

# Response to Reviewer #1

We would like to thank the anonymous Reviewer for their careful review. We have taken all of the comments into account, improving the accuracy and presentation of the results in our manuscript. Below is our response to the comments on a point-by-point basis. The questions and remarks from the Reviewer are recalled in italic.

## *## Overall*

*This is a nice technical note expanding on the work of Wang et al 2025. I think the manuscript deserves to be published after improving the clarity of the presentation and considering the major comments below. Ideally the manuscript would be accompanied by example training code in the author's language of choice.*

We thank the Reviewer for the overall positive assessment of our manuscript. We will attach to the article an example training code written in python.

## *## Major Comments*

*- L3 & L46: Many publications on this topic often make some form of the statement "[There are] considerable errors in RH<sub>i</sub> estimates" which makes "accurate forecast of ISSRs [difficult]." I'm interested to see more analysis on the type and distribution of errors to better understand how ISSR forecast errors will result in ineffective (or inefficient) avoidance measures. In our experience, RH<sub>i</sub> (and ISSRs) have high pointwise error, but overall ISSR regions are (generally) spatially and temporally correlated with ISSR forecasts.*

The Reviewer emphasises an essential distinction between correctly predicting RH<sub>i</sub> and correctly predicting ISSRs. It is true that errors on the (vertical and horizontal) location of ISSRs may be less than errors on the intensity of the supersaturation pointwise. In our study, we are taking care of the latter and, therefore, we are only addressing part of the problem. As a consequence, we weaken our statement and modify the line 3 as follows

“However, accurate prediction of the Relative Humidity with respect to Ice, RH<sub>i</sub>, distribution within ISSRs at cruising altitude remains difficult.”

*- L253: What are the requirements to support effective contrail avoidance strategies?*

Effective contrail avoidance strategies will require efficient, quick, and accurate meteorological predictions of ISSRs. However, the specific performance requirements for such predictions remain unclear. Several strategies have been proposed (avoiding ISSRs entirely or avoiding both ISSRs and their surrounding regions) each of which may impose different accuracy requirements on the forecasts. We qualify our statement and modify the line 253 as follows:

“Although ISSR detection has improved, reanalysis products still exhibit limitations to support potential contrail avoidance strategies, especially in the absence of clearly established performance requirements.”

*- L51: Wang et al 2025 published a ANN humidity correction methodology. This publication adds an XGBoost regression for RH<sub>i</sub> < 85%, and a different training/validation data split. Given the*

*similarities, this line deserves a whole paragraph describing the differences with Wang 2025, and how this methodology aims to improve on the previous work.*

Line 51: This study is building on the methodology by Wang et al. (2025). One of the key differences is the rigorous sampling strategy to strictly prevent temporal leakage between the training and testing/validation sets. This new splitting strategy revealed much weaker performance from the standalone ANN, especially in terms of MAE on dry values. To address this, we developed the hybrid architecture presented in this study. Thus, the main differences from the original work are the strict data splitting strategy, the introduction of new engineered features (detailed in Section 2.4), and the integration of an XGBoost model. Additionally, our study focuses on the North Atlantic region as it is a hot spot for air traffic and a potential region for contrail avoidance trials. Furthermore, it offers a better coverage by the IAGOS dataset.

We think that an additional paragraph highlighting the main differences with Wang et al. (2025) would be redundant. We note that additional point of comparison have been added:

- line 73: “Unlike Wang et al. (2025), we focus on the North Atlantic region, spanning latitudes from 40°N to 70°N and longitudes from 65°W to 5°E (Fig. 1).”
- line 75: “For these reasons, it is thought to be a potential region for contrail avoidance trials.”
- line 79: “Our data processing is implemented based on methodology from Wang et al., 2025, with some modifications to ensure proper separation of the training and test datasets to avoid temporal leakage.”
- line 126: “This differs from Wang et al. (2025) in two ways: i) we use five days for training (four days in Wang et al. (2025)), ii) an additional day gap is used after the validation or testing day to prevent correlations with the following training period.”

*- L74: What kind of biases in the weather might this domain selection introduce? Have you tested how well your models apply outside this domain?*

Our study focuses on the North Atlantic, adapting the model to other regions of interest would require tuning both the model and the engineered features to the specific thermodynamic properties of the new region, as these differences can impact performance. However, this expansion presents a major challenge because we rely on the IAGOS dataset as ground truth, and measurements are sparse or non-existent in many other regions. This lack of data complicates the training process and severely limits our ability to robustly assess the actual performance of the model globally.

*L83: Did you consider model levels? It may be worth exploring if the higher vertical resolution would improve your results.*

We acknowledge that using ERA5 model levels could potentially benefit the results due to their higher vertical resolution. However, this approach was not implemented in the final analysis due to the large size of the dataset and the significant computational resources required for processing. We consider the current approach to offer good tradeoff between precision and computational performance, without compromising the main conclusions.

*- L117-121: How did you interpolate the values for T and q? Linear interpolation in q introduces bias when working with coarse pressure levels.*

We performed a linear interpolation in the vertical dimension for T and q to map the ERA5 pressure levels to the aircraft altitude. While we acknowledge that linear interpolation of humidity can introduce biases over large vertical distances, we found this method to be sufficient given the vertical resolution

of ERA5 in the UTLS region and the scope of our study. We experimented with polynomial interpolation but the resulting datapoints did not match the IAGOS observations as closely as the linear approach. Leveraging the higher vertical resolution from ERA5 model levels in future work could mitigate the bias introduced by interpolation.

- *Table 1: Teoh et al 2024 introduced a latitude correction for the humidity correction. Should latitude be a feature?*

During our data analysis, we explored a wide range of potential features, including latitude, to evaluate their impact on the model. However, our feature selection process revealed no significant statistical relationship between latitude and RHi in our dataset, and its inclusion did not lead to a meaningful improvement in predictive performance. Hence, latitude was filtered out at the feature selection process.

## *Minor Comments*

- *L31: "are spending" -> "spend"*

We have corrected the typo.

- *L33: Suggest using stats from more recent Teoh, R. et al. (2024) "Global aviation contrail climate effects from 2019 to 2021," Atmospheric Chemistry and Physics, 24(10), pp. 6071–6093. Available at: <https://doi.org/10.5194/acp-24-6071-2024>.*

We have changed to a more recent publication and adapted the text accordingly:  
"For instance, it has been estimated that only 2.7% of total flights are responsible for 80% of the total contrail RF in 2019 (Teoh et al., 2024)."

- *L37: Its worth motivating why we need to detect ISSRs. Its presumed that the reader knows "to meteorologically forecast ISSRs with enough accuracy" we need ISSR detections. May want to add context e.g. "Global ISSR forecasts are generally derived for numerical weather forecasting systems, or nowcast from in situ measurements or inferred from remote sensing. Both approaches rely on accurate detections of ISSRs, in the first case to validate models, or in the second through measurements"*

We have added context to motivate the ISSR detection by including your suggested sentence.

- *L44: Not just ERA5 - any numerical weather prediction system. I'd flip this around - numerical weather prediction models provide a comprehensive prediction across the global atmosphere. ERA5 is a highly trusted source of numerical weather prediction.*

We have modified the statement as follows:

"Hence we must rely on numerical weather prediction models to obtain RHi values across the global atmosphere. We use the ERA5 reanalysis data (Hersbach et al., 2020) as it is highly trusted source of numerical weather prediction."

In addition, we modify line 5 as follows:

"On the contrary, reanalyses offer a global estimate of RHi, but it is known to exhibit a dry bias near the tropopause where ISSRs are located as well as significant random errors."

- *L45: Define what a dry-bias means*

We have defined the dry bias as follows:

“However, it is known to suffer from a dry bias in the Upper Troposphere (UT) and Lower Stratosphere (LS) (Rolf et al., 2023), which means that the model often predict lower relative humidity values compared to the observed values.”

- L48: *Other publications with humidity correction: (constant) Schumann, 2012; Schumann et al., 2015; Teoh et al., 2020; Schumann et al., 2021; (piecewise function) Teoh et al 2022; Teoh et al 2024; (quantile mapping) Platt et al 2024*

We have included additional references for humidity corrections (Teoh et al., 2022; Platt et al., 2024).

- L55: *This sentence sounds like an LLM. I'd move L59-L61 up front, remove this sentence, and then have L57-58. Can you be more specific as to why you chose the hybrid model? From this description it sounds like you used XGBoost for compute performance reasons rather than accuracy.*

We think that the sentence at line 55 is important as it describes the ANN and three-based models in a general manner, before focusing on the case of ISSRs.

Concerning the decision to use XGBoost for the hybrid model, it was driven by accuracy rather than computational performance. As discussed in section 4 and illustrated in Figure 4, XGBoost significantly outperforms the ANN in reducing the MAE in drier conditions. However, the ANN excels at capturing the non-linear dynamics associated with high humidity and achieves a much higher ETS (resulting in a better classification of ISSRs). Therefore, the hybrid approach was selected to leverage the complementary strengths of the models to achieve a better classification performance while maintaining a strong overall accuracy.

- L94: *How long is the "longer period"?*

We have specified “a longer period of few minutes”

- L104: *Just confirming that IAGOS accuracy is a function of RH<sub>i</sub> or of absolute humidity. I had remembered that humidity sensor accuracy was a function of absolute humidity.*

The IAGOS accuracy is indeed a function of the relative humidity (<https://www.iagos.org/iagos-core-instruments/h2o/>)

- L126: *How does this compare to Wang 2025?*

To further compare with Wang et al. 2025, we have included the following sentence:

“This differs from Wang et al. (2025) in two ways: i) we use five days for training (four days in Wang et al. (2025)), ii) an additional day gap is used after the validation or testing day to prevent correlations with the following training period.”

- L156: *Is it possible the ANN is overfitting these engineered features? You acknowledge the proper data split, but could you use additional data outside the domain to gain confidence?*

We are confident that our ‘by day’ data splitting strategy avoids overfitting by preventing temporal leakage between the training and test sets. Our objective is to optimize predictions within the North Atlantic region. Applying the model to other domains requires tuning the ANN to fit the specific dynamics of the given region.

- L160: This criteria sounds more like "No existing cirrus" rather than "clear sky." Could also look at the IAGOS ice crystal measurements to judge pre-existing cirrus (Petzoldt 2025)

We agree that "No existing cirrus" is a physically accurate description of the condition but we have chosen to use "clear-sky" to stay consistent with previous work in the field (e.g. Wang et al 2025).

- L170: (Re)Introduce acronym MAE

We have reintroduced the acronym MAE.

- L182: Add citation? Where does this baseline come from?

Table 3 is now referred at the end of the sentence.

- L186-188: Its not clear to me why "structured input data" ~ drier regimes. Its more clear to me that "high humidity conditions" ~ complex non-linear dependencies.

We have restructured the paragraph as suggested:

" The ANN is well-suited for capturing complex, non-linear dependencies, particularly under high humidity conditions, while XGBoost performs robustly in drier regimes."

- L221-222: This is first clear explanation of why XGBoost is preferable to ANN for the drier regimes. L230 - L233 is also great. Bring this language up front!

- Table 3 is super helpful - It would be helpful to use this language up front when describing the benefits of the hybrid architecture.

We are grateful to Reviewer #1 for these comments. To address them, we modify the abstract to introduce the following sentence at line 11:

"This hybrid approach significantly outperforms raw ERA5 data, leveraging the ANN's ability to capture non-linear relationships and the XGBoost's robustness in handling drier conditions."

This additional sentence in the abstract necessarily comes with other minor changes to respect the word number limitation, but these do not change the overall content.

- L223: Repeats the previous line

- Line 223: We have deleted the previous line (222): "It effectively corrects low RH<sub>i</sub> values while maintaining a strong improvement in the ETS".

- Table 3, Table 4: How do these results compare with Wolf et al 2025 or Platt et al 2024 (quantile mapping)

We have added a sentence at line 236 which compare our results with the ones using quantile mapping by Wolf et al (2025):

" In comparison, while QM successfully removes bias from the distribution, it does not improve point-wise predictive accuracy. Wolf et al (2025) explicitly report that after applying QM, the RMSE, MAE, and MSE increase unnoticeably". In contrast, our model significantly reduces the MAE from \$13.70\%\$ to \$11.37\%\$."

Sincerely,

J r mie Juvin-Quarroz, on the behalf of the authors

## Response to Reviewer #2

We thank Reviewer #2 for the positive assessment of our manuscript. The comments provided valuable clarifications and have helped improve both the content and presentation of the article. All comments have been carefully addressed. Below, we reproduce the comments from Reviewer 2 in *italics*.

*In this study, the authors build up a hybrid machine learning model to correct bias of relative humidity over ice in ERA5 in the UTLs over North Atlantic. The model consists of a XGBoost regressor in dry conditions and an ANN in more humid regions, with the threshold of 85% RHi. In the test data, the model can reduce the dry bias of ERA5, and increase the number of correctly predicted ISSRs.*

*The paper is well written, with a clear presentation of technical details. The figures and tables effectively illustrate the comparisons.*

*I have a few additional comments that should be clarified as part of a minor revision before the study is published.*

We are grateful to Reviewer #2 for the careful reading and comments. We address the additional comments below.

*In Sect 2.3 or Fig 2, the ERA5 variables are linearly interpolated to match the mean altitude of the IAGOS data points. This is a reasonable approach to reduce discrepancies between model level and pressure level input data. However, I am thinking if log-pressure linear interpolation—as used in tools like pycontrails—might be more appropriate, since it better represents the vertical structure of the atmosphere.*

Reviewer #2 is raising a good point about the linear interpolation of ERA5 variables to match the altitude of the IAGOS data points, as did Reviewer #1. We indeed performed a linear interpolation using pressure as a coordinate for T and q to map the ERA5 pressure levels to the aircraft altitude. We experimented with polynomial interpolation but the resulting datapoints did not match the IAGOS observations as closely as the linear approach. We did not try a log-pressure linear interpolation as done in pycontrail python module. This may indeed be a better vertical interpolation. We add the following sentence at line 116:

“Future work should test log-pressure linear interpolation to reflect the atmosphere’s vertical structure.”

*In Table A1, in addition to XGBoost, LightGBM also performs well in predicting RHi, particularly in terms of MAE and ETS. It would be helpful to include a sentence or two explaining the choice of XGBoost over LightGBM, or whether the results could be generalized to gradient boosting decision trees as a class of models.*

The performance of several machine-learning models is reported in Table A1. XGBoost and LightGBM indeed achieve a similar MAE. However, we selected XGBoost as it offered better classification performance. While LightGBM achieved a slightly lower Test MAE (11.55% vs. 11.56%), XGBoost outperformed LightGBM in the ETS (0.35 vs. 0.32). The two models use Gradient Boosting Decision

Trees, explaining the very similar results. Since our dataset size is limited, the computational speed offered by LightGBM was not a limiting factor, allowing us to prioritize the model giving the best balance over all evaluation metrics (MAE,  $R^2$  and ETS). We acknowledge that LightGBM can be considered in the Hybrid Architecture especially when using a significantly larger dataset (millions of samples), or when dealing with very limited computational resources.

*Other comments:*

*The definition of contrail formation and persistence could be clarified. Contrails form when hot and humid jet engine exhaust mixes with the cold ambient atmosphere, resulting in local liquid saturation within the plume. Contrails persist when the surrounding air is ISSRs, allowing the ice crystals to grow and spread into contrail-induced cirrus. This clarification relates to the first sentence of the abstract and Lines 35–36.*

We have modified the sentence at line 35-36 to address the Reviewer's remark on contrail formation and persistence as follows:

“The largest contributor to non-CO<sub>2</sub> effective radiative forcing (ERF) is the radiative forcing from condensation trails---contrails---which are linear ice clouds formed behind aircrafts as hot exhaust gases mix with cold ambient air. The physical processes governing contrail formation are well understood (Schmidt, 1941; Appleman, 1953; Schumann, 1996).”

We think that the first sentence of the abstract is not misleading as we are speaking about persistent contrail cirrus, and not just contrails.

*The description of Fig 4a as the baseline ERA5 reanalysis could be introduced earlier*

We add the following sentence at line 120:

“A comparison between observed RHi from IAGOS and RHi predictions from ERA5 reanalysis is shown in Fig. 4a.”

*Lines 26: full name of MOZAIC*

We now give the full name of the acronym MOZAIC.

*Line 70: the measurements of RHi are not direct; they are calculated from temperature and water vapor measurements*

Indeed, RHi is estimated from the temperature and the relative humidity by IAGOS. We have modified the manuscript as follows:

“Combining these datasets enables a supervised learning approach where the ERA5 features serve as model inputs, while the estimation of RHi from IAGOS from temperature and relative humidity measurements serve as the target variable.”

Sincerely,

J r mie Juv n-Quarroz, on the behalf of the authors