In this paper, the authors describe the creation of a global temperature dataset derived from IASI satellite observations. A new retrieval method based on using an Artficial Neutral Network (ANN) to retrieve vertical profiles of atmospheric temperature is applied globally for 13 years of IASI data from 2006-2018 at 11 pressure levels throughout the troposphere and stratosphere. They then verify their derived temperatures against ERA5 reanalyses globally and ARSA radiosonde observations in selected locations. They find that their new retrieval typically deviates less than 0.5 - 1K from these reference data. Finally they fit linear trends to their retrieved temperatures to show global warming/cooling rates in the troposphere and stratosphere, providing further important evidence and quantification of our changing climate.

Overall, the paper and its dataset appear to be of very high quality and are potentially very useful to the scientific community. The manuscript is well written and in general there is sufficient detail to describe the method. Figures are generally clear and the analysis is sufficient to support the conclusions. The temperature dataset itself is very well presented and archived on a freely accessible website and easily downloaded from an FTP server.

I have some minor comments below that the authors may like to consider in a revised manuscript.

Below these, I have also included some typographic and stylistic suggestions that they may also like to consider. I do not require a point-by-point response to these typographic/style suggestions.

We thank the reviewer for their positive review and useful comments. We answer each of the questions raised in blue hereafter.

Minor Comments

The manuscript could benefit from a short "Data" section where the IASI instruments and radiances, the ERA5 reanalyses and the ARSA radiosonde data are briefly described. This would greatly improve flow in the results sections where the ATPs are compared to ERA5 and IASI, without needing to introduce and describe them in the text there.

This is indeed a good suggestion. A data section presenting IASI radiances, ERA5 reanalysis and ARSA measurements was added as Section 2, and the paragraphs describing the data later in the article were removed:

"2 Data

2.1 IASI radiances

Each of the three IASI instruments are mounted on-board the Metop platform flying on a polar orbit at an altitude of 817 km. The IASI swath contains 30 fields of view with 4 pixels in each field of view. This observation mode allows each IASI instrument to observe the entire Earth twice a day, between 9:15 and 9:45 AM and PM local time. IASI measures the radiation of the Earth-atmosphere system in the thermal infrared in 8461 channels between 645 and 2760 cm-1 (resolution of 0.25 cm-1, 0.5 cm-1 apodized; Clerbaux et al., 2009).

2.2 ERA5 reanalysis

The European Center for Medium Weather Forecast (ECMWF) reanalysis, ERA5 (Hersbach et al. 2018, Copernicus Climate Change Services) is a 4D-Var data assimilation product. It is part of the Integrated Forecast System (IFS) that provides variables relevant to the atmosphere, land and ocean. ERA5 atmospheric temperature product used in this work is hourly, and is available on 37 pressure levels (from the surface up to 0.01 hPa). ERA5 actually assimilates IASI radiances from Metop-A and Metop-B, as well as high spectral resolution radiances from other instruments such as AIRS on Aqua, and CrIs from S-NPP and NOAA-20. Note that IASI is the largest contributor to error reduction for global numerical weather prediction in the thermal infrared spectral band (Borman et al., 2016).

2.3 Analysed RadioSoundings Archive

The Analysed RadioSoundings Archive (ARSA) is a 41-year (1979-2019) database of radiosonde temperature profiles measurements from different stations around the globe (Scott et al., 2015). ARSA provides 43 pressure-level profiles (from the surface to 0.0026 hPa) of temperature, water vapour and ozone, and surface temperature. The raw radiosonde observations go through severe multistep quality controls, to eliminate gross errors. If the selected radiosonde measurement is unable to provide forward radiative transfer modelers with the required information (above 30 hPa for temperature), ARSA combines existing radiosonde measurements with other reliable data sources to complete the description of the atmospheric state as high 0.0026 hPa. Temperature profiles are thus extrapolated with ERA-Interim (Dee et al., 2011) outputs between 30 hPa and 0.1 hPa for temperature. Above 0.1 hPa, the profiles are extrapolated up to 0.0026 hPa using a climatology of ACE/Scisat Level 2 temperature products. ARSA was validated against IASI observations by simulating spectra from the 4A/OP forward model (Scott and Chédin, 1981) with ARSA profiles as inputs, and comparing them with space-time colocated IASI observations. The pertinence of the requested modifications after this validation has been also assessed against the TIROS Operational Vertical Sounder (TOVS), the Advance TIROS Operational Vertical Sounder (ATOVS, Reale et al., 2008), the Atmospheric InfraRed Sounder (AIRS, Lambrigtsen et al., 2004), the High resolution Infrared Radiation Sounder (HIRS4, EUMETSAT, 2013b), and the Microwave Humidity Sounder (MHS, Hans et al., 2020). Based on these validations, incorrect or unreliable data inherent to the quality of the radiosondes were completed other relevant auxiliary datasets (in particular Level 2 results of ACE-FTS temperature profiles above 10 hPa) measurement data. It is usefull to recall that ARSA is being reprocessed to replace ERA-Interim with ERA5. This will allow, among other things, to extend the period beyond summer 2019, when the production of ERA-Interim stopped.

2.4 EUMETSAT CDR of all-sky temperature profiles

In 2020, EUMETSAT released a climate data record (CDR) of all-sky IASI temperature (doi:10.15770/EUM_SEC_CLM_0027), so the temperature is homogeneous over the whole IASI time series (EUMETSAT, 2020). The reprocessed temperatures were computed with a Piece-Wise Linear Regression Cube (PWLR3) algorithm, using all IASI observations in input (clear and cloudy scenes), and observations from two other microwave instruments flying onboard the Metop-A and -B satellites: the Microwave Humidity Sounding (MHS) and the Advanced Microwave Sounding Unit (AMSU-A).

The basic principle of this algorithm is a linear regression between IASI radiance observations and real atmospheric temperatures. To take into account the non-linearity between the observations and the temperatures, the training dataset is divided into several sub-datasets, resulting from a k-mean clustering. This ensure that, in each sub-dataset, a linear relationship is a good approximation between the observations and the temperature and different linear regression coefficients are computed for each sub-dataset."

Did the authors notice and difference in the accuracy or noise levels in their retrieval during day or night conditions? Possibly, I might expect a local nighttime retrieval to be better constrained than daytime because local thermodynamic equilibrium can be assumed at night, however this might not be a problem for this retrieval because of the machine learning method applied. The authors could simply regenerate Fig. 3 once for ATPs derived during local daytime and again for local nighttime conditions, and see if there are any differences. They can just report this in the manuscript if the is/isn't any difference, no need to include the figures. There is another temperature retrieval for the NASA AIRS satellite developed by Hoffmann and Alexander (2009) who found a large difference between day/night, but theirs was a different method to the one applied here.

The two figures below show the differences between IASI-ANN and ERA5 temperatures in 2016 for day and night observations separately. There are no significant changes in the differences when looking at day or night observations. We add this in Section 4.1 as follows:

"When taking day and night observations separately (not shown), there is no significant change of the differences."

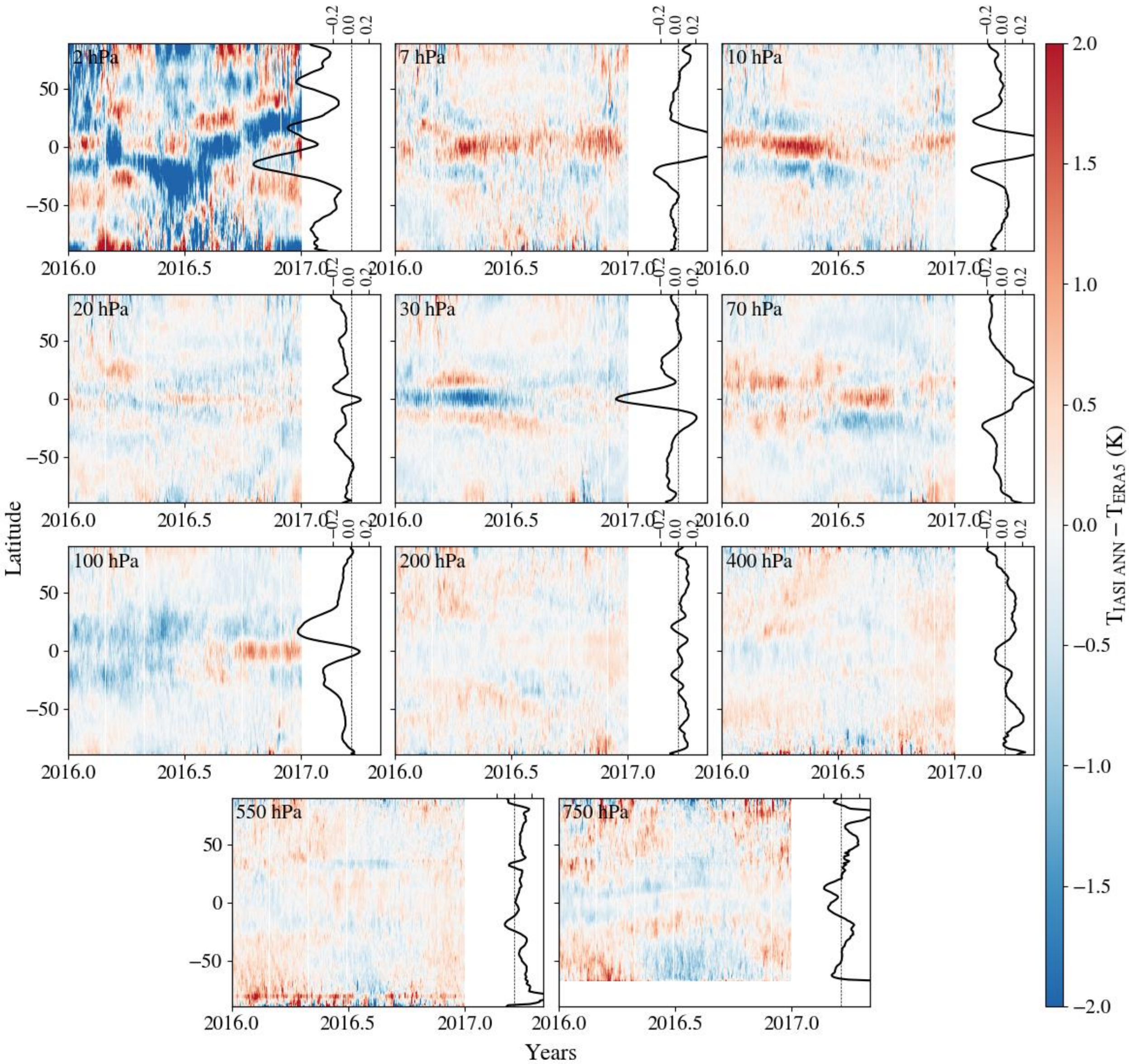


Figure: Daily zonal mean differences between IASI-ANN and ERA5 temperatures for day observations only.

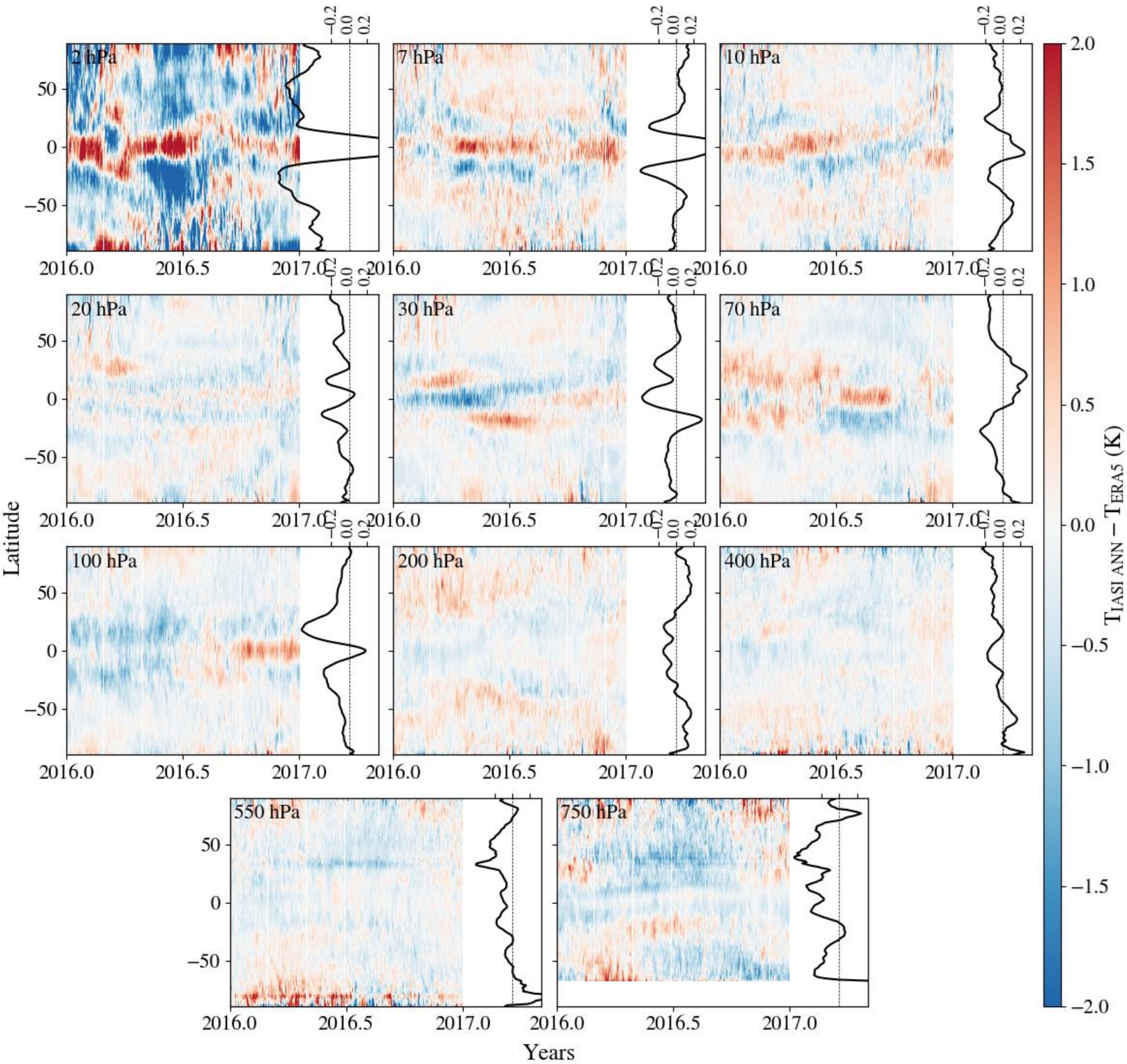


Figure: Daily zonal mean differences between IASI-ANN and ERA5 temperatures for night observations only.

Sect. 2.2, see also l.171-174 - Is it a problem that the training dataset output (ERA5) actually assimilates IASI radiances, so it sounds like the later comparison to ERA5 could be a bit circular? I don't think this presents a problem, but the authors should briefly explain why this is not a problem.

The following sentence was added at the beginning of Section 4.1:

"Althought the neural network was trained with ERA5, it does not reproduce the same temperatures. The output of the retrieval is mainly governed by the variations of observed radiances, and ERA5 can be used for validation."

l.144 - "More information on ERA5 temperatures..." this is where a dedicated Data section would be more helpful.

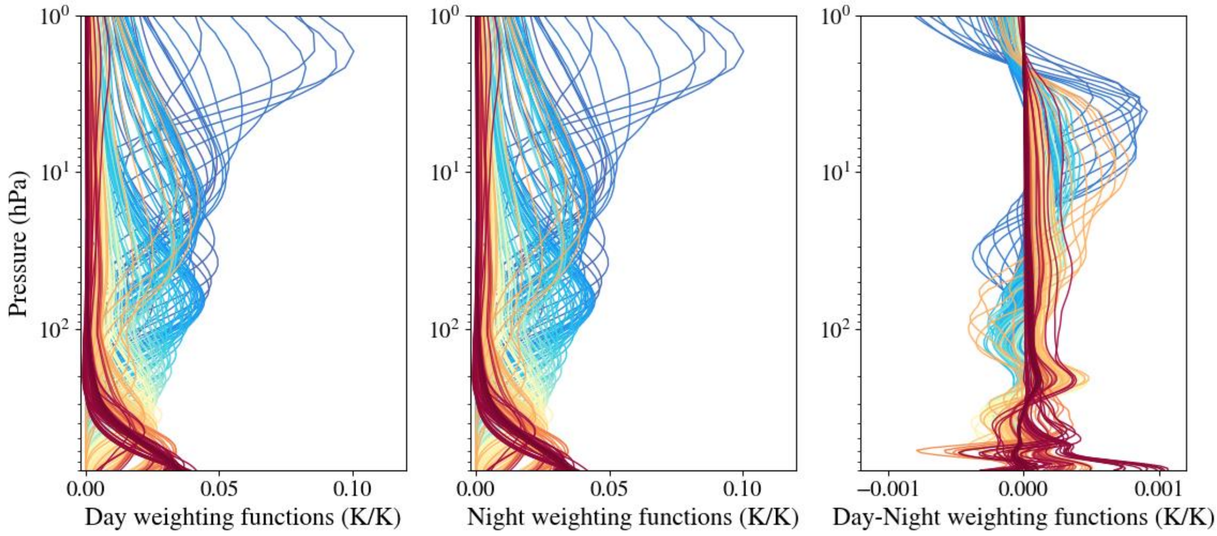
This sentence was removed, as the information about ERA5 was provided before in Section 2.

l.149 - I could be wrong about this, but I think what is plotted in Fig. 2 would be better described as the "weighting functions", "kernel functions" or "sensitivity functions" of the selected IASI channels. These channels, or rather their sensitivity functions with pressure, are arranged into a Jacobian matrix, but the channel sensitivity profiles themselves are not necessarily "Jacobians". Terms like "weighting function" are commonly used to describe these sensitivity functions for hyperspectral imagers, so I might suggest either using this, or clearly explaining why they are being described as Jacobians here. As I said, I may be wrong, but some more explanation around the chosen terminology is required to clear this up.

All the occurrences of the word "jacobian" were replaced by "weighting function".

Fig. 2 - Related to point 2 above, do these sensitivity functions change significantly under day/night conditions? If so, could this affect the true height of the retrieved temperature? I don't expect this effect to be very large, but the authors could comments on this.

The figure below shows the weighting functions averaged for 1000 random day observation (left), for 1000 random night observations (middle) and the differences between the two (right). There are no significant differences between day and night weighting functions. For one given channel, the differences between day and night range from -0.001 to 0.001 K/K while the maximum value of the weighting functions go from 0.04 to 0.11 K/K.



The following sentence was added:

"The weighting functions do not change significantly under day or night conditions and this does not have an impact on the retrieval."

Sect. 3.1, l.168-171 - This information for example would be better in a new "Data" section.

This paragraph was moved to Section 2.

l.178 and elsewhere - [Important] I cannot find anywhere in the manuscript where it is clearly stated whether these differences are "IASI minus ERA5" or "ERA5 minus IASI". As a result, it's not clear for sure which dataset has a warm/cold bias with respect to the other. This is the same for Figs. 3, 6, 7 and S1. I would suggest simply writing (IASI minus ERA5) or similar in the figure caption, that would be enough to clarify.

The figures showing the differences were modified so that the label of the colorbar now indicates "$T_{IASI\ ANN} - T_{ERA5}$ (K)" instead of "ΔT (K)". (see Figure 3 below for example)

l.191 - This first sentence is not clear, please rephrase. How about "At 2hPa, differences range from -2K to 2K globally", or similar?

The sentence was modified as suggested.

l.196-199 and Fig. 3 - The information in these lines and Figure 3 could be easily summarised in a line plot that could be included neatly into the bottom right hand corner of Fig. 3. The additional panel could show the time-averaged difference (x-axis) against latitude (y-axis) for each of the 11 pressure levels considered, which could be colour coded. This would be a very useful summary of the information contained in Fig. 3.

Showing the time averaged differences of the 11 pressure levels on a single plot was making the figure difficult to read and interpret. For more clarity, the time averaged differences were added on the right of each subplot.

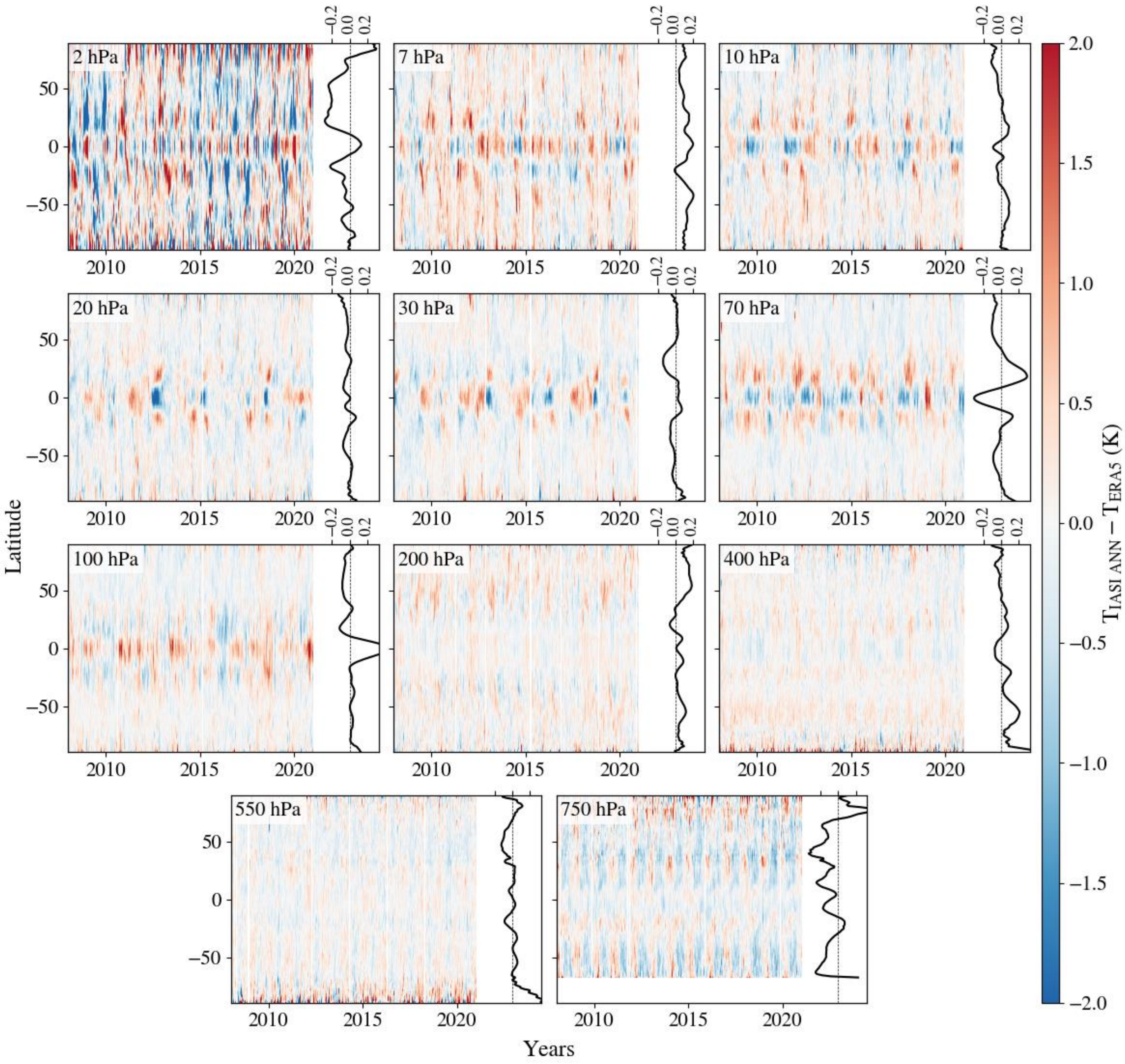


Figure 3: Daily zonal mean differences between IASI and ERA5 zonal mean temperature for the 11 pressure levels of the ANN, with the time averaged differences on the right of each subplot.

l.201-204 - Does the fact that a latitude-longitude gridding is being used affect the RMS calculation? Obviously, a 1x1 degree lat-lon bin is much smaller at high latitudes than in the tropics, so there will be fewer IASI data points going into it. Are there sufficient numbers of points in each bin that the RMS is not likely to be affected by this?

On average, there are between 20 and 25 points per bin each day, and it can go up to 27 for latitudes around 75°N or S. At latitudes larger than 75°, the number of points per bin decreases and there are bins with less than 5 points, but these only concern latitudes larger than 88°.

l.206-207 - The increased RMS above mountain ranges could also be due to atmospheric gravity wave (GW) activity. If there are orographic GWs present over the mountains in the IASI measurements that are even slightly different from those simulated in ERA5 in terms of phase, amplitude, intermittency or location, this will likely result in a higher RMS value than a region with low GW activity.

The sentence was modified:

"At 750 hPa, RMS values are small at the equator (about 0.5 K) and larger at higher latitude (between 1 and 2 K), especially around mountain ranges, where they reach 3 K and can be due to gravity wave activity."

Sect 3.2 l.214-234 - Firstly, the description of the ARSA radisondes could go in a Data section. Secondly, there is a long paragraph here from l.216-231 where it is not entirely clear what processes are applied to the radisonde datasets by whom and which of these steps are relevant for the present study. l.231 onwards "ARSA provides a 43 pressure-level profile..." should be near the top of the paragraph for readability, or even in the Data section. The authors should also decide how best to describe the relevant quality controls and extrapolation steps applied to the ARSA data for readability, because at the moment it is a little confusing which parts are relevant.

Thank you. Indeed, you are correct. The description of ARSA was adapted and moved to Section 2 (see first comment).

l.233 - I think the authors mean "substitute ERA-Interim for ERA5"?

The sentence was modified: *"... ARSA is being reprocessed to replace ERA-Interim with ERA5".*

l.250 and Fig. 6 - Normally it is good to show the highest time resolution possible, but would these figures be better simply showing the monthly averaged differences? This also would help to overcome the poor time coverage of the high latitude radisonde stations. The figures are also not high enough resolution to see individual daily differences anyway.

The poor time resolution is not a problem to overcome. On the contrary, averaging the differences over a month would probably underestimate the differences, as averaging the differences over the stations in each region already does.

Fig. 6 - As mentioned above, the information in Fig. 6 would be very well summarised by a line plot showing time-averaged temperature differences (x-axis) against altitude (y-axis) for each radisonde region. The different lines for the different regions could be neatly colour-coded like the authors have done in Fig. S3. I think this could be very clear and may be worth including.

As for the comparison with ERA5, putting all the profiles on the same plot was not very clear, so the profiles were added on the right of each subplot:

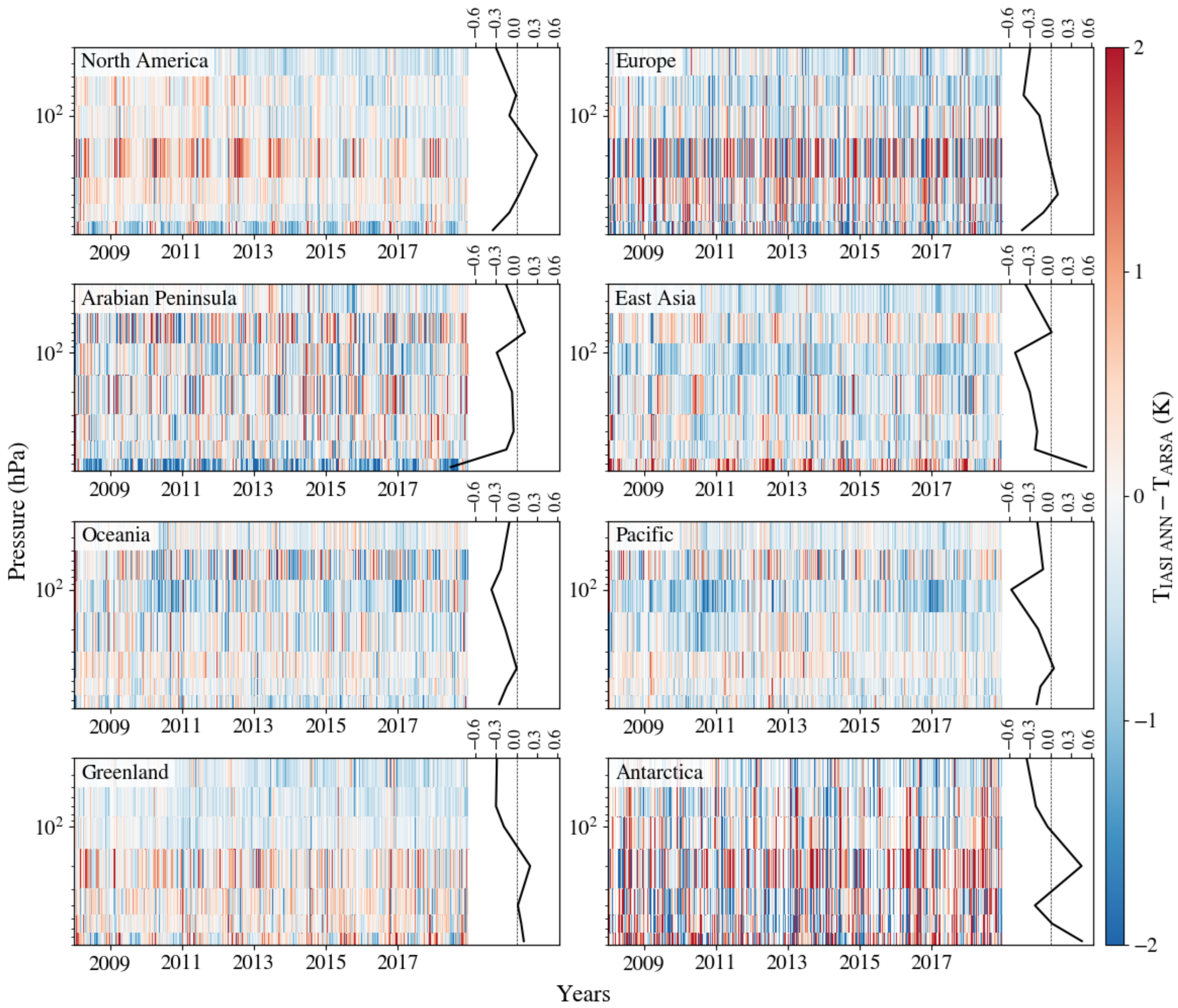


Figure 7: Daily differences between IASI and ARSA temperatures between 2008 and 2018 in North America, Europe, the Arabian Peninsula, East Asia, Oceania, the Pacific, Greenland and Antarctica, with the time average differences profiles on the rigt of each subplot.

l.264-265 - How much of this persistent positive temperature difference at 2hPa could be due to biases in the ARSA dataset, and not due the IASI retrieval? Looking at Fig. S1, there is a very similar positive bias when the ARSA radisondes are compared to ERA5. There could therefore be a small temperature bias at these altitudes in the ARSA data, perhaps due to the additional datasets that are used to extrapolate or constrain the radisonde data at these altitudes?

Yes, as suggested by the first reviewer, we removed the comparison above 30 hPa in the figure, because the bias probably comes from the extrapolation data so the comparison is not with radiosoundings anymore. This was clarified in Section 4.2:

"We only show differences between 750 and 30 hPa as ARSA data above 30 hPa does not always come from radiosounding measurement but from the extrapolation datasets."

l.269-272 - Related to the point above, it would be very useful to include in the supplementary material the exact same figure as Figs. 6 and S1 but for the differences between IASI and ERA5 for each region. This could help the authors to more confidently assess some of the observed temperature differences in different the regions.

Figure S2 shows the differences between IASI and ERA5 at the time and location of ARSA observation. The following sentences were added in Section 4.2:

"Figure S2 shows the differences between IASI and ERA5 temperatures interpolated to the time and locations of ARSA observations. In most regions, the differences are less than 0.5 K. In Antarctica, in Europe (troposphere only) and in the Arabian Peninsula and Oceania (stratosphere only), the differences can reach 1 K."

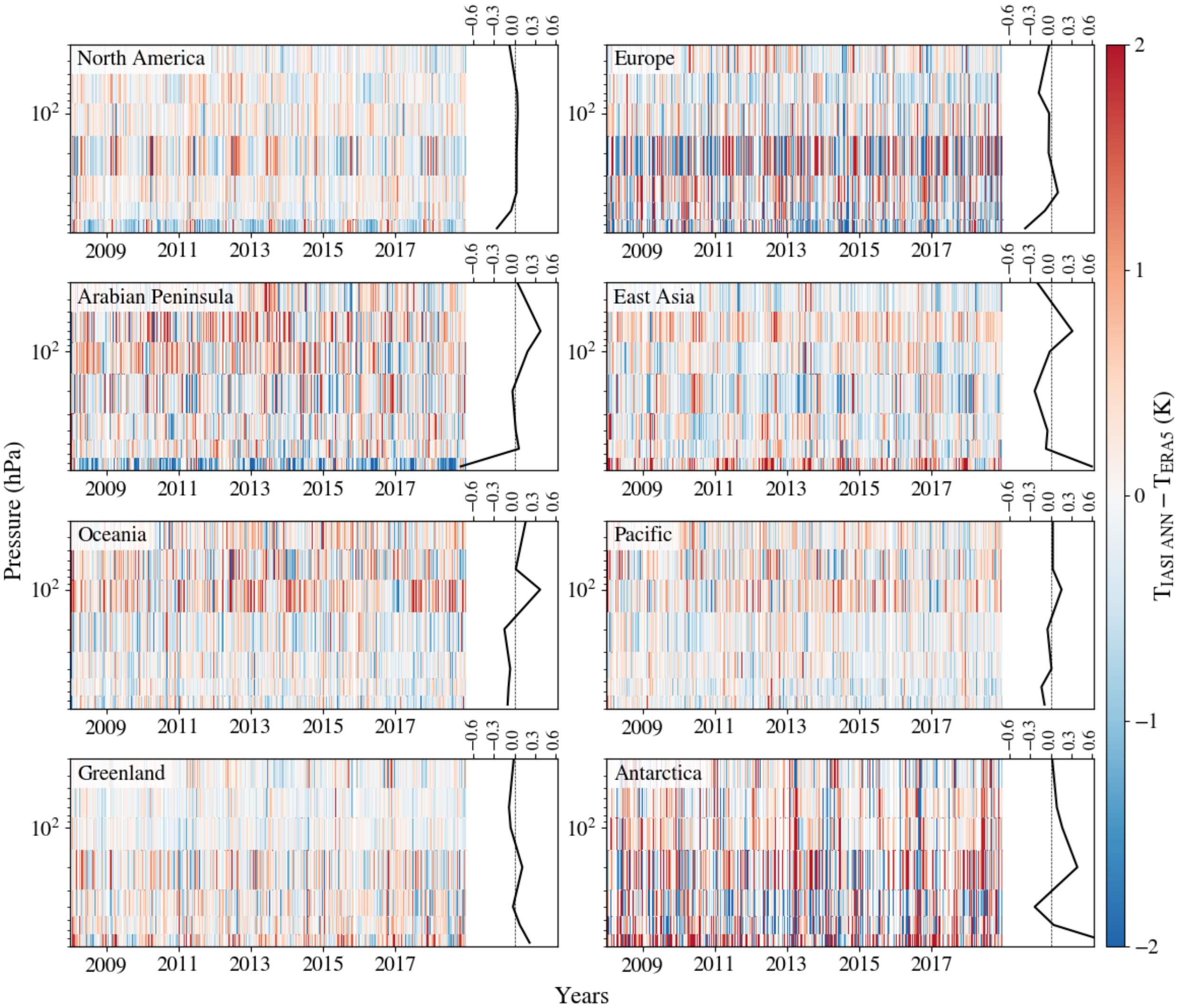


Figure S2: Differences between IASI-ANN and ERA5 temperatures at the time and location of ARSA observations, averaged daily in each of the regions.

Fig. 7 - [Important] The authors should explain, or at least discuss, the thin vertical red stripes that appear in some of the panels in Fig. 7 (such as in the 30hPa panel). They should clarify whether these are artefacts that result from their analysis or if they are physical. I did wonder if they were due to re-initialisations of the EUMETSAT retrieval, or perhaps even due to sudden stratospheric warmings. The authors should discuss.

The red stripes are an artefact of the analysis: this comparison was done using monthly files of ERA5 data (as the comparison between IASI-ANN, EUMETSAT CDR and ERA5 was done all at once), and for the last day of each month, when both the file of one month and the next were needed for the interpolation, there was a problem in concatenating the two files. Due to the large computing time, we did not redo the comparison. We clarified this in the description of the figure:

"Figure 7: Daily differences between IASI and ARSA temperatures between 2008 and 2018 in North America, Europe, the Arabian Peninsula, East Asia, Oceania, the Pacific, Greenland and Antarctica, with

the time average differences profiles on the rigt of each subplot. The red stripes seen in some panels are artefacts from the analysis and the do not reflect a physical phenomenon."

l.320 - "...although the areas of strongest warming are slightly different." It would be useful to briefly describe what these differences are if the authors are going to mention the Shangguan et al. (2019) study.

As suggested by the first reviewer, the contributions of ENSO of MEI were removed before computing the trends and tropospheric trends are now very similar to those found by Shangguan et al. The sentence is now:

"The values of the trends we found between 45°S and 45°N are similar to those found by Shangguan et al. (2019)."

l.342-345 - The DOIs listed do not appear to be working correctly, please check. Also, the authors could consider using the accepted short doi service for readability (https://shortdoi.org/).

The DOIs seem to be working, but for more clarity they were shortened to https://doi.org/hbxm (Metop-A) and https://doi.org/hbxn (Metop-B):

"This dataset is available from https://iasi-ft.eu/products/atmospheric-temperature-profiles/ (doi for Metop-A temperatures: https://doi.org/hbxm and doi for Metop-B temperatures: https://doi.org/hbxn)."

l.354 - Are both of these southern warming regions in Fig. 8 due to ozone hole recovery or just the region over the pole?

The warming over the pole is due to ozone hole recovery and the warming at 50°S is most likely due to the Sudden Stratospheric Warming that happened in September 2019, but both regions are impacted by the two phenomena in different proportions. The sentence was changed to "… there are two regions with important warming due to the ozone hole recovery and a SSW that happened in 2019."

One final general point, it might be very useful for the community if the authors could say something about the vertical resolution of their retrieval. Naturally, the retrieval is evaluated on 11 pressure levels, but if the authors were able to estimate the vertical resolution of the retrieved temperature at for each of these levels that would be very useful if other researchers wanted to investigate gravity wave observations in the dataset, in a similar to what has been done in many studies for AIRS (Hoffmann and Alexander, 2009). Further to this, are the retrieved ATPs also archived on the satellite scantrack or is only the global 1x1 degree grid available?

The vertical resolution goes from 5 to 12 km from the lower to the upper troposphere. In the stratosphere, the vertical resolution goes from 12 km in the lower stratosphere to 25 km above 7 hPa. This was added in Section 3.2.:

"At the selected pressure levels, the vertical resolution goes from 5 to 12 km from the lower to the upper troposphere. In the stratosphere, the resolution goes from 12 km in the lower stratosphere to 25 km above 7 hPa."

Only the 1°x1° grids are archived on the website.