S1 Effect of sample size on results

In the evaluation of ERA5 using IAGOS, we are highly dependent on the number of available measurements from IAGOS. If we do not have enough measurements, our conclusions may not be representative. Hence, in this appendix, we explore the effect of the sample size on temperature, relative humidity over ice (RHi), the ice supersaturated region (ISSR) fraction, and the equitable threat score (ETS).

The effect of the sample size on temperature and RHi is shown in Fig. S1 and S2, respectively. A maximum sample size of 2000 was chosen due to stabilisation of mean and standard deviation and 1000 tests were drawn in total for all sample sizes based on literature (Johnson, 2001). We see that as the number of sample size increases, the standard deviation decreases. At a sample size of 100, we find that the mean plus or minus one standard deviation is within the accuracy range of the IAGOS ICH sensor. The mean appears to stabilise at around 250 to 500 samples. However, from 500 samples on, the standard deviation range also seems to stabilise, but this is already within the accuracy range. Hence, for the evaluation of mean temperature and RHi, it is recommended to use at least 250 to 500 samples for a given region and vertical level.

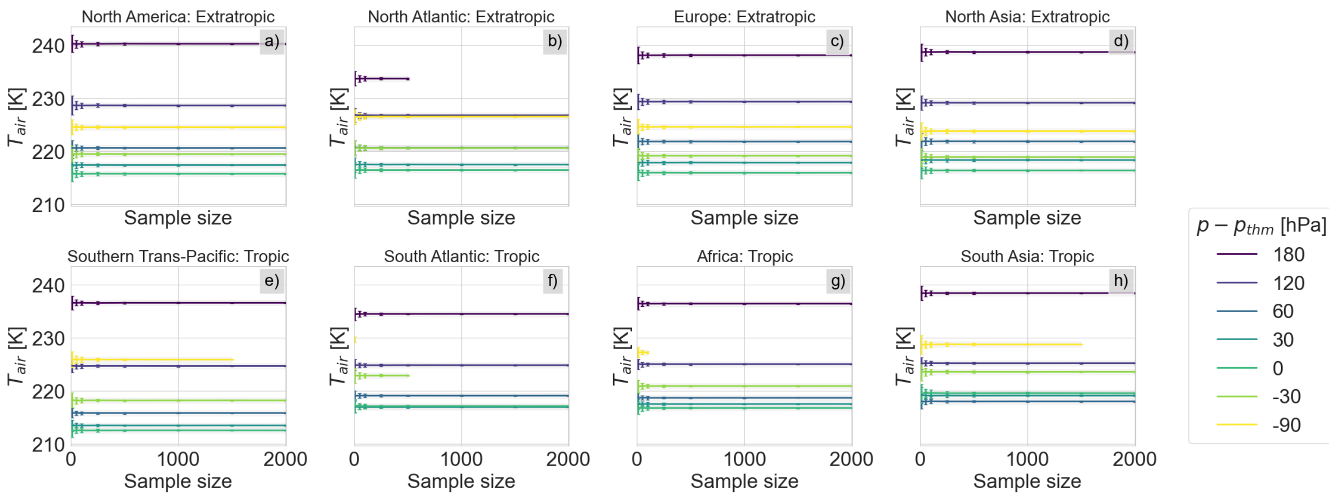

Figure S1. (a-h) Effect of sample size on IAGOS mean temperature for the different geographic regions at different vertical levels from the tropopause. Vertical lines indicate the standard deviation from the mean and grey shading shows the accuracy of the IAGOS ICH sensor for temperature. For each sample size, 1000 tests were drawn.

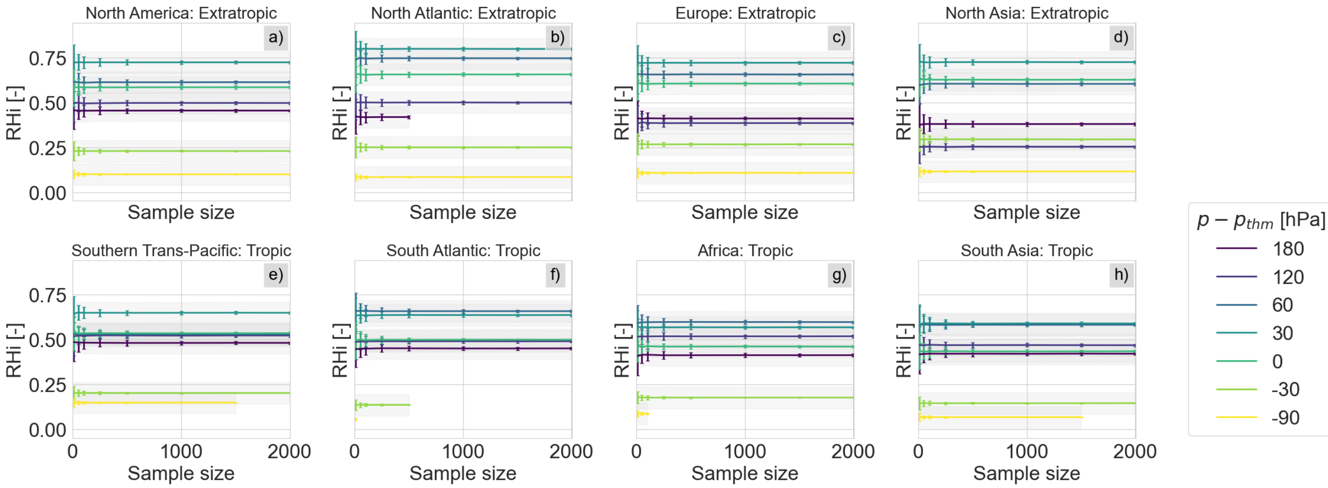

Figure S2. (a-h) Effect of sample size on IAGOS mean relative humidity over ice for the different geographic regions at different vertical levels from the tropopause. Vertical lines indicate the standard deviation from the mean and grey shading shows the accuracy of the IAGOS ICH sensor for relative humidity. For each sample size, 1000 tests were drawn.

The effect of the sample size on the ISSR fraction is shown in Fig. S3. It is important to investigate this as the ISSR fraction is a fraction of the number of points that show ISS to the total number of points available for the given conditions. We also see this sensitivity in Fig. S3. The mean ISSR fraction appears to stabilise already at around 250 samples, in which the standard deviation is between 2% and 2.5%, approximately. Given the natural variability of ISSRs, this is not statistically significant. Furthermore, as the number of samples increases, we do not see a large impact on the reduction of the standard deviation; generally, the change in the standard deviation is less than 1% going from 250 to 500 samples. Therefore, 250 to 500 samples is considered sufficient.

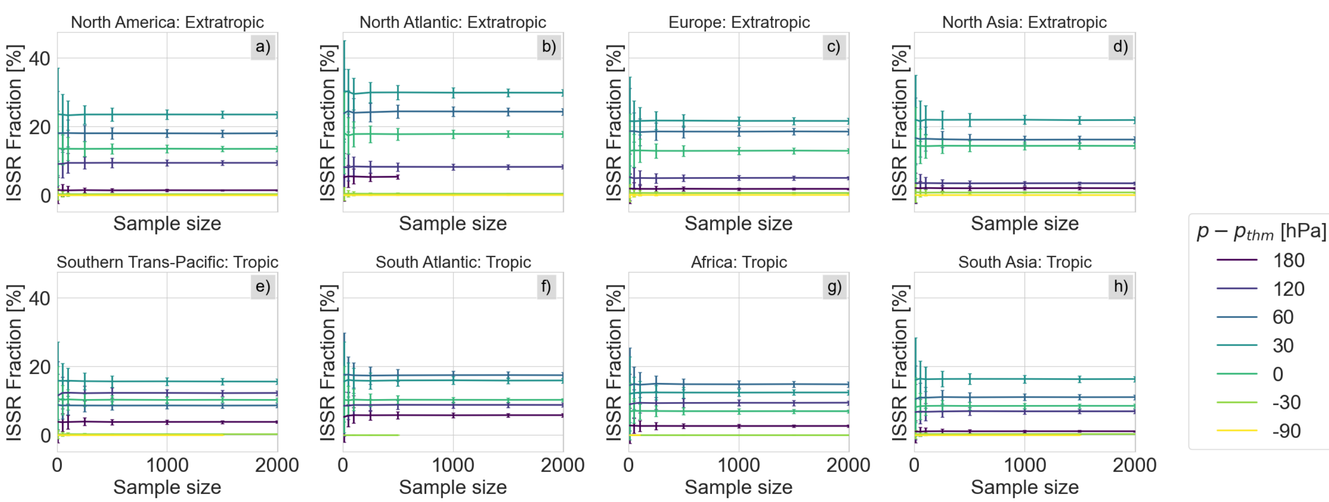

Figure S3. (a-h) Effect of sample size on IAGOS ice supersaturated region fraction for the different geographic regions at different vertical levels from the tropopause. Vertical lines indicate the standard deviation from the mean. For each sample size, 1000 tests were drawn.

We also analyse the effect of sample size on the ETS, shown in Fig. S4. However, in this case, we only take 100 tests per sample size for each region and vertical level combination due to the time it takes to calculate the ETS. The mean ETS seems to stabilise around 250 to 500, meanwhile the standard deviation first appears to stabilise between 500 and 1000 samples. However, in the lower stratosphere, the stabilisation occurs close to 1000 samples. At 250 samples, the 95% confidence interval has already reduced significantly for all vertical levels except for in the lower stratosphere. Here, the 95% confidence interval with 250 samples is approximately 0.015, which is still small, but is also dependent on the region. Hence, for the analysis of the ETS, we also recommended using at least 250 to 500 samples, but a larger sample size would result in less variations of the ETS.

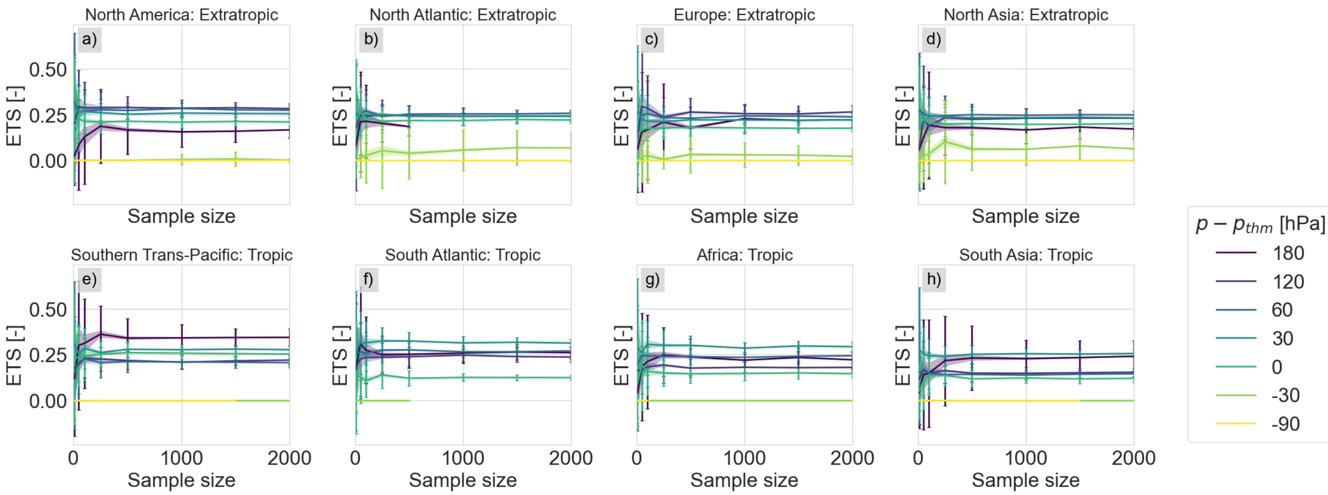

Figure S4. (a-h) Effect of sample size on ice supersaturated region fraction hit rate for the different geographic regions at different vertical levels from the tropopause. Vertical lines indicate the standard deviation from the mean. Shading shows the 95% confidence interval. For each sample size, 100 tests were drawn.

S2 Distribution of differences in temperature and relative humidity over ice between IAGOS and ERA5

This appendix presents the distribution of the differences in temperature and RHi between IAGOS and ERA5. The differences in temperature are displayed in Fig. S5. The differences in RHi are displayed in Fig. S6.

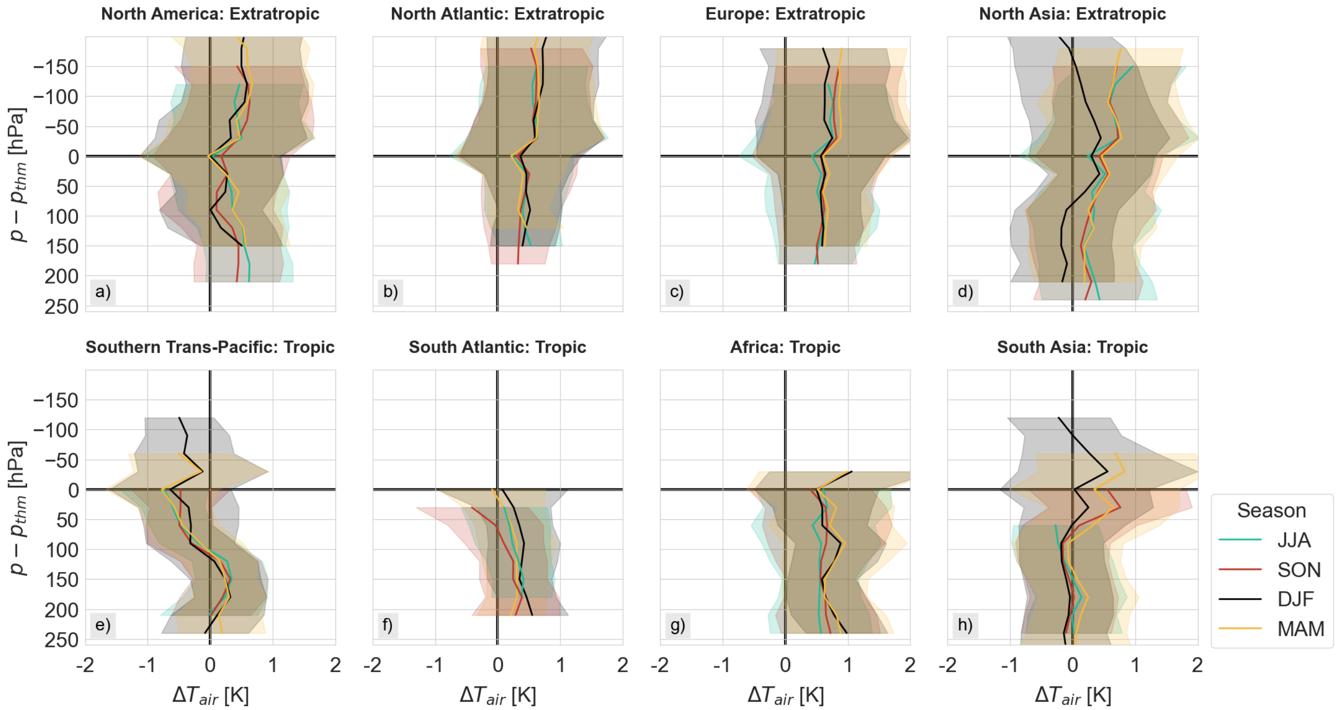

Figure S5. (a-h) Vertical distribution of temperature differences between IAGOS and ERA5 ($\Delta T = T_{\text{IAGOS}} - T_{\text{ERA5}}$) per season and per region, using levels based on distance to thermal tropopause, only considering levels with 500+ samples. Shading shows ± 1 standard deviation around the mean difference.

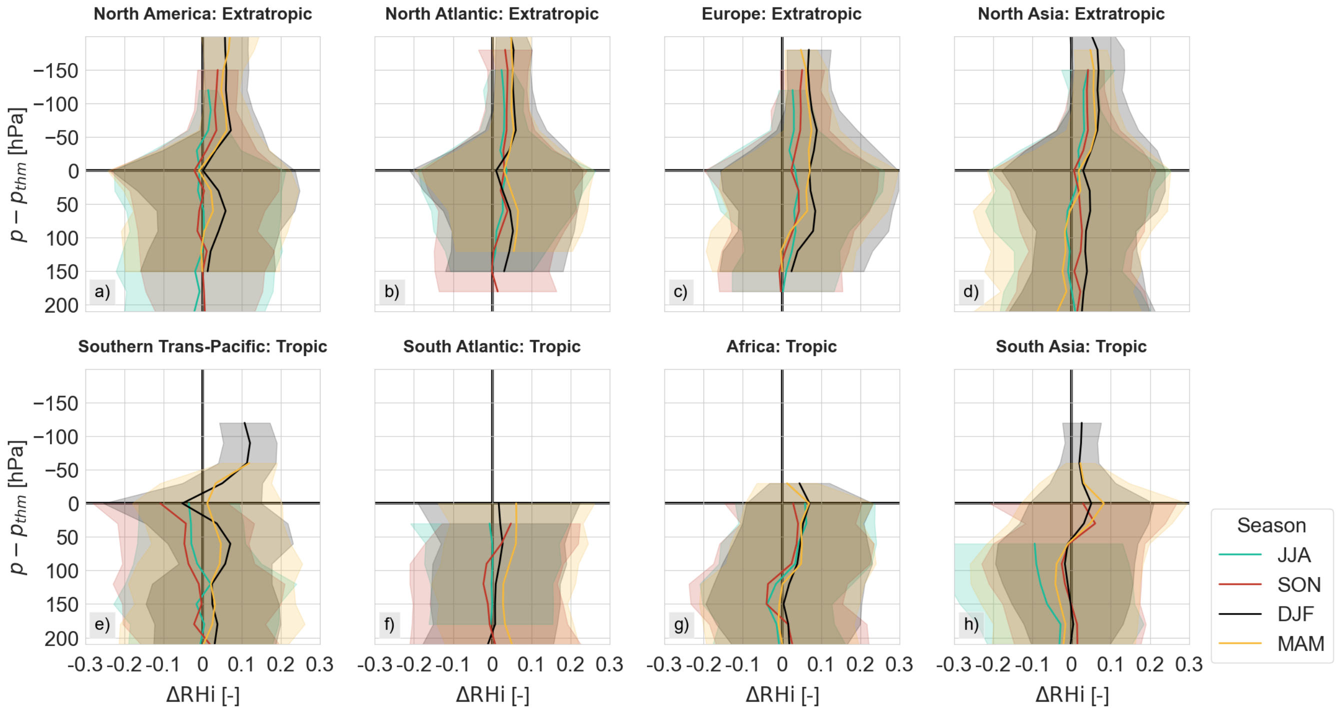

Figure S6. **(a-h)**Vertical distribution of relative humidity over ice differences between IAGOS and ERA5 ($\Delta RHi = RHi_{IAGOS} - RHi_{ERA5}$) per season and per region, using levels based on distance to thermal tropopause, only considering levels with 500+ samples. Shading shows $\pm$ 1 standard deviation around the mean difference.

References

Johnson, R. W.: An Introduction to the Bootstrap, Teaching Statistics, 23, 49–54, https://doi.org/10.1111/1467-9639.00050, 2001.