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.

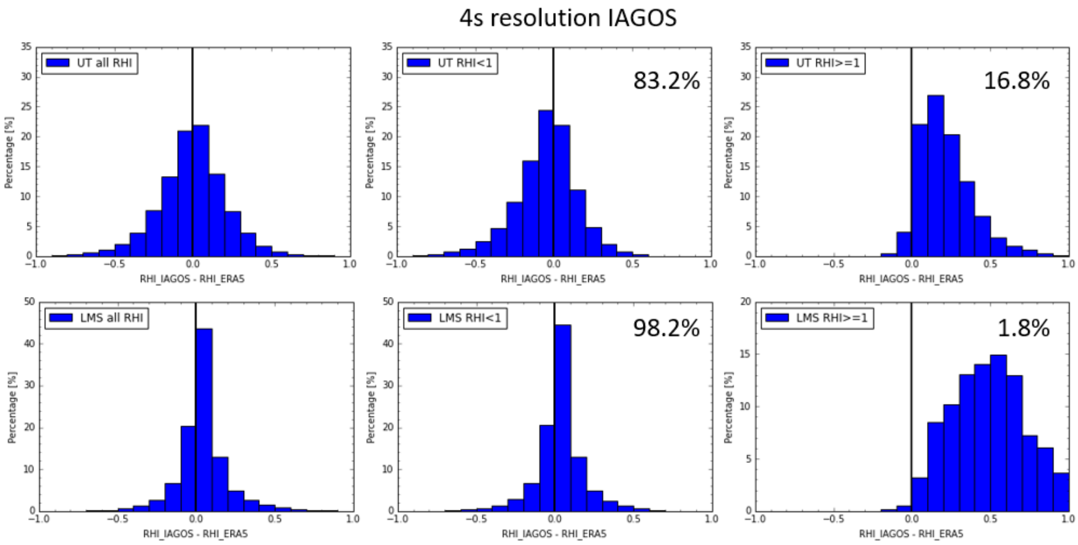
We partly agree with the Reviewer. We now mention that ERA5 is generated from a spectral model that internally operates with an approximate resolution of 31 km. Indeed, smoothing of the IAGOS along-track data lowers the natural variability and reduces extreme values, e.g., low and high relative humidity. However, the applied smoothing modifies the calculated mean values only slightly. The text sections have been rephrased and now read as follows:

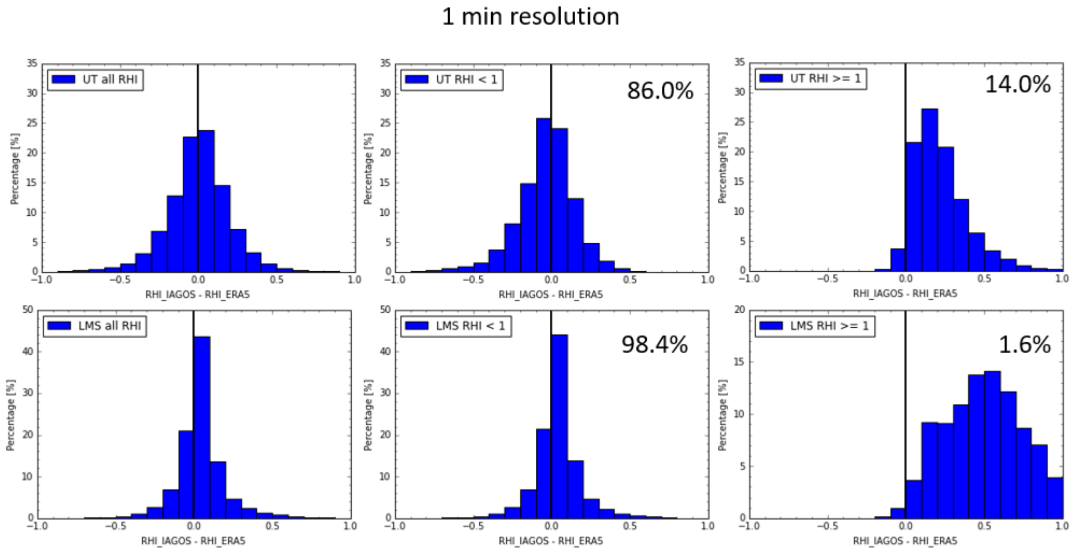
“The fixed (Cartesian) grid resolution of 0.25° of ERA5 does not correspond to a constant longitudinal grid box size in km which instead depends on the latitude. Considering the three sub-domains between 30°N and 70°N, the spatial resolution of one ERA5 grid-box ranges between 24 km (30°N) and 14 km (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.





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.

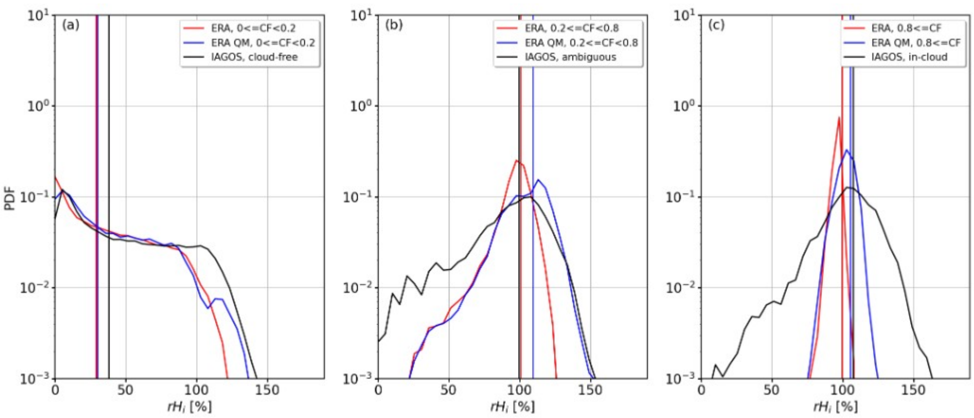


Figure 6. a–c Probability density functions of relative humidity rice w.r.t ice (in %) from IAGOS (black), original ERA5 (red), and corrected.

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