Reply to Reviewer #1 (3rd round)
(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), 2023)

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

For better legibility, the Reviewer's comments are highlighted in **bold**, comments from our side are in normal font, and changes in the manuscript are in *italics*.

This is my first review of "Correction of temperature and relative humidity biases in ERA5 by bivariate quantile mapping: Implications for contrail classification" by Kevin Wolf et al. The authors present a thorough analysis of a correction of ERA5 relative humidity and temperature using a quantile mapping approach, with the aim to obtain an improved assessment of the contrail formation potential. The presented figures are nice and clear and thoroughly explained. Especially the illustration of the effect on the contrail formation potential based on the Schmidt-Appleman criterion is well-made.
After the preceding rounds of reviews the content of the paper is in good shape. However, I have a few comments that I hope can help the authors to improve their manuscript and, especially, to make it easier to access the relevance of the work.

1) I think you should better clarify the purpose of the work and add more information about the added value. In some places this information is simply missing (abstract, summary) and in others unprecise formulations caused confusion (introduction).
The first sentence of the abstract implies a different topic to be addressed. In combination with the word "prediction" in the title, I anticipated that forecast skill of T and RH is addressed (see also comments below), which is not the case. Only a few lines later the topic shifts to contrails. I recommend better introducing the key topic of this work. After that, partly redundant details are linked without a conclusion. So, I suggest better connecting these sentences and adding a precise conclusion about the added value of this work, which I believe is a more reliable identification of the contrail regions using post-processed ERA5 T and TH fields.
We agree with the Reviewer. We have restructured and rephrased the abstract to avoid giving the impression that we are trying to correct ERA5 and state clearly that we instead are post-processing the ERA5 temperature and relative humidity fields to obtain a better estimate of potential contrail formation.
We also had to significantly reduce the length of the abstract to meet ACP's word limit of 250 words.
Due to the comprehensive changes, we would like to direct the Reviewer to the track-changes version of the revised manuscript to review the modifications.

The title "correction (…) in ERA5" may imply that ERA5 is improved within the model, which is not the case. It is rather a post processing with the aim to improve temperature and rHi to better analyse the potential of contrail formation. In addition, I struggle with the word "prediction" as no forecast data are used in this study. I would be careful when using "prediction" in the context of NWP. Hence, I would suggest changing the title to e.g. "Correction of ERA5 temperature and relative

humidity biases (…) for contrail analysis and classification".
We followed the suggestion of the Reviewer and changed the title to the following:
"Correction of ERA5 temperature and relative humidity biases by bivariate quantile mapping for contrail formation analysis."

In the introduction a clear introduction of the approach and goals are missing. Different approaches to investigate contrail formation/impact are listed and mixed with the explanation of the chosen approach. Especially after L62 the structure is not clear: Approach three on contrail modeling is followed by ERA5 and its biases. The fourth approach, which starts with campaign/IAGOS measurements, is abruptly followed by ERA5 correction attempts (L92 "Comparing and bias correcting"). I do not understand the causality to i, ii, and iii. Additionally, both approaches overlap somehow. The introduction contains all information, but should be better structured and more precisely formulated with a focus on your own work and open questions. For what is the correction actually needed?
The Reviewer is right and we restructured the introduction to be clearer. We also have refrained from mentioning different approaches as it is better to separate into measurements (in-situ and remote) and simulations by models.
We also added a section at the end of the introduction to better articulate the goal of the manuscript and the application of the QM-correction. We would also like to direct the Reviewer to the track-changes version of the revised manuscript for a full overview of the modifications.
The added text reads as following:
"The QM-correction aims to remove possible temperature and humidity biases in ERA5 post-processed data to better estimate the contrail formation potential beyond the common locations of the IAGOS flight tracks. The advantage of such a correction is that ERA5 data away from the IAGOS flight tracks can be used to estimate the large-scale contrail formation potential, thus providing a broader perspective on potential contrail formation in space and time over the Atlantic. Potential applications include the study of temporal and spatial patterns of contrail formation, and the development of rerouting options based on statistical distributions of contrail formation potential."

In the summary section, I wonder about the added value of the presented corrected ERA5 data and the applied methods? Can the method be used to correct grid points away from IAGOS flight routes and to actual prediction data in order to support flight routing to avoid contrail formation? What is the added value to the IAGOS observations alone? Who could make use of such data? What are the ways forward?
We do agree with the Reviewer and added the following sentences to the summary:
"[…] Overall, the presented QM–correction allows to remove the systematic bias in temperature and relative humidity in ERA5 using IAGOS as a reference. Therefore, the method can be applied to ERA5 data to estimate the contrail formation potential away from IAGOS flight tracks under the constraint that the correction is applied to grid points within a specified domain between 105°W and 30°E and from 30°N to 70°N. This provides a broader perspective on potential contrail formation in space and time over the Atlantic region. This allows to study of temporal and spatial patterns of contrail formation over the North Atlantic region to develop statistically based rerouting options."

2) I think that the ERA5 reanalysis and applied methods could be explained more clearly:
In the introduction (L62ff) the explanation of ERA5 is unprecise (e.g. "the prediction and reanalysis"). ERA5 is based on a particular version of the IFS that is kept constant over the time period from 1940. The constant version including the data assimilation system allows trends to be better accessed and separated from model changes. Make sure that the reader understands the difference between analyses, reanalysis and predictions (see also L94). I suggest in restricting the discussion to the terminology reanalysis. It is not clear how the accuracy of relative humidity and temperature can be compared ("less accurate").
All instances, where the word "prediction" was used incorrectly, were removed. We also removed the statement about the lower accuracy of relative humidity and rephrased the sentences to the following:
"Higher uncertainties with respect to IAGOS observations were found in the re-analysis of relative humidity, which is generally challenging due to the high temporal and spatial variability of WV. Specific issues have been identified in the upper troposphere and lower stratosphere, as well as with the general representation of ice supersaturation."

In L68ff, there is an abrupt transition from the simulations of contrail occurrence to "A frequent source of information (…)". What does the latter mean and how are these parts connected? I think ERA5 has no dedicated information about contrails and that should be clarified. Avoid contrail representation and speak e.g. of ability to estimate the contrail formation.
We followed the suggestions of the Reviewer. All instances, where prediction was used incorrectly, have been changed to "estimation".
The discussion about contrails models and the use of ERA5 data is now better linked. Furthermore, the role of ERA5 is better explained. The following text has been added to the introduction.
"Offline models such as CoCiP require meteorological data, e.g., temperature and humidity, as input. A well established data set of meteorological data is provided by ERA5 (Hersbach et al., 2020), which stems from a state-of-the-art global modeling system of the European Centre for Medium-Range Weather Forecasts (ECMWF) and a large number of observational data streams. The ERA5 output is based on simulations with a specific, constant version of the Integrated Forecasting System (IFS) of ECMWF. Thus, the ERA5 data set from 1940 to present can provide some insight into trends in the Earth's atmosphere."

In the discussion of model biases in ERA5 (LL72ff) absolute and relative humidity biases are mixed. Studies like Bland et al. or Krüger et al. and possibly a few others discuss absolute humidity biases. This cannot lead to a conclusion "no consensus" with studies looking at rHi in ISSRs. I think there is no knowledge about systematic absolute humidity errors in ISSRs. I find it also confusing that throughout the paper you simply talk about a dry bias (L98, L231). It should be made clear that and where you refer to rHi.
We only partly agree with the Reviewer that we mix the discussion about absolute and relative humidity biases in the introduction. While it is true that the studies compared absolute, specific, and relative humidity, within the the cited studies either absolute, specific, or relative humidity from ERA5 were compared with respective observations. Hence, equal properties were compared. Dry and moist biases were identified in absolute, specific, and relative humidity. Therefore, we argue that there is no real consensus on the

sign of the bias. Furthermore, it is known that the supersaturation adjustment in ERA5 has drawbacks in the representation of ice clouds in the subsaturated and supersaturated state (e.g., ECMWF, IFS Docuementaion – Cy47r3, Part IV, 2021).
Based on the available ERA5 data and the IAGOS observations that are available to us, we identified a slight but systematic rHi dry-bias in ERA5 with respect to IAGOS. While we aimed to bias correct rHi in general, the main focus is on ISSRs and to improve the relative humidity distribution in the post-processed ERA5 data to better identify potential contrail formation regions.

The Reviewer is right that we need to be clearer and have to mention that we refer to a bias in relative humidity. We now explicitly call it relative humidity in the text, where we previously wrote about a moisture bias. Due to the scattering of the changes over the length of the manuscript, the individual corrections are not copied here. Instead we like to direct the Reviewer to the track changes version of the revised manuscript.

In the description of ERA5 (Sec 2.2), the "native vertical resolution" should be revised. Neither is the spacing between the pressure levels equal to the resolution nor is the pressure level data native data.
The word "native" is removed and the sentence is rephrased as follows:
"We make use of the full vertical resolution with a 50-hPa spacing between 350 and 300 hPa, and a 25-hPa spacing between 300 and 150 hPa."

In fact, I wonder why the authors did not use the full model level data.
Previous contrail related studies that used ERA5 data in one way also used pressure level data, for example Agarwal et al. (2022), Wilhelm et al. (2022), and Hofer et al. (2024). Also from a aircraft perspective it is convenient to use pressure levels as aircraft fly along constant pressure levels.

Related to that, I cannot follow the justification of the nearest neighbor sampling. RH is not a prognostic variable and I do not understand why T and q cannot be interpolated due to the C-C-Eq. I guess the same is done when interpolating model levels to pressure levels!? I miss a discussion about the treatment of the vertical sampling or interpolation. You talk about levels but sometimes rather mean layers. I guess that a vertical difference of up to 25 hPa, when the vertically nearest grid point is used, may relate to quite a difference in RH as it is known that these ISSRs are shallow and vertical gradients near the TP can be strong. May some of the differences between model and observations be related to that fact? Were sensitivity tests made for different sampling / interpolation strategies? There is no "current version of ERA5" instead ERA5 is, as you mention, using a constant model cycle for creating the reanalysis data set.
We did not perform a sensitivity study on the sampling method in the vertical nor in the horizontal direction.
In section "ERA5" we added "(horizontally and vertically)" to clarify that we used the closest ERA5 value on the Longitude-Latitude grid as well as the value from the closest pressure level. The respective sentence now reads as follows:
".. selecting the ERA5 grid points that are temporally and spatially (horizontally and vertically) closest to the IAGOS observations."

We also removed "current version of ERA5" and the sentence now reads as follows:
"The ERA5 data set was generated with the ECMWF Integrated Forecasting System (IFS)

cycle Cy41r2, which was operational in 2016."

In the description of the quantile mapping (2.3) I do not understand the discussion in L265-268. What is the purpose of this? What do you mean with ERA5 is invariant? Do you mean potential biases?
The presented version of the QM correction relies on a constant bias between model (ERA5) and observation (IAGOS). This means that the bias between ERA5 and IAGOS must be invariant in time. ERA5 is constant in the sense that only one IFS Cycle is used to generate the data set. Thus, the same implementation of dynamics, cloud microphysics, and others are used. However, there may be potential changes due to changes in the observation system that feeds into the 4D-VAR assimilation procedure and those due to a changing climate. Only the latter changes would also be represented in the IAGOS measurements. What may also change over the years are the IAGOS measurements themselves, e.g., through calibration and maintenance procedures, and updated post-processing algorithms.
We have rephrased the text to make this clearer.

"The presented version of the QM-correction assumes a time-invariant bias between model and observations. On the model side, we assume that the ERA5 data set is constant in time, since it is generated only with the IFS Cycle 41r2 and therefore has the same implementation of the dynamical core, cloud microphysics representation, etc. However there may be some changes in the quantity and quality of observations feeding into the ERA5 data assimilation system. The IAGOS reference observations may also vary over time due to changes in instrument calibration and maintenance procedures."

It would be good to have some more information about the training data set and the verification period. What is the strategy behind selecting these time periods? (Why) Is the training and verification period overlapping?
Several sentences in the section "Quantile mapping" have been rephrased to be more precise in the selection of the training and validation data set.
"[…] The subscript 'h' commonly refers to historical data, which can also be understood as training data. The training data makes use of IAGOS measurements from January 2018 to June 2021, as this period was considered to be stable in the IAGOS post-processing. [...]"

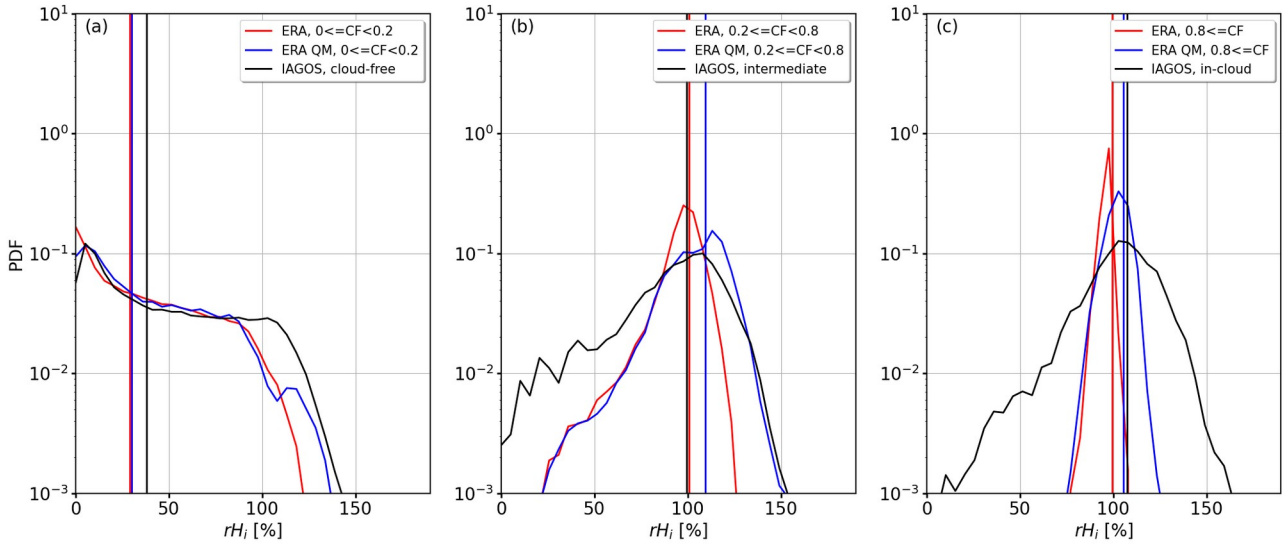
Concerning the selected, overlapping periods we added the following sentences :
"[…] The QM technique is applied to the entire reference period from January 2015 to June 2021, which includes but exceeds the training period. The periods were chosen to a) verify the general applicability of the bi-variate QM-correction method with the same data set and to b) test the stability of the bias correction in years outside the training period."

What data period is shown throughout Section 3?
We have clarified the time period that is analyzed at the beginning of section 3. The first sentence now reads:
"In a first step, along-track temperature and relative humidity from January 2015 to June 2021 from IAGOS and ERA5 are compared in terms of probability density functions (PDFs), mean values, and mean difference (MD). [...]"

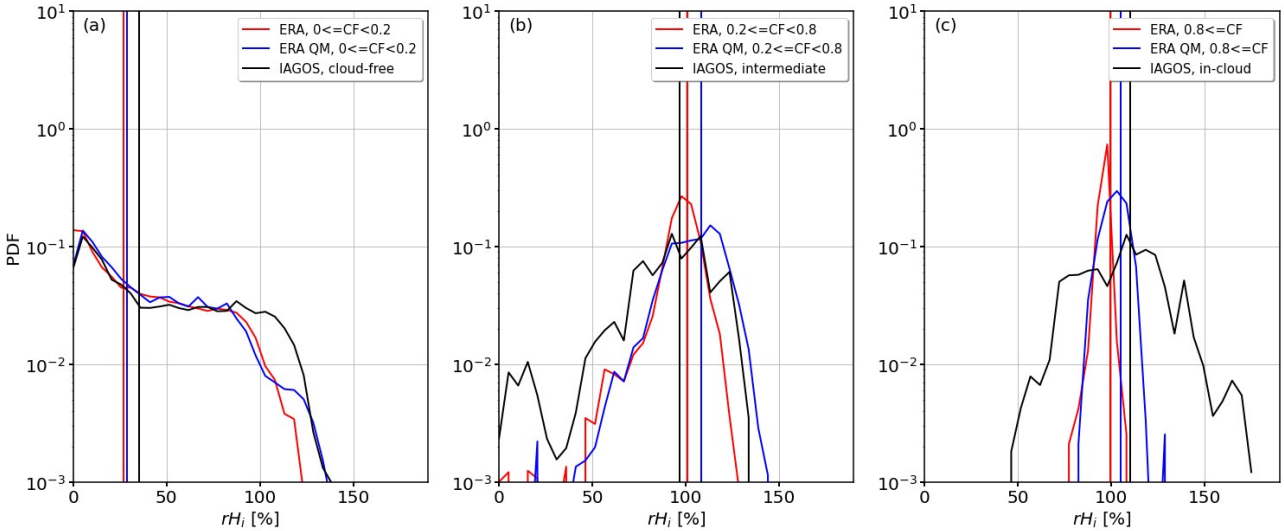
Can the QM method only be used to correct at observed locations or also used at other model grid points?
The model can be used outside the IAGOS flight tracks, but within the study area, which extends from the Eastern US across the North Atlantic to the central EU. We have made this clearer by restructuring and rephrasing the abstract, introduction, and summary. We direct the Reviewer to our responses given below and to the track changes file.

Sometimes you talk about “grid boxes” (e.g. Sec. 3.2) and I wonder if these are the nearest grid points?
The manuscript was screened for the use of the word “grid-box(es)” and, if incorrectly used, the word grid-box(es) was replaced with the word “data points” or an equivalent word. We direct the Reviewer to our responses given below and to the track changes version of the revised manuscript.

I think I understand the purpose of analyzing differences of cloudy and cloud-free condition in Sec. 3.2. However, the reader would profit from a few clarifications. Is it correct that you compare PDFs from completely different data sets? How does a comparison of data points look where ERA5 and IAGOS both have no cloud simulated/observed? To what extent are results influenced by situation of cloud simulated but not observed and vice versa? In addition, I wonder about the use of cloud-coverage? Wouldn’t ice water content be more straightforward to compare cloudy situations.
The Reviewer is partially correct in the sense that we compared different data sets, whereas we would prefer to say that we compared different data points from the same data set.
However, even by requiring cloud-free, intermediate, or cloudy conditions to be present at the same time does not mean that same data is compared as cloud conditions derived from IAGOS and ERA5 do not have to agree in space and time.
However, we have followed the Reviewer’s suggestion and repeated the analysis so that only data points are compared where in-cloud and cloud-free conditions were determined in ERA5 and based on the IAGOS measurements. For comparison, we first provide the plot of the previous version of the manuscript:



and below the updated version with the new analysis:



The distributions for the cloud-free conditions (panel a) remain largely unchanged compared to the plot in the previous version of the manuscript. Differences appear for the intermediate (b) and the cloudy conditions (c). In panel (b), the broad distribution from IAGOS becomes narrower and more similar to the distributions of the original and QM-corrected ERA5 data. While the medians of the original and the QM-corrected ERA5 data remain unchanged compared to the plot in the last version of the manuscript, the median from IAGOS is slightly shifted to lower values by about 3%. In panel (c), the distributions of the original and corrected ERA5 data remain unchanged compared to the plot in the last version of the manuscript. Similar to panel (b), the distribution of IAGOS rHi becomes narrower. However, the IAGOS distribution is still much broader compared to both ERA5 distributions, indicating that neither the original ERA5 nor the QM-corrected ERA5 can represent the natural variability in rHi observed by IAGOS. The medians of the original and QM-corrected ERA5 data are slightly shifted to lower values of rHi shown in the plot in the previous version of the manuscript.

We decided to use the cloud fraction as a criterion, since IAGOS does not provide the ice water content. Furthermore, the saturation adjustment is applied depending on the presence or absence of a cloud. We selected a cloud fraction value threshold of 0.2 to determine cloud-free samples, assuming that these have a high probability of being cloud-free. Similarly, we chose 0.8 to select points with a high probability of cloud cover.

Minor comments:

L82: The sentence “It is noted (…)” should be revised. I guess you don’t want to say that the observations have a bias, right?

We rephrased the sentence to be more specific and to say that the spatial sampling might be biased and not necessarily the measurements itself.

“It is noted that in-situ aircraft observations are potentially biased in terms of spatial sampling because aircraft typically avoid deep convective clouds and the outflow of such clouds.”

L229: Please revise “problematic for contrail and cirrus representation”. Contrails

are not at all represented in ERA5. What is the problem for Cirrus?
The sentence has been clarified and rephrased to:
“While the adjustment approach proved to be suitable for most atmospheric conditions (Gierens et al., 1999; Tompkins et al., 2007; Lamquin et al., 2009). However, the use of ERA5 relative humidity data, which is subject of the adjustment in the context of contrail formation analysis, leads to an underestimation of ISSR in the upper troposphere (Gierens et al., 2020).”

L396 “revels” should be “reveals”?
The typo has been corrected.

L410 What means “smaller”?
The word has been replaced by “narrower”, which is the correct term to describe the width of a distribution. The sentence now reads as following:
“All three distributions of r_{ice} are narrower compared to the cloud-free or intermediate conditions, with $r_{P1,ice}$ being broadest.”

L495: This sentence makes no sense.
The sentence has been rephrased to the following:
“A confusion matrix is a table that is used to visualize the classification performance of an algorithm. Table 4 provides a schematic for a binary event. ”

L577-582: Be more specific about the identified RH biases, especially, whether it is a cold/warm or moist/dry bias. Avoid redundant information.
The Reviewer is correct and we now specify the bias. Redundant information was removed. Due to the length of the section, we would like to direct the Reviewer to the track changes version of the revised manuscript.

Agarwal, A. / Meijer, V. R. / Eastham, S. D. / Speth, R. L. / Barrett, S. R. H., Reanalysis-driven simulations may overestimate persistent contrail formation by 100%–250% 2022-01, Environ. Res. Lett. , Vol. 17, No. 1, IOP Publishing, p. 014045

ECMWF, IFS Documentation CY47R3 - Part IV Physical processes, 2021, IFS Documentation CY47R3 , No. 4, ECMWF

Hofer, S. / Gierens, K. / Rohs, S., How well can persistent contrails be predicted? An update, 2024, Atmos. Chem. Phys. , Vol. 24, No. 13, p. 7911-7925

Wilhelm, L. / Gierens, K. / Rohs, S., Meteorological conditions that promote persistent contrails, 2022, Applied Sciences , Vol. 12, No. 9