Review of „Variability of ice supersaturated regions at flight altitudes: evaluation of ERA5 reanalysis using IAGOS in situ measurements by Hildebrandt et al., 2025

The manuscript tackles an important topic, namely the representation of ice supersaturated regions (ISSRs) in ERA5 and their comparison with in situ observations from IAGOS. The study is not restricted to the North Atlantic but also includes several other regions, including the tropics. The paper is generally well written and technically sound.
That said, the degree of novelty is limited. Key results such as the ERA5 dry bias, the underestimation of ISSR occurrence, and the generally low ETS have already been documented in recent studies. While the authors expand the analysis in both spatial and temporal scope—which is certainly valuable—the added contribution beyond confirming previous findings is not fully evident.

Therefore, I recommend **major revisions** before the manuscript can be accepted.

Major Comments

This paper presents in detail the results of the comparison between ERA5 and IAGOS with regard to relative humidity over ice. Most of the results confirm earlier studies and do not provide any truly new insights. Additionally, the paper is very long.

It should be noted, however, that the tropics are also coming more into focus here. Perhaps this would be a way to make the manuscript more exclusive. I think there are many interesting findings in it, but more attention should be paid to the influence of aircraft routes on the results.
For example, while in the Southern Trans-Pacific region most of the measurement points are north of the equator and only at the edges of the area, the South Atlantic and Africa cover a very good area in a north-south direction up to 30°S. How does this influence the results?
This also raises the question of whether the division of the seasons into DJF, MAM, JJA and SON makes sense here. Perhaps the tropics should be evaluated using different methods than the extra tropics?

Many methods are taken from previous work, such as the distinction between cloud free and cloudy conditions. It would be interesting to see how the results change when these definitions change. Or are there perhaps even better methods for doing this?

Another example that comes to mind is the use of ETS. Although ETS is a frequently used measure in meteorology and is also used in the studies cited, I wonder how meaningful the results are or what added value they provide when we already know that ERA5 has problems with ice supersaturation.
How do the results change if ISSRs in ERA are defined as, for example, RHi=90%? Similarly, in the later analysis in Sec 3.5: how sensitive are the results to the thresholds used to distinguish between cloudy and clear skies?
Could there be another measure that does not penalize spatial or temporal shifts as heavily as ETS?

The consideration of weather patterns over the North Atlantic does not fit well into the manuscript. Why are other patterns over different regions not also taken into account? First, it is pointed out that many other regions are also being studied, only to then undertake a very specific study of the North Atlantic. This part is also rather shallow.

Some of the figures are difficult to read, and fine details are sometimes hardly visible. The axis labels are also incorrect in some cases (saturation ration or RHi?). More emphasis should be placed on consistency (always the same order in the legend, each subplot with vertical axis labels or only the first column?).

Specific comments

Line 5: „data set" missing after *reanalysis*?

Line 16: Use another word than *important*. Maybe „significant"?

Line 37: I think ERA-Interim was compared to IAGOS

Line 100 & 108: Formatting of reference is broken

Line 145: how were RHi and RHI calculated in detail? This is an important information

Line 150f: Why is the stronger gradient an argument for the use of the thermal tropopause? This is not clear to me. Did you consider the cold point tropopause (CPT) for the tropics?

Figure 2: X-axis bottom row: What kind of frequency? Relative to what?

Line 168 & 180: What does „no measurement" mean? Data point of a flight without BCP? Please clarify.

Figure 3: See Fig. 2: What kind of frequency relative to what?

Line 175: Why was the approach chosen from Wolf et al. (2025) using the cloud cover instead of the IWC used in Wang et al. (2025)? This definition is crucial to the results later on. How sensitive are the results to these values?

Sec. 2.5: What was the influence of the COVID-19 pandemic on the statistics?
During the pandemic, the number of commercial flights decreased substantially. Did this reduction in flight activity affected the amount of IAGOS data collected during that period. How does this reduced sampling influence the representativeness of the results, in particular for the distribution across different weather patterns? So e.g. Winter patterns 1 and 4 show high values in Winter 2020/2021.
How many measurements fall into each weather pattern?

Line 207: „*maybe due to sensor error*": be more specific.

Line 219: what is meant by „*better approximation of temperature*"?

Figure 7: shading of 95% confidence interval not visible. Also, the order of the seasons is different in the legend from figure to figure (JJA, SON, DJF, MAM, in Fig 7 but DJF, MAM… in Fig. 12)

Line 225f and Figure. 7: Why are the minima of temperatures for the tropic cases below the thermal tropopause and not as distinct as in the extra tropics? Also not only JJA in South Asia is below the thermal tropopause. The minimum of DJF is even lower (approximately 900 hPa, Fig. 7e). You only discuss it for JJA in South Asia.

Figure 8: X-axis. Either Saturation ratio without units or RHi with [%]. You show values for saturation ratio not RHi.

Line 245: „distribution" missing after *vertical*?

Line 257: I don't understand the sentence starting with „The lower values of RHi…."?

Figure 9: X-axis: Same as Fig. 8. You show values for saturation ratio, not RHi.

Line 261. Be more precise. What do you mean with the *highest RHi*? Peak value? Or throughout the profile?

Line 272: I do not agree. I think a seasonal cycle is also clearly visible for Fig. 9 e)

Line 273: Is the argument for the high RHi values in South Asia in JJA due to deep convection not also valid for other tropical regions? Why is it so different for South Asia? Is it an observation of the meteorological characteristics of the region or rather due to the course of the flight routes in this area?

Line 275f: where can I see this?

Line 302. The beginning of this sentence sounds odd. Why should it be dry in the first place?

FIgure 10: saturation ratio or RHi

Figure 11: Subfigure e): Are there no cloudy or indeterminate measurements from IAGOS? I can't see solid lines for this cases.

Line 340/341: *ERA5 may predict less … than ERA5…*

Figure 12: subplot f): There are no labels for height levels above the tropopause. In g) there is only „-50".

Line 386: „*outliers*": What might be the reason for these outliers? Physical or sampling? I think this should be more specific.

General: The figures are sometimes hard to read with a lot of different lines. Also please check if your are plotting saturation ratio or RHi!

Line448/449: The reader is left alone with the question raised. What would the authors' suggestions be?

Line 457: I'm not sure I can agree with the authors' conclusion here. In some cases, the weather patterns cover different altitudes but are still all within a comparable range.

Line 476f: It should be clarified whether the lower performance of ERA-Interim implies that the seasonality found by Irvine et al. (2012) is no longer valid, or whether the differences mainly reflect dataset characteristics and methodological choices. The current phrasing ("but") may unintentionally suggest that the earlier results are invalid.

Line 550: You recommend a standard method but do not specify which one.

Code availability: I strongly recommend making the code publicly available via platforms such as zenodo.

Line 588: Which time frame?

Line 605: Why is a sharper gradient the basis for selection?

Figure A1: Saturation ratio or RHi?

Figure S1: The important Information is very small in this figure, grey shading is not visible.

Figure S6: What is the x-Axis? Is ΔRHi in units of RHi? Or is it a relative difference between RHi_iagos and RHi_era?