variations captured by IAGOS that are not represented by ERA5 due to temporal and spatial resolution."

19.) L 539/540: Isn't this a trivial statement? Did you expect something else? Perhaps there are more interesting results in this section that could be more elaborated. For instance, which error usually dominates, is it the error in T or the error in rice. It would also be interesting to see how the differences in absolute humidity contribute to the misclassifications. What is here the conclusion? What needs to be fixed in the ECMWF model more urgently, is it T or the water vapour field? What is the effect of the QM correction in this analysis? I suggest you add a second set of points (squares or empty points), perhaps with arrows, to show how the QM correction shrinks the errors. Please think also on getting rid of the "default" class in the figure. It seems as if a couple of dots are hidden behind the big black dot in the middle.

We agree with the Reviewer that the sentence is trivial, but we meant to express that it is worth identifying whether the misclassification in ERA5 with respect to IAGOS is most often due to biases in temperature or in humidity. The sentence and paragraph have been rephrased and extended in this regard. The paragraph now reads as follows:

"It is worth identifying whether the misclassification in ERA5 with respect to IAGOS is most often due to biases in temperature or in humidity. Focusing on the PC representation in ERA5, the primary reasons for a misclassification after the correction is the deviation in rice. This is visualized by the proximity of the violet and green dots to the y-axis (small ΔT), while the differences in rice are larger than ±20 %. Hence, the underestimation (green dot) or overestimation (violet dot) of potential contrail formation is primary related to the underlying humidity field in ERA5."

20.) L 575: "So from a statistical perspective, that the original ERA5 model output is able to adequately represent contrail formation." First, the sentence is gramatically incorrect. Second, please explain what means that contrail formation is adequately represented FROM A STATISTICAL PERSPECTIVE? Assume, the model would only predict ISS and PC over Antarctica and nowhere else, but with the correct frequency of occurrence averaged globally, then one could also state that the formation is represented adequately from a statistical perspective. So my question is, whether your statement makes sense.

In the course of revising the manuscript this sentence has been removed.

21.) L 590/591: Isn't it the other way, that is, the special bias combination leads to the corresponding entry in the contingency table?

Following the suggestion of the Reviewer the sentence has been removed as it is a trivial statement. The sentences that follow in the paragraph are sufficient.

22.) Section 4: should not be entitled "Summary and discussion", since it does not contain any discussion.

"Discussion" has been removed from the subtitle.

Miscellaneous

23.) Line 3: upper tropospheric

The sentence has been rephrased as follows*:*

"The skill of the atmospheric reanalysis ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) at simulating temperature and relative humidity in the upper troposphere and lower stratosphere is assessed by using five years of In-service Aircraft for a Global Observing System (IAGOS) observations."

24.) L 19: underlying

The typo has been corrected.

25.) L 30: delete "bonds"

"Bonds" was deleted.

26.) L 31: What is WC?

This was a typo and is now corrected to "WV", which is defined on the line before.

27.) L 41-42: The word "defined" is too strong. In fact, this was never defined. There is only a vague understanding that contrails that survive the vortex phase are somehow persistent.

The sentence was rephrased as follows:

"For a contrail to be persistent (with the common meaning that is has a lifetime longer than 10 minutes), the ambient air has to fulfill the SAc and must also be supersaturated with respect to ice."

28.) L 67: "due to the high temporal and spatial distribution of WV" ???

The sentence was rephrased as follows:

"Slightly less accurate is the prediction and re-analysis of relative humidity, which is generally challenging due to the high temporal and spatial variability of WV."

29.)L 73: What is "contrail estimation"?

To be clearer the sentence has been rephrased as follows:

"To mitigate the dry bias under conditions close to ice-supersaturation in ERA-interim and ERA5, studies have applied either multiplication factors (Schumann et al., 2013, Schumann et al., 2015) or parameterized corrections (Teoh et al., 2022)."

30.) L 126: Replave "multiple" by "many". I believe "multiple years" is nonsense.

Multiple has been replaced by many.

31.) L 199: The expression "minimizing the C... test" is confusing. How can a test be minimized. I think this sentence can be dropped since it does not provide essential information.

We meant that the test statistic was minimized. However, we followed the Reviewer's suggestion and removed the sentence as details about how the parameters were detected can be found in Teoh et al. (2022).

32.) L 241: criterion (singular).

Thank you for pointing this out. The word has been corrected.

33.) L 244: slow

The typo was corrected.

34.) L 335: r_ice IS

The typo was corrected.

35.) L 433: remove one instance of correction

The word was removed.

36.) Table 6: the caption says "TO07" while the table headline says "T07"

Based on the suggestion from the Reviewer we use the equitable threat score to compare the measurements. Doing so removed the reference to Tompkins.

37.) L 513: delete comma after ERA5

The comma was removed.

38.) L 658: require

The typo was corrected.

39.) L 671: showS and remainS

Both words have been corrected.

40.) Table A1: eta should be labelled "overall propulsion efficiency". It is not necessary (and not good) to introduce new notions.

We agree with the Reviewer and added "overall propulsion efficiency" in the text and unified it in the table.

41.) Figure A1(e): The y-label contains [*2e6]. Please check.

The label is correct and scales the y-axis, which shows the number of individual IAGOS measurements (every 4 seconds) within a given month.

References

Appleman, H.: **The formation of exhaust condensation trails by jet aircraft**, *B. Am. Meteor. Soc.*, 34, 14–20, 1953.

Busen, R. and Schumann, U.: **Visible contrail formation from fuels with different sulfur contents**, *Geophys. Res. Lett.*, 22, 1357–1360, 1995

Heymsfield, A., Baumgardner, D., DeMott, P., Forster, P., Gierens, K., and Kärcher, B.: **Contrail microphysics**, *B. Am. Meteor. Soc*., 91, 465–472, 2010.

Jensen, E., Toon, O., Kinne, S., Sachse, G., Anderson, B., Chan, K., Twohy, C., Gandrud, B., Heymsfield, A., and Miake-Lye R.: **Environmental conditions required for contrail formation and persistence**, *J. Geophys. Res*., 103, 3929–3936, 1998

Kärcher, B., Peter, T., Biermann, U. M., and Schumann, U.: **The initial composition of jet condensation trails**, *J. Atmos. Sci*., 53, 3066–3083, 1996.

Kaufmann, S. / Voigt, C. / Heller, R. / Jurkat-Witschas, T. / Krämer, M. / Rolf, C. / Zöger, M. / Giez, A. / Buchholz, B. / Ebert, V. / Thornberry, T. / Schumann, U. , **Intercomparison of midlatitude tropospheric and lower-stratospheric water vapor measurements and comparison to ECMWF humidity data**, 2018, *Atmos. Chem. Phys*. , Vol. 18, No. 22, p. 16729-16745

Kiemle, C. / Schäfler, A. / Voigt, C. , **Detection and analysis of water vapor transport.**

2012, *Atmospheric Physics: Background -- Methods – Trends,* Springer Berlin Heidelberg: Berlin, Heidelberg, p. 169-184

Krüger, K. / Schäfler, A. / Wirth, M. / Weissmann, M. / Craig, G. C., **Vertical structure of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection to mixing processes**, 2022, *Atmos. Chem. Phys.* , Vol. 22, No. 23, p. 15559-15577

Penner, J. E., Lister, D. H., Griggs, D. J., Dokken, D. J., and McFarland, M. (Eds.): **Aviation and the global atmosphere,** *Cambridge University Press*, Cambridge, UK, 1999

Schmidt, E.: **Die Entstehung von Eisnebel aus den Auspuffgasen von Flugmotoren**, Schriften der Deutschen Akademie der Luftfahrtforschung, 44, 1–15, 1941.

Schumann, U., Ström, J., Busen, R., Baumann, R., Gierens, K., Krautstrunk, M., Schröder, F. P., and Stingl, J.: **In situ observations of particles in jet aircraft exhausts and contrails for different sulfur-containing fuels**, *J. Geophys. Res.*, 101, 6853–6869, 1996.

Schumann, U., Arnold, F., Busen, R., Curtius, J., Kärcher, B., Kiendler, A., Petzold, A., Schlager, H., Schröder, F., and Wohlfrom, K.-H.: **Influence of fuel sulfur on the composition of aircraft exhaust plumes: The experiments SULFUR 1–**7, *J. Geophys. Res.*, 107, 4247, doi:10.1029/2001JD000813, 2002.

Schumann, U.: **Formation, properties and climate effects of contrails,** *C. R. Phys*., 6, 549–565, 2005.

Teoh, R. / Schumann, U. / Majumdar, A. / Stettler, M. E. J., **Mitigating the climate forcing of aircraft contrails by small-scale diversions and technology adoption**, 2020, *Environ. Sci. Technol.* , Vol. 54, No. 5, p. 2941-2950

Teoh, R., Schumann, U., Gryspeerdt, E., Shapiro, M., Molloy, J., Koudis, G., Voigt, C., and Stettler, M. E. J.: **Aviation contrail climate effects in the North Atlantic from 2016 to 2021**, *Atmos. Chem. Phys*., 22, 10 919–10 935, 10.5194/acp-22-10919-2022, 2022.

Reply to Reviewer #2
(Referee comment on "Correction of temperature and relative humidity biases in ERA5 by bivariate quantile mapping: Implications for contrail classification" by K. Wolf et al. (egusphere-2023-2356), https://doi.org/10.5194/egusphere-2023-2356-RC2, 2023)

We thank the Reviewer for the time she/he spent on the manuscript and for the useful comments. Addressing the comments has improved the manuscript.

For better legibility, the Reviewer's comments are highlighted in **bold** and changes in the manuscript are in *italics*.

General Comments:

The authors compare ERA5 temperature and relative humidity at pressure levels of 175hPa to 300hPa with collocated IAGOS observations. A slight cold bias and a dry bias are found that peak at 200hPa and 250hPa, respectively. A bias correction method is applied that reduces the cold and dry bias. Finally, the impact of the correction on the probability of contrail formation is analyzed and compared with simpler corrections used in the literature. The probability of non-persistent and persistent contrail formation in uncorrected ERA5 data are found to agree quite well with the respective estimates from IAGOS data. This agreement improves very slightly for persistent contrail formation (correction reduces the probability of contrail formation) and deteriorates for non-persistent contrail formation after the correction of ERA5 data.

My main objection to the paper is that I am not convinced of the UT dry bias in ERA5 (see my point 1 below). I appreciate that the authors are careful to compare ERA5 and IAGOS data on a similar scale but I worry that the resulting (close to in-cloud) RH PDF (their figure 6) is unphysical. Nevertheless, as the authors point out, many studies in the literature have used a 'correction', often multiplying ERA5 RH with a constant factor. The factor is often determined by comparing ERA5 RH interpolated onto the aircraft track to IAGOS. The present study is designed to call into question such approaches and additionally shows that if the comparison between ERA5 and IAGOS is performed on a similar scale, the correction that can be inferred has close to no impact on contrail formation.

We thank the Reviewer for this faithful summary of our work.

Major comments:

1.)The ERA5 dry bias in extratropical UT RH:

1.1)The authors cite Dyroff et al. (2015) and Bland et al. (2021) as showing the LS moist and cold bias and Kunz et al., 2014; Gierens et al., 2020; Schumann et al., 2021 as presenting the UT underestimation of water vapor concentrations and ice supersaturation in ERA-interim and ERA5.

1.2) Kunz et al. (2014) analyzed ERA-interim and operational IFS data from the years 2001 to 2011 and finds a LS moist and UT dry bias. They note that the agreement with measurement data was improving within the analyzed time period due to model

updates. Bland et al. (2021) evaluate the ERA5 UT humidity fields using radiosonde data and find a slight moist bias or no bias depending on the type of radiosonde. Krüger et al. (2022) report a slight moist bias in the ERA5 UT when comparing to the active Differential Absorption Lidar (DIAL) WAter vapour and Lidar Experiment in Space (WALES; Wirth et al., 2009). This slight moist bias in water vapor mixing ratio coupled with the cold bias should lead to an UT moist bias in relative humidity (RH). Gierens et al., 2020 and Schumann et al., 2021 both compare ERA5 to MOZAIC / IAGOS RH data partly at relatively low resolution and interpolate ERA5 data to the flight track which leads to comparing observed and simulated RH at different resolutions.

1.3) Based on the above literature I am not convinced that there is an UT dry bias in ERA5.

Subsequently, we address Reviewer 2's comments 1 through 1.3.

Obtaining upper tropospheric / lower stratospheric measurements is an inherently challenging task and inter-comparisons of different instruments and measurement principles exist. However, it is not so easy to compare different studies because the observations or the meteorological data (ERA-interim, IFS operational cycles, ERA5) or the method for the comparison (e.g., use of a tropopause-relative coordinate in Krüger et al. (2022) vs the use of pressure levels in our study) differ. To answer the question of the Reviewer, it is not clear whether ERA5 data is too dry or too moist for conditions close to saturation and/or high cloud cover. To our knowledge, there is no consensus as inter-comparisons lead to contradictory results. We modified the introduction in this regard in several places and would like to direct the Reviewer to the revised manuscript with tracked changes.

We referenced the studies by Kunz et al. (2014), Dyroff et al. (2015), Bland et al. (2021), and Gierens et al. (2020) to show previous attempts that evaluated the performance of ECMWF products (ERA-interim, and the operational forecast). In those studies various observations have been used, ranging from aircraft measurements from dedicated campaigns, radiosonde observations, to satellite products.

Indeed, the recent paper by Krüger et al. (2022) shows a wet bias in ERA5 on average on a wide range of meteorological conditions and seasons. However, Krüger et al. (2022) also found a minor wet bias below the thermal tropopause, which would support our results. Having said that, caution is needed when comparing our study to Krüger et al. (2022) because they use the thermal tropopause as the vertical reference, while our investigation uses fixed pressure levels. Variations in the altitude of the thermal tropopause mean that it is difficult to directly compare both studies. One also needs to consider potential uncertainties and biases in the LIDAR / WALES retrieved relative humidity that is presented in Krüger et al. (2022). Lidar is a remote sensing technique, and the retrieved relative humidity relies on temperature fields that are inferred from ERA5 reanalysis data (Groß et al., 2014), which itself could introduce a small bias.

In this regard we would like to mention the study from Reutter et al. (2020), who compared IAGOS with the ERA5 predecessor ERA-interim, and found a slight dry bias below the tropopause, which would also support our findings. They also found that a reduction in the ice-supersaturation threshold from the physical 100% (w.r.t liquid water) to 85% (lower stratosphere) or 95% (upper troposphere) – to compensate a bias at around 100% - lead to an improved estimation of the potential contrail formation.

Another sampling-specific factor that needs to be considered are targeted measurement campaigns. While Krüger et al. (2022) investigated a large variety of meteorological conditions, the investigated measurement campaigns still target specific weather patterns. This selected sampling might bias the data set to certain atmospheric and cloud conditions. A similar problem might influence the IAGOS data set, as pilots tend to avoid clouds during the flight (Petzold et al., 2020).

While we cannot rule out that IAGOS data themselves are biased, comparisons to more exact instruments do not indicate a bias. Here we refer to the validation study by Petzold et al. (2020), who compared CARIBIC and MOZAIC measurements (both predecessor of IAGOS) with dedicated humidity measurements. Petzold et al. (2020) demonstrated a good statistical agreement of CARIBIC and MOZAIC measurements with respect to the dedicated humidity measurements. They found only small differences around 100 % relative humidity, which they argue can be explained by differences in sampling (Petzold et al., 2020). A more recent study by Konjari et al (2022) showed that, from the instrumentation and data quality side, IAGOS measurements agree very well with other instruments' data, such as the airborne Fast In-situ Stratospheric Hygrometer (FISH, Meyer et al. 2015), in the upper troposphere. We also acknowledge that IAGOS is known to have a wet bias in the lower stratosphere and now explicitly mention this in the manuscript. Due to the agreement in the upper troposphere and in spite of the bias in the lower stratosphere, we argue that IAGOS data in the upper troposphere is reliable.

"IAGOS measurements in the lower stratosphere that are typically characterized by low values of relative humidity (≈ rP1 < 20 %), are subject to a moist bias. This moist bias is a non-linear function of the relative humidity and requires a multi-dimensional regression correction that is currently under development (Konjari et al., 2022). Therefore, this known moist bias in IAGOS is not corrected in our analysis and it should be kept in mind that subsequent differences between ERA5 and IAGOS for low values of relative humidity may also be attributable to artifacts in the IAGOS measurements. However, since the focus of this analysis is to investigate contrail formation and persistence, only high values of relative humidity are relevant. Consequently, the moist bias for low relative humidity values in the IAGOS observations has little impact, on our analysis."

In summary, there is a strong rationale for comparing and bias correcting ERA5 with IAGOS because: i) IAGOS data has shown to be reliable; ii) IAGOS samples temperature and relative humidity at exactly at the locations and pressure levels that are relevant for aviation studies; and iii) ERA5 is often used for predicting potential contrail formation.

2.) When comparing IAGOS and ERA5 data you filter the IAGOS data because of their higher resolution. In order to determine the resolution of the comparison you estimate the resolution of the ERA5 data from the grid point distance of the Gaussian grid at the latitude of interest. In a spectral model, such as IFS, this is not the model resolution. The model resolution (grid length at the equator) of ERA5 is 31km. Even then the ERA5 grid box value is of course representative of the 3D volume and the IAGOS observation of a small subset of this volume so that ERA5 should be expected to display a lower variability.

We partly agree with the Reviewer. We now mention that ERA5 is generated from a spectral model that internally operates with an approximate resolution of 31 km. Indeed, smoothing of the IAGOS along-track data lowers the natural variability and reduces extreme values, e.g., low and high relative humidity. However, the applied smoothing

modifies the calculated mean values only slightly. The text sections have been rephrased and now read as follows:

"The fixed (Cartesian) grid resolution of 0.25° of ERA5 does not correspond to a constant longitudinal grid box size in km which instead depends on the latitude. Considering the three sub-domains between 30°N and 70°N, the spatial resolution of one ERA5 grid-box ranges between 24 km (30°N) and 14 km (70°N). Therefore, we assume an average grid box size of 19 km. However, it is emphasized that ERA5 is a spectral model with an internal Gaussian resolution of around 31 km and, thus, the effective resolution is coarser than the Cartesian grid resolution (Hersbach et al. (2020))."

"Smoothing the IAGOS data, as explained in Section 2.2, leads to mean values of T_{P1} and $r_{P1,ice}$ for the native and the smoothed data that are similar by 0.1°C and 1 %, respectively. As the smoothing did not change the mean values significantly, the differences in the PDFs of ERA5 and IAGOS, as well as the bias in mean $r_{ERA,ice}$ compared to $r_{P1,ice}$ cannot be attributed to differences in the spatial resolutions. However, the smoothing of the IAGOS data leads to a reduction in the variability as well as in the extreme values in measured T_{P1} and $r_{P1,ice}$ (not shown here)."

It should be noted that we use ERA5 data that is interpolated from the model levels onto pressure levels (e.g. 250 hPa). We do not think it is appropriate to say that temperature and humidity data are representative of a volume. In this case the vertical resolution is not the same as the vertical sampling. However we compare measurements made in a given pressure layer centered onto a pressure level to this particular pressure level.

To further answer the question about the impact of spatial averaging of IAGOS data, we present IAGOS data at the original 4-s resolution and temporally smoothed data (60 s) in the plots below. The obtained distributions change little between the original (top) and the smoothed (bottom) data and, thus, we argue that averaging IAGOS does not have a significant impact on the obtained statistics.
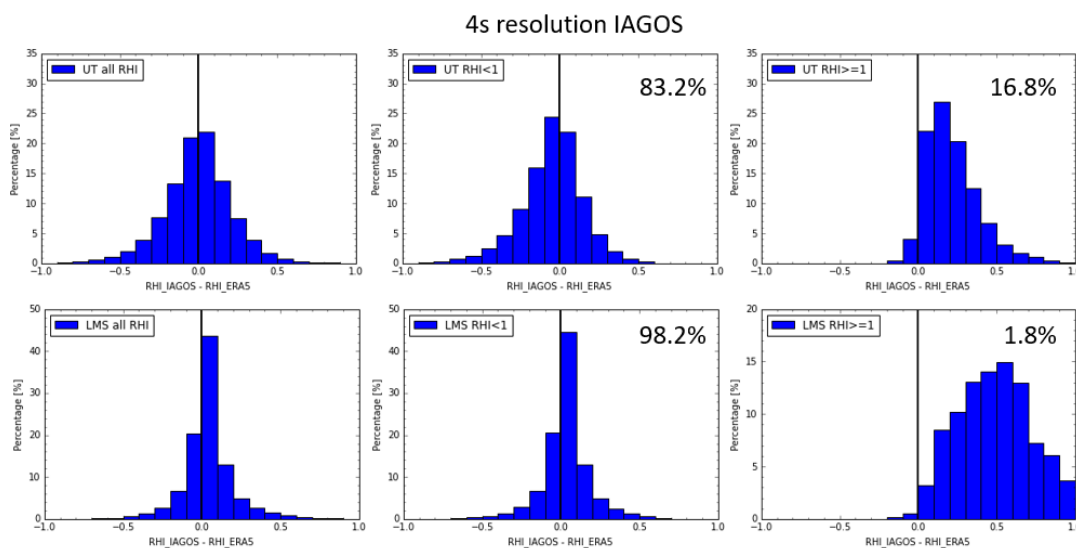
1 min resolution
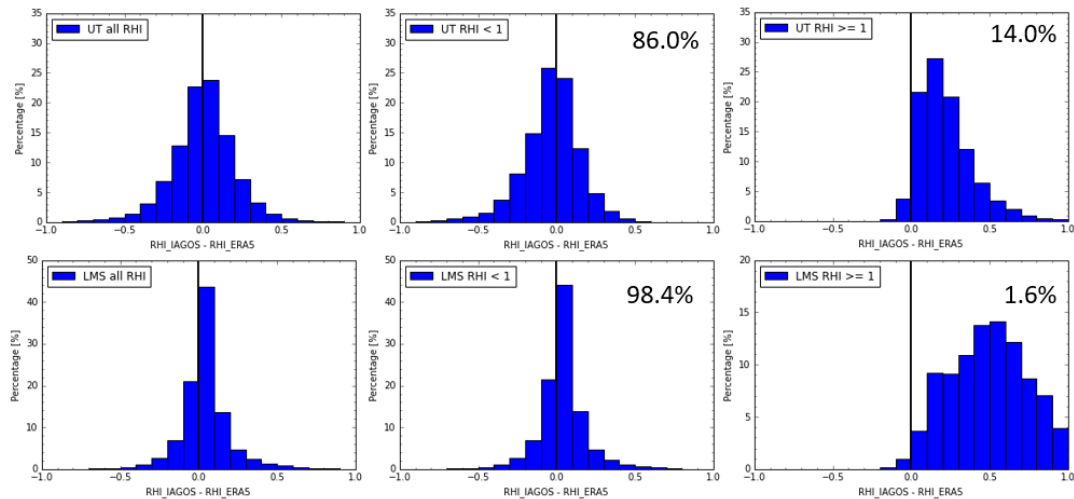
3.) The only reason given for the dry bias in ERA5 RH is saturation adjustment within clouds (section 2.2.1 and in lines 278-279). If this is your hypothesis then I would suggest using the IAGOS particle number concentration N_ice and splitting up the data set into cloudy and cloud free measurements and doing the same for ERA5 (using cloud cover) and then comparing the RH CDF for cloudy and cloud-free instances separately. The comparison could also be extended to include the data of Krämer et al (2009, 2020) that show an increased probability at RH = 100% (unlike IAGOS data). Corrections could and should be done for cloudy cases only.

3.1) You do something similar in section 3.2 but sort both ERA5 and IAGOS data dependent on ERA5 cloudiness into in-cloud and cloud-free data. Since you do not check in figure 6 whether IAGOS measurements are representative of cloudy or cloud-free conditions (see also my comment 4) the differences between ERA5 and IAGOS PDFs worsen with increasing ERA5 cloud cover. Since you have calculated already a 'IAGOS cloud cover' it is not clear to me why you use it only in order to correlate the IAGOS and ERA cloud cover and don't do the IAGOS in-cloud with ERA in-cloud comparison.

3.2) Nevertheless, it is also not clear to me why a correction of in-cloud RH is needed since contrail formation studies are mainly interested in contrail formation within cloud free air.

In this section, we answer comments 3 through 3.2.

Intrigued by the suggestion of Reviewer 2, we performed additional analysis and modified Figure 6 and its corresponding discussion in Section 3.2 "Distribution of relative humidity under cloud-free and in-cloud conditions".

We now separate ERA5 simulations by cloud fraction (CF), where CF < 0.2 are considered as cloud-free, 0.2<CF<0.8 as intermediate, and 0.8<CF as cloudy or in-cloud. Similarly, the IAGOS observations are separated for cloud-free, intermediate, and in-cloud measurements using the detection threshold for cloud particle number concentration given by Petzold et al. (2017).

We replaced the old figure with a new plot that is also given here. Filtering for cloud-free conditions results in similar PDF for ERA5, corrected ERA5, and IAGOS measurement for RHi< 90%. Larger RHi are observed much more frequently compared to ERA5. The QM-correction partly enhances RHi such that the PDF approaches the one from IAGOS. For intermediate cloud conditions the distributions form ERA5 and QM-corrected ERA5 are similar, however, there is a better agreement of ERA5 QM-corrected data with the IAGOS observations. Both, ERA5 and ERA5 QM-corrected data miss the variability towards RHi< 80%. For data points considered to be in-cloud, the PDF of IAGOS is much broader compared to ERA5 and ERA5 QM-corrected data. The QM-correction broadens the distribution of enhancing the right wing of the distribution towards larger RHi values. Considering the three different categories, there is an improvement in the statistical distribution of RHi after the QM-correction. Particularly important is the improvement in RH representation in the intermediate category, where clouds are most likely to form.

The newly added analysis is described in section "Distribution of relative humidity under cloud-free and in-cloud conditions". Due to the length of the revised section, we refer the Reviewer to the revised manuscript with track changes.
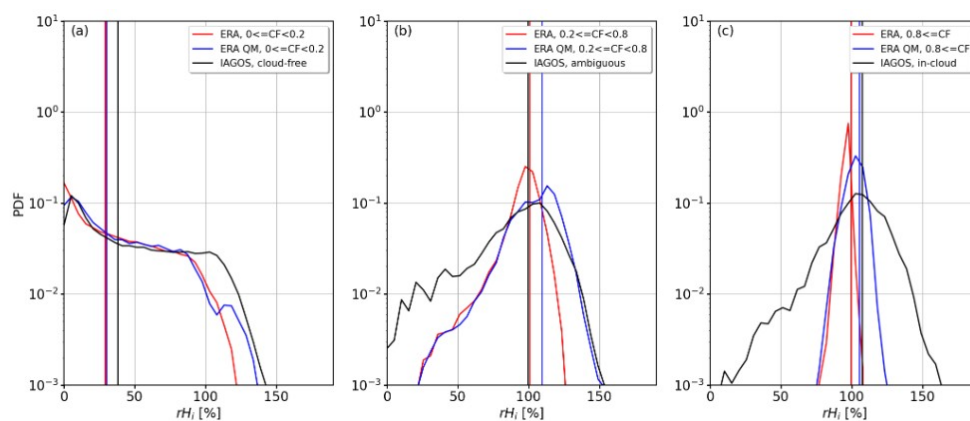

Figure 6. a–c Probability density functions of relative humidity rice w.r.t ice (in %) from IAGOS (black), original ERA5 (red), and corrected.

4.) I believe there could be another reason for differences in the PDF of RH between IAGOS and ERA5 which is connected with sampling. Petzold et al. (2020) discuss sampling issues in their figure 5 when comparing MOZAIC measurements with measurements from research flights (Krämer et al. 2009). They say that the reason why the research flights show a much higher probability of RH around 100% than IAGOS is because campaign measurements often target clouds in which RH is often around 100%. In the same way, it is likely that IAGOS pilots tend to avoid clouds and rather fly through cloud free air or very thin clouds while ERA5 data represent cloudy, partly cloudy and cloud free situations purely based on their probability of occurrence.

We agree with the Reviewer. In addition to mentioning the sampling issue in the introduction and the IAGOS data section, we further highlight this potential shortcoming later in the text.

The following text was added to the introduction.
"In situ measurement campaigns are a potential fourth approach, during which contrails are directly probed and contrail properties are investigated. Dedicated measurement campaigns, for instance by Krämer et al. (2009, 2020) and Voigt et al. (2017), are rare.

Furthermore, they may lack spatial representation by targeting specific atmospheric features as well as cloud conditions, which may bias the results (Petzold et al., 2020). "

The following text was added to the section "In-service Aircraft for a Global Observing System":
"In addition, the sampling is biased, i.e., by avoiding severe weather and by avoiding or favoring specific atmospheric circulation patterns, such as the jet stream (Petzold et al. (2020))."

5.) Figure 6 shows that the ERA QM data display a maximum probability of RH of ~105% for close to full cloud cover and of ~115% for cloud cover of between 0.6 and 0.8. Uncorrected ERA data show that RH = 100% has the highest probability in cloudy situations in line with Krämer et al. 2009 and 2020 and in line with theory that predicts significantly large in-cloud supersaturation only if supersaturation forcing is very large or if ice crystal number concentrations are very small. Both conditions are not the most probable inside clouds. To be fair, the other methods that you mention within your paper, e.g. scaling up RH with a constant factor, will have exactly the same (or possibly a worse) impact on in-cloud RH.

To answer this comment, we would like to direct the Reviewer to the answers to comment 3 and the revised Section 3.2.

While it might be true that such conditions are rare, the conclusions given in Krämer et al. (2009, 2020) are based on sophisticated measurements in terms of instrumentation and accuracy. In contrast, IAGOS are routine measurements onboard passenger aircraft that are reliable but cannot achieve the same accuracy. They sometimes provide information on particle number but only some IAGOS aircraft were equipped with particle counters. Selecting only those measurements where the particle number concentration is provided would drastically reduce the available dataset. The advantage of IAGOS data is the large number of measurements and the relatively uniform spatial distribution, considering the constraints given by Petzold et al. (2017).

6.) I am surprised that the averaging of IAGOS data does not change the RH PDF. Averaging should always decrease the variability of data and I would expect the probability of very high ice supersaturation to be reduced. Since this result is used as an argument to claim that mixing ratios are too low I think the figure should be included. Note also that the result that averaging of IAGOS does not improve the comparability to IAGOS data is in contradiction to Reutter et al. (2020) who finds that the IAGOS data show a very high percentage of small-scale ice supersaturated areas which ERA-interim cannot resolve but that ERA-interim and IAGOS fit well once the IAGOS resolution is reduced to the resolution of ERA-interim.

To answer this question, we would like to direct the Reviewer to our answer to major comment 2 and the provided figure. The presented figure shows distributions of relative humidity for the raw measurements with 4-s temporal resolution as well as distributions based on 1-minute averages. Even after averaging over such a long period the distributions did not change significantly. There is an exception for extreme values but they contribute only a very small fraction to the total number of samples.

7.) Despite the fact that I am not convinced of the need to 'correct' ERA RH data, I still think that your work points out an important point in relation to the earlier attempts to study contrail formation using ERA5 data. You show that if you take

care to compare ERA5 and IAGOS data on a similar scale that correction has hardly any impact on the contrail formation probability. This means that the ERA5 RH scaling of earlier studies that is calculated by interpolating ERA5 data to the aircraft track and comparing to IAGOS data, lead to an increase in contrail formation probability that is not supported by IAGOS data.

We thank the Reviewer for this comment. Concerning the bias in ERA5 we would like to direct the Reviewer to our answers to comments 1 and 3.

Minor:

1.) The text is generally difficult to read because it is often not clear which figure is being discussed. E.g. at the beginning of section 3.1 you mention fig. 2 and fig. 5e in the first 10 lines but actually you are discussing probably figure 3 which you fail to say. This problem is repeated in other places.

We partly agree with the Reviewer. The reference to Fig. 2 is correct. In the sentence before the reference we discuss the number of flights / measurements per pressure layer. The distribution of flights per pressure level is shown and indicated in Fig. 2.
The reference to Fig. 5e is also correct as we discussed the reduction in the mean deviation of temperature, which is given in Fig 5e. However, we added an additional reference to Fig 3. The sentence now reads the following:

"[...] distributions of T_{P1} (see second column in Fig. 3 and Fig. 5e)."

Several further annotations have been made in the text by adding references to direct the reader to the respective figures.

2.) line5 (abstract): You extract wind speed but I can't find the place where you use this variable.

The Reviewer is right. "Wind speed" accidentally remained from an earlier revision of the text. Wind speed has been removed and the sentence reads as following:

"IAGOS flight trajectories are used to extract co-located meteorological conditions from ERA5, namely temperature and relative humidity, which are compared with the IAGOS measurements."

3.) Line 31: You did not define 'WC'.

"WC" was a typo and is now corrected. The correct abbreviation is WV, which is defined in the line above.

4.) Line 59: Bickel et al. (2020) discusses the difference between radiative forcing and effective radiative forcing. The model that is used in Bickel et al. is described in Bock and Burkhardt (2016).

We thank the Reviewer for pointing this out and we took the suggestion to better explain the different definitions of radiative forcing earlier in the manuscript. The revised and new section now reads:

"[….]The influence of a perturbation, e.g., clouds, aerosols, or gases, on the Earth's atmosphere and its radiative transfer is quantified by the radiative forcing (RF). By definition, RF is defined as the difference in the net irradiance at the top of atmosphere under perturbed and unperturbed conditions (Ramanathan et al., (1989)). In the context of climate studies the RF is understood as the difference of the Earth energy budget due to a contributor to climate change (Bickel et al., 2020). For example, the aviation-induced global CO_2-related RF is estimated to be around 30 mWm^{-2} (Boucher et al., 2021). Contrail RF is estimated to be stronger, at about 60 mWm^{-2} but is subject to much larger uncertainties (Burkhardt and Kärcher, 2011)."

5.) Line 67: Instead of the word 'distribution' you mean probably 'variability'.

The Reviewer is correct and the words have been exchanged. The sentence now reads:

"[...] which is generally challenging due to the high temporal and spatial variability of WV."

6.) Line 68 ff: Please include Krüger et al (2021) and improve the description of findings of the papers (see above).

We revised the section of the text and added the reference to Krüger et al. (2022). The section has been modified to the following:

"Specific issues have been identified in the upper troposphere and lower stratosphere, as well as with the general representation of ice supersaturation. For example, Bland et al. (2021) compared radiosonde observations with operational ECMWF IFS weather forecast and identified a lower stratosphere moist bias. Similarly, Krüger et al. (2022) compared measurements from a differential absorption Lidar with ECMWF ERA5-reanalysis data (on a relative-tropopause coordinate) and identified a slight moist bias in the upper troposphere. A moderate to significant moist bias was found in the lower stratosphere. Contrarily, studies that compared water vapor concentrations and ice supersaturation in ERA-interim and ERA5 with aircraft in-situ observations found that ice supersaturation is insufficiently represented (in frequency and magnitude) in those re-analysis products (Kunz et al., 2014; Dyroff et al., 2015; Gierens et al., 2020; Reutter et al., 2020; Schumann et al., 2021). Consequently, there is no consensus whether ECMWF re-analysis products are subject to a moist or dry bias in the upper troposphere."

7.) Line 75 ff: You may want to include Krämer et al. (2009, 2020)

We followed the suggestion of the Reviewer and added the two references to the manuscript.

8.) Line 196: The multiplication of ERA5 RH with a factor is only common to the studies that you cite afterwards and not generally common.

We followed the Reviewer's comment and made this point clearer by highlighting that scaling is primarily applied in studies from, e.g., Schumann et al. (2013) or Schumann et al. (2015), and is not common practice.

"To compensate for the dry bias in ERA5 for contrail detection applications, $r_{ERA,ice}$ values are scaled sometimes in some studies by multiplication factors between 0.8 or 0.9, particularly in Schumann and Graf (2013) and Schumann et al. (2015)."

9.) Line 276-277: I am not sure what you are talking about. I don't see any rectangular shapes.

Thank you for mentioning this. The word rectangular was replaced by triangular and the text now reads:

„*The PDF of r_{ERA},ice close to 100 % is characterized by a triangular shape, while the distribution of [...]*"

10.) Line 340-341: 'r_ice close to 100% are likely associated with cloud formation' – no this is not the case. Cirrus formation happens at high relative humidity relative to ice.

This section the sentence has been removed during revisions.

11.) Line 366-367: 'All data points that do not belong to any of the categories are flagged for no contrail formation (NoC).' Isn't you R category also no contrail formation?

Agreed. We removed the word "formation" from the sentence that was highlighted by the Reviewer. The sentence now reads:

"The SAc and the ISS threshold are used to flag the IAGOS measurements and the along-track ERA5 for NPC, PC, and R conditions. Samples that belong to none of these three categories are flagged as "no contrails" (NoC)."

12.) Line 395-396: please reformulate the sentence 'but the saturation is insufficient to reach supersaturation to …'

In the course of the revision of the manuscript this sentence has been removed.

13.) Line 671-672: 'The mode in Fm,h is primarily caused by the saturation adjustment in ERA5 (see Sec. 2.2.1)'. As said above, Krämer et al shows that the mode around 100% RH is a naturally occurring mode and not a modelling feature.

We agree and have rephrased the sentence to the following.

"The mode in Fm,h is a superposition of two effects. While the peak is of natural origin, as reported by Krämer et al. (2016, 2020), it is also caused by the saturation adjustment in ERA5 (see Section 2.2.1)."

References
Groß, S. / Wirth, M. / Schäfler, A. / Fix, A. / Kaufmann, S. / Voigt, C. , **Potential of airborne lidar measurements for cirrus cloud studies**, 2014, *Atmos. Meas. Tech.* , Vol. 7, No. 8, p. 2745-2755

Meyer, J. / Rolf, C. / Schiller, C. / Rohs, S. / Spelten, N. / Afchine, A. / Zöger, M. / Sitnikov, N. / Thornberry, T. D. / Rollins, A. W. / Bozóki, Z. / Tátrai, D. / Ebert, V. / Kühnreich, B. / Mackrodt, P. / Möhler, O. / Saathoff, H. / Rosenlof, K. H. / Krämer, M., **Two decades of water vapor measurements with the FISH fluorescence hygrometer: a review**, 2015, *Atmos. Chem. Phys.* , Vol. 15, No. 14, p. 8521-8538