Response to Referee #1

The authors would like to thank the referee for her/his general comments about the manuscript and her/his useful suggestions and corrections, which have helped us clarifying several points and improving the manuscript. Below are our responses to the comments brought up by the referee. The referee's comments and our responses are marked in blue and in black, respectively. In italic are the changes made in the manuscript.

General: The article "Validation of the IASI FORLI/Eumetsat ozone products using satellite (GOME-2), ground-based (Brewer-Dobson, SAOZ) and ozonesonde measurements" submitted to AMT by A. Boynard et al. describes comparisons between IASIO3 retrieval data products and several other O3 data sets. They employ the FORLI version v20151001 retrievals. They used various time periods within the overall periods of observations IASI-A (2008-2017) and IASI-B (2013-2017) for overlap with other instrumental observations. The data have global coverage and is used in intercomparison between –A & -B. Other latitudinal or hemispheric comparisons are made with sparse ground based observations. Comparisons are also made with total and partial column products.

Overall the article describes comparisons with 4 different datasets in a logical manner. There is considerable detail for any one of the comparisons that could be clarified better. That descriptions are brief may be necessary since there are several datasets to describe and there are references to previous work. Still the main points are not as forthcoming as they could/should be. There are no new techniques nor sophisticated procedures or concepts hence it should be clear where and especially why the comparisons are the state they are in. This intercomparison is nearly identical to a previous comparison by the same author Boynard et al., 2016 yet not mentioned much. This current paper describes the latest FORLI version and the previous paper a previous FORLI version. But that hardly makes this new work, in fact many plots and tables are identical. This work should be cast as an update and a comparison to the previous FORLI version. In this way specific details on how the new version improves or changes O3 columns and partial column data are explicit. This is a large shortcoming of this submission and should be remedied before publication.

In order to take into account Reviewer #1's main concern, we better quantify the improvements of the new version of FORLI in comparison with the previous one in the revised manuscript. The improvement for ozone partial columns is also discussed. We also highlight the fact that the improvement is rather constant over the globe and therefore issues are still persisting over some regions such as high latitudes, mountain region and desert. Specific studies have been initiated with different validation groups in order to assess the reasons for the larger differences and will be the object of an independent study. Here are the changes made in the revised manuscript:

<u>GOME-2 comparison section:</u>

"*Globally, IASI-A (IASI-B) TOC product are slightly higher than GOME-2A TOC product, with a global mean bias of 0.3±0.8 % (0.4±0.8 %). It is worth noting that the previous IASI TOC product (v20140922) was in disagreement by more than 5 % (Boynard et al., 2016). The global mean bias is now within total errors of GOME-2 estimated to 3-7 % (Valks et al., 2017) and IASI, which demonstrates the good consistency between IASI and GOME-2 TOC products.*"

<u>Brewer/Dobson comparison section:</u>

"*Nevertheless the overall comparison with Dobson and Brewer TOCs shows that IASI new TOC product is improved by 4% in comparison with the previous IASI TOC product (v20140922; see Boynard et al. (2016)) and is within IASI and GB TOC total error bars.*"

SAOZ comparison section:

"*The results are consistent with those found for the comparison with GOME-2A along with Brewer and Dobson measurements (see Sections 5.1 and 5.2, respectively). An improvement of 3-4 % is found when compared to the previous IASI product (v20140922).*"

Ozonesonde comparison section:

"*In comparison with the previous IASI partial ozone column products reported in Boynard et al. (2016), the new IASI ozone product is significantly improved in the MS by 8-12 % for the mid latitudes and tropics. The improvement is less significant for the LMS except in Antarctic where an improvement of 6 % is found. As for the TROPO and UTLS columns, no or slight improvement (<2 %) is found, and the agreement between IASI and sonde data is even worse compared to the previous IASI ozone product, especially for the southern tropical TROPO column (by 7 %) and the UTLS column (by 10-18 %)*"

The IASI O3 retrieval is performed in the 1025-1075cm-1 IR region. Yet there are no comparisons with IR derived data sets. Such a comparison would diminish any discrepancy with cross section differences between IR and UV / Vis. This comparison would be seen as more thorough and results very interesting to take advantage of IR ground based datasets. Further, in particular NDACC IR data have vertical information comparable to IASI (DOFS ~>4) to use for partial columns. Secondly, there is little discussion given to any contribution of cross section differences.

We thank the referee for his/her suggestion of adding a comparison with IR remote sensed data. We followed his/her suggestion by comparing IASI and FTIR TOCs and partial ozone columns for several FTIR stations. As for the discussion about any contribution of cross section differences, we discuss it in the new IASI/FTIR section. Here are the changes made in the revised manuscript regarding the new IASI/FTIR comparison :

2. IASI measurements and independent datasets used for the validation

2.3 Ground-based data

Daily TOC measurements from Dobson and Brewer UV spectrophotometers available […] are used for IASI-A and IASI-B TOC validation.

Regular ozone measurements from high-resolution solar absorption spectra recorded by GB FTIR (Fourier transform infrared) spectrometers available for the period 2008 – 2017 were downloaded from NDACC. The ozone FTIR retrieval principle, which is based on the optimal estimation method (Rodgers, 2000), as for FORLI, is detailed in Vigouroux et al. (2008). Such measurements have the advantage to provide not only TOCs with a precision of 2 %, but also low vertical resolution profiles with about four independent partial columns, one in the troposphere and three in the stratosphere up to about 45 km, with a precision of about 5-6 % (Vigouroux et al., 2015). Therefore, the FTIR measurements are used to validate not only IASI TOCs but also IASI partial ozone columns. The stations considered in the present work were used in several papers for trend analyses (Vigouroux et al., 2008, 2015; García et al., 2012; Wespes et al., 2016) and validation studies (Dupuy et al., 2009; Viatte et al., 2011). The latitudinal coverage ranges from 67.8°N to 45°S, so only the southern high latitudes are not covered. The location of the six FTIR stations used in the comparison is given in Table 1 and presented in Fig. 2. Since these solar absorption measurements requires daylight conditions, there is no measurement at Kiruna during polar winter. All stations use the high-resolution spectrometers Bruker, which can achieve a resolution of 0.0035 cm-1 or better. Details on the harmonized retrieval parameters can be found in Vigouroux et al. (2015). For all stations, the 10µm spectral region is fitted to retrieved O_3 using two retrieval algorithms: either PROFFIT9 at Kiruna and

Izaña or SFIT2/4 at the other stations. The two algorithms have been compared in Hase et al. (2004). The spectroscopic database used is HITRAN 2008 (Rothman et al., 2009). Each station is using the daily pressure and temperature profiles from NCEP (National Centers for Environmental Prediction) and has one a priori profile, which is obtained from the same model WACCM4 (Whole Atmosphere Community Climate Mode; Garcia et al., 2007).

3. Comparison methodology

3.2 Comparison with FTIR and ozonesonde data

For the comparison between IASI data against FTIR and sonde TOCs and partial ozone columns, the coincidence criteria used in this study are the same as those defined in Boynard et al. (2016), except for the time coincidence which is slightly different in order to be more consistent with the temporal variability of tropospheric ozone: we apply coincidence criteria of 100 km search radius and ±6 h. As the ozonesonde measurements are mainly performed in the morning (local time), this implies that most of the pixels meeting these coincidence criteria correspond to pixels of the IASI morning overpass, which is not the case for FTIR measurements that can be performed all day long.

In the comparison with FTIR data, the FTIR retrieved profiles are adjusted following Rodger and Connor (2003, their Eq. 10) in order to take into account the different a priori profiles used in both IASI and FTIR retrievals:

$$\boldsymbol{x}_{adjusted,FTIR} = \boldsymbol{x}_{FTIR} + (\boldsymbol{A}_{FTIR} - \boldsymbol{I})(\boldsymbol{x}_{a,FTIR} - \boldsymbol{x}_{a,IASI}) \quad (1)$$

where $\boldsymbol{A}_{FTIR}$ is the FTIR AK matrix, $\boldsymbol{I}$ the unity matrix, and $\boldsymbol{x}_{a,FTIR}$ and $\boldsymbol{x}_{a,IASI}$ the FTIR and IASI O_3 a priori profiles, respectively.

In addition, when validating satellite profile products, a proper comparison method is to account for the difference in vertical resolution. In the present work, the ozonesonde and adjusted FTIR profiles are first interpolated on the corresponding IASI vertical grid and then degraded to the IASI vertical resolution by applying the IASI AKs and a priori O_3 profile according to Rodgers (2000):

$$\boldsymbol{x}_s = \boldsymbol{x}_a + \boldsymbol{A}(\boldsymbol{x}_{raw} - \boldsymbol{x}_a) \quad (2)$$

where $\boldsymbol{x}_s$ is the smoothed ozonesonde/FTIR profile, $\boldsymbol{x}_{raw}$ is the ozonesonde/adjusted FTIR profile interpolated on the IASI vertical grid (referred as "raw" FTIR), $\boldsymbol{x}_a$ is the IASI a priori profile and $\boldsymbol{A}$ the IASI AK matrix. Incomplete ozonesonde profiles above ozonesonde burst altitude are filled with the a priori profile.

For each ozonesonde/ FTIR measurement, we calculate the TOCs (only for the FTIR data) and the four partial columns defined above from all IASI and smoothed ozonesonde/FTIR profiles meeting the coincidence criteria, then we average all IASI and smoothed ozonesonde/FTIR total and partial columns. In the end there is one IASI-DATA profile pair per ozonesonde/FTIR measurement. To avoid unrealistic statistics skewed by extremely unrealistic low values in the UTLS O_3 columns found in the smoothed ozonesonde data, we filter out extreme outliers exceeding 200 % relative differences with IASI (which can be up to ~8 % of the data in the tropics).

5. Validation results

5.4 Comparison with FTIR TOCs and partial ozone columns

Figure 12 shows the temporal variation of the monthly mean relative differences between IASI-A and IASI-B against FTIR TOCs convolved with the IASI averaging kernels according to Eq. (2) for the six FTIR stations (see Table 1 and Fig. 2 for their location) for the period 2008 – 2017. Compared to FTIR, the IASI-A and IASI-B TOCs *are negatively biased by 0.8-6.2 % with the largest biases (-4.1 % and -6.2 %) at Jungfraujoch and Lauder, respectively. At Lauder, mean biases of 5.7±5.4 % and 0.6±6.4 % between FTIR and IASI against Dobson TOCs, respectively, are found, suggesting that the FTIR data might be biased high at that station, but 4 % of this bias between FTIR and Dobson is likely due to the known inconsistency between IR and UV cross-sections (Gratien et al., 2010) (note that the bias is calculated as [100x(FTIR-DOBSON)/DOBSON] or [100x(IASI-DOBSON)/DOBSON]). It can be noted that the bias between FTIR and IASI-A, and SAOZ and IASI-A for close latitude stations are very consistent, if one takes this spectroscopic bias into account (i.e. UV Sodankyla lower than IASI-A by 3.9%, FTIR Kiruna higher by 1.1 %; UV OHP lower than IASI-A by 1.0 %, FTIR Jungfraujoch higher by 3 %; UV Kerguelen higher than IASI-A by 0.9 %, FTIR Lauder higher by 6.2 %).*

At Zugspitze and more particularly at Jungfraujoch, two jumps are visible in 2010 and 2014, with larger biases before 2011 and after 2014 with respect to the period in between. It is worth noting that these two jumps seem to coincide with changes in IASI L2 temperature (in September 2010 and September 2014). The analysis of surface temperatures used in both IASI (Eumetsat) and FTIR (NCEP) retrievals (IASI L2 Eumetsat and NCEP, respectively) shows that the differences between Eumetsat and NCEP can reach up to 20 K for the surface temperature and vary between -10 and 10 K along the temperature vertical profile at both Jungfraujoch and Zugspitze while at the other stations the differences are much lower (less than |5| K), which suggests that IASI L2 Eumetsat temperatures are less reliable above elevated areas. However a more in-depth analysis is needed and for that matter is in progress in order to understand the exact origin of the jumps found in the differences between IASI and FTIR TOCs at these stations.

The dominant systematic uncertainty in FTIR O_3 retrievals is due to the spectroscopic parameters (García et al., 2012). The IASI retrieval algorithm uses HITRAN 2012 and the FTIR retrieval algorithm uses HITRAN 2008, however no differences were found in the O_3 absorption band, respectively (Boynard et al., 2016). We do not expect a significant bias between the IASI and FTIR total columns due to ozone spectroscopy, because both retrieval algorithms use the same ozone spectroscopic parameters and the same fitting spectral range. Except at Lauder and Jungfraujoch, the mean biases between IASI and FTIR TOCs are relatively low and within total errors of FTIR (e.g. García et al., 2012) and IASI, which shows again the good quality of IASI TOC data.

Except at Jungfraujoch and Zugspitze, the IASI-A and FTIR TOC monthly relative differences show insignificant drift less than |0.9| % decade^{-1} (see Fig.12 and Table 2), which is among the 1 – 3 % decade^{-1} Ozone_cci requirements for the long-term stability for total ozone measurements (Van Weele et al., 2016), demonstrating that the current IASI-A TOC products are homogeneous and reliable for trend studies. The significant negative drifts found at Jungfraujoch and Zugspitze, are explained by the bias drop observed from 2014 that is discussed above.

Since FTIR data also provide up to four independent pieces of information in the vertical ozone profile, we now assess four IASI partial ozone columns characterized by a DOFS of ~1 (surface-300 hPa, 300-150 hPa, 150-25 hPa and 25-3 hPa), which should make such assessment meaningful. The comparisons of the four partial ozone columns between IASI-A and FTIR performed for the period 2008 – 2017 are presented in Fig. 13. The correlation coefficients between FTIR and IASI-A partial columns are good to excellent (from 0.72 to 0.98), with the highest correlations found in the UTLS and LMS.

For all stations except Kiruna, IASI tropospheric column is negatively biased by 5-14 %. The comparison for the UTLS O_3 columns shows that IASI-A O_3 product is positively biased at all stations (except at Izaña), with the largest bia found at Wollongong (21.1±19.9 %) and the lowest bias found at Jungfraujoch (3.7±15.0 %). The standard deviation is maximum in the UTLS at Izaña and Lauder, which is due to strong O_3 variability and large total retrieval error in this region as shown in Wespes et al. (2016). It should be noted that IASI is positively biased in the UTLS region, as reported in previous studies comparing IASI to ozonesonde data (e.g. Boynard et al., 2016; Dufour et al., 2012; Gazeaux et al., 2013). Although Dufour et al. (2012) attempted to give some explanations for this particular feature, the exact reason for this overestimation is still not clear. One reason could be the use of inadequate a priori information. Note that FORLI uses only one single a priori profile (Hurtmans et al., 2012) that *is the global mean profile of the McPeters/Labow/Logan climatology (McPeters et al., 2007). As shown by Bak et al. (2013), using tropopause-based ozone profile climatology can significantly improve the a priori. However, using dynamical a priori makes the comparison on a global scale less straightforward since a different a priori profile would be used at each IASI pixel. The best correlation coefficients and smaller standard deviations (in %) between IASI-A and FTIR data are found for the LMS column. The small standard deviations in the LMS comparisons allow the detection of consistent IASI-A negative biases at all stations (5-9%). This consistent negative bias in the LMS, where the ozone partial column contributes the most to the total column, is reflected in the observed negative bias on TOC discussed above. These better correlation coefficients and standard deviations in LMS are due to the better IASI sensitivity to this column (mean DOFS ~1.2 – 1.5 as indicated in Fig. 13) compared to the other partial columns. The smallest biases between FTIR and IASI-A columns are found in the MS column (-0.2 / +4.9%), except at Kiruna where the bias reaches 13 %. This higher bias at Kiruna might be due to a bad collocation of sounded air masses which can be in different in or out polar vortex conditions for the two instruments. The FTIR instrument sounds the atmosphere along the line-of sight instrument-sun, therefore the sounded air masses at this higher partial column and for high solar zenith angles measurements might be far away from the station itself (few hundreds kilometers). A collocation with the satellite that would take the FTIR line-of sight into account, would improve the comparisons.*

A similar picture us found for the comparison between IASI-B against FTIR partial ozone columns over the period May 2013- 2017 (not shown).

The stability of IASI-A partial ozone columns is also assessed based on the time series of monthly relative differences between IASI-A and FTIR data over the period 2008 – July 2017. Table 3 gives the decadal drift values along with their 2-σ standard deviations in % decade^{-1} as well as the P-value. As a reminder the trend is considered significant if the drift value is higher than its 2-σ standard deviation. For the TROPO column, we clearly see a significant negative drift at all stations ranging from -5.0±4.8 % decade^{-1} (Izaña) to -16.1±8.1 % decade^{-1} (Kiruna). Smaller or insignificant drifts are found in the UTLS and LMS. Regarding the MS, insignificant positive drifts are found, except at Izaña where a positive drift is found (3.7±2.5 % decade^{-1}). As a consequence, the stability of the IASI-A partial O_3 columns when compared to the six FTIR GB measurements that cover the IASI measurement period and that are characterized by limited vertical sensitivity cannot be confirmed.

To answer that question, comparisons of IASI partial O_3 columns with ozonesonde measurements that provide numerous highly resolved vertical O_3 profiles is performed in the section below.

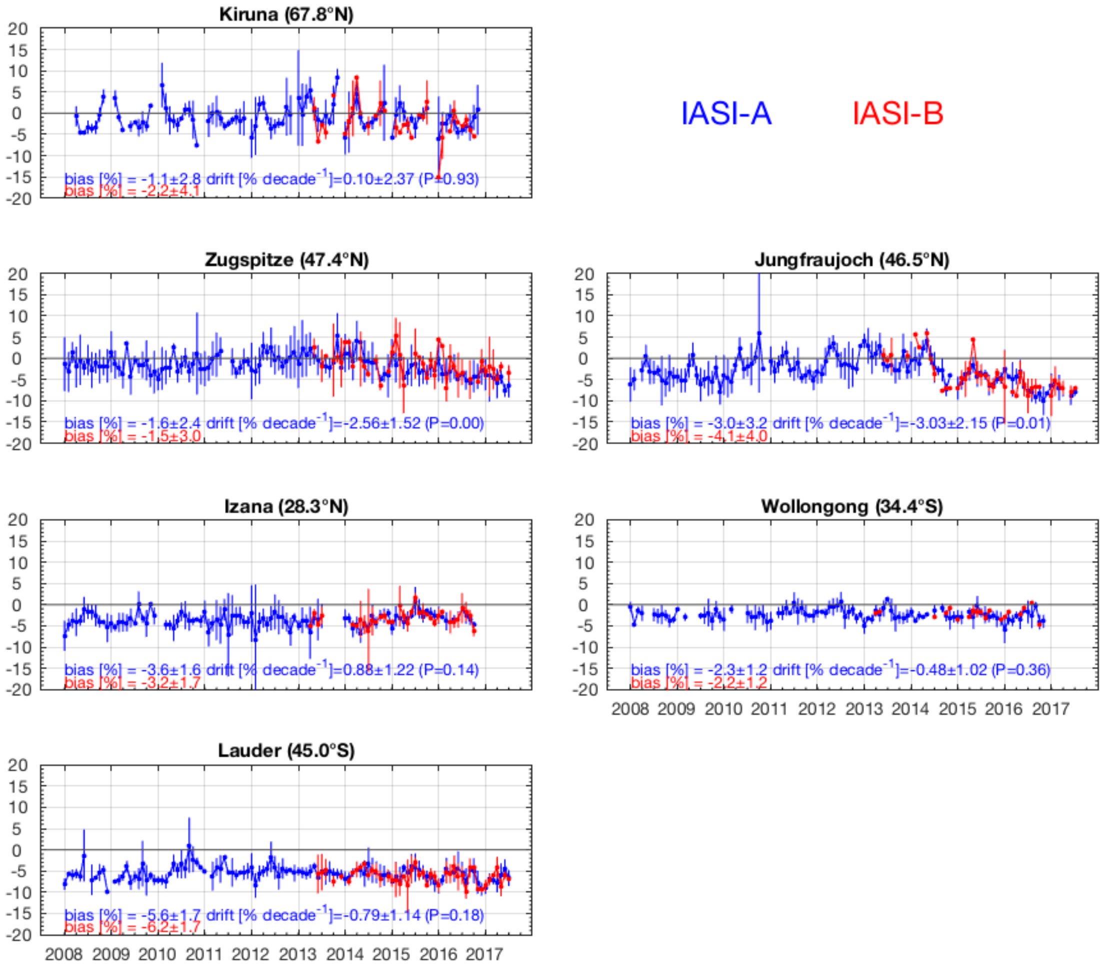

Figure 12: Time series of the monthly relative differences (in percent) between IASI-A (blue) and IASI-B (red) against collocated FTIR TOC measurements for five stations from North to South. For each daily FTIR measurement, a relative difference is calculated as 100 x (IASI – FTIR) / FTIR [%]. All the relative differences are then monthly averaged. For the period May 2013 onwards, only the common collocations between IASI-A and IASI-B are shown. The standard deviation of the average is also displayed (vertical bars). Comparison statistics including mean biases and standard deviations in percent for the common period May 2013 – July 2017, the decadal drift (in %) and its 2-σ standard deviation along with the *P* value for the IASI-A time series are indicated on each panel.

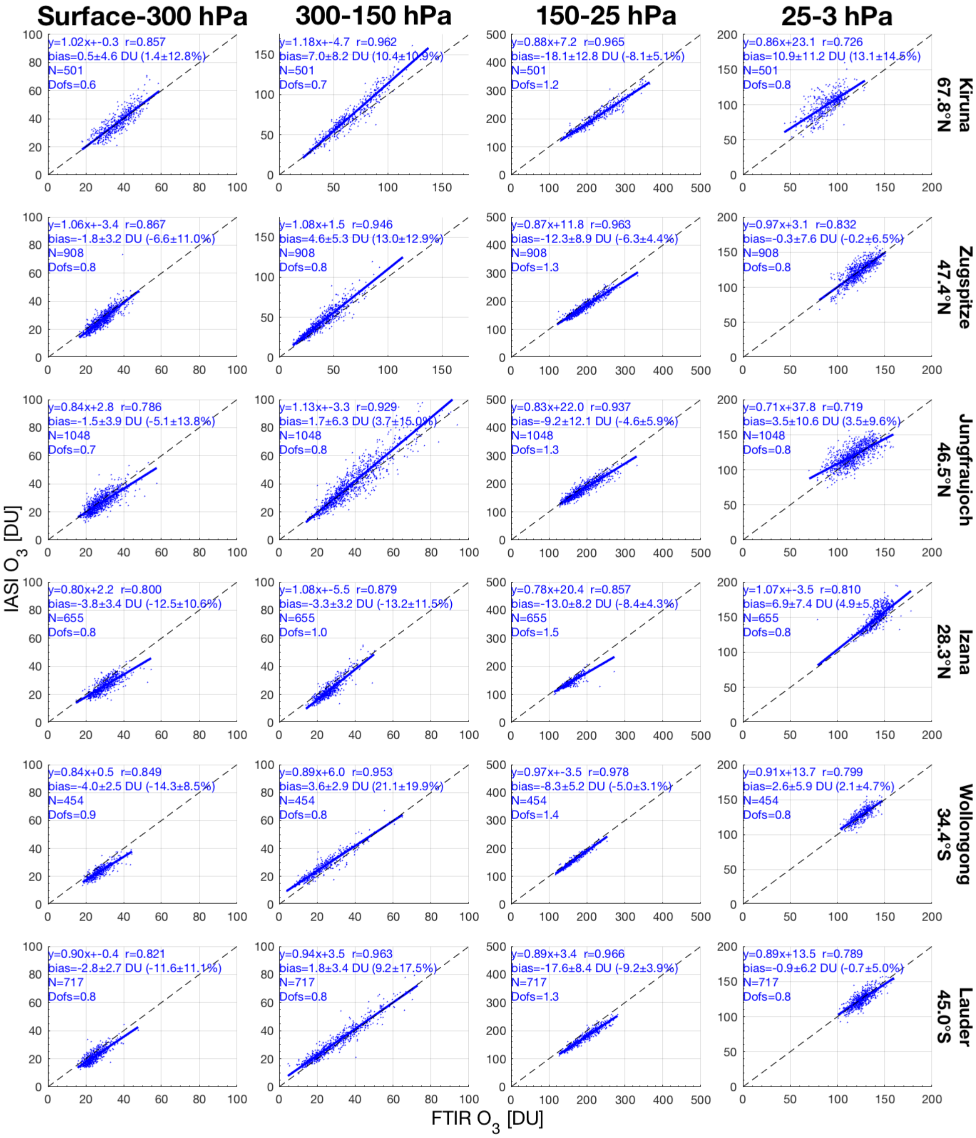

Figure 13: Scatter plots of IASI-A against smoothed FTIR O_3 partial columns at six FTIR stations for the period 2008 –2017. Comparison statistics including the linear regression, the mean differences and standard deviation in both DU and %, the number of collocations and the mean DOFS for each partial column are shown on each panel.

Specific: Throughout the text the adjectives 'good' or 'generally' are used in descriptions of a comparison. These qualitative comments do not help the reader nor are they appropriate. They are subjective and are detrimental to a real grasp of the state of the IASI data with regard to other pertinent datasets. There are many uses of approximately (~) or less then (<) that seem inconsistent and hence then to obfuscate the real quality of the data.

We removed the adjectives 'good', 'generally', ~ and < as much as possible.

Here are specific issues with the scientific points being made.
1. P1, L23 Brewer & Dobson TOC are not retrievals per se.
We removed the word 'retrieved'.

2. P1 L25 to wit "shows good long term stability" good relative to what?
We removed the word 'good'.

3. P1 L 29 "Compared to ozonesonde data, IASI-A and IASI-B O3 products overestimate the O3 abundance in the stratosphere 30 (up to 20 % for the 150-25 hPa column) and underestimates the O3 abundance in the troposphere (within 10 % for the mid- latitudes and ~18 % for the tropics). This sentence is needlessly confusing mixing zonal and altitude comparisons and using hPa layers an "troposphere".
We changed the sentence to:
"*Compared to ozonesonde data, IASI-A and IASI-B O_3 TROPO column (defined as the column between the surface and 300 hPa) is positively biased in the high latitudes (4-5 %) and negatively biased in the mid-latitudes and tropics (11-13 % and 16-19 %, respectively).*"

4. P1 L32 "small" compared to what?
We removed the word "small".

5. P2 L21 "180 shift" is not clear. It's a shift in what?
Metop-A and Metop-B satellites are 180° out of phase and thus for one specific location one satellite may be before or after the other. We made this part clearer as follows:

"The two Metop satellites are on the same orbit with Equator crossing times of 09:30 (21:30) local mean time solar time for the descending (ascending) part of the orbit. There are therefore numerous common observations between two consecutive tracks. However, since Metop-A and Metop-B are 180° out of phase, there is a ~50 min temporal difference between both instruments (one satellite might be before or after the other); thus the observations are never quite simultaneous."

6. P4 L24 what is a O3 profile "C-shape"
A C-shape O_3 profile is a profile characterized by an abnormal increase in O_3 at the surface, as shown in the figure below. We rephrased the sentence in the revised manuscript as follows:
"(vii) the O_3 profiles have an unrealistic C-shape *(i.e. abnormal increase in O_3 at the surface, e.g. over desert due to emissivity issue)*, with a ratio of the surface – 6 km column to the total column higher or equal to 0.085;"

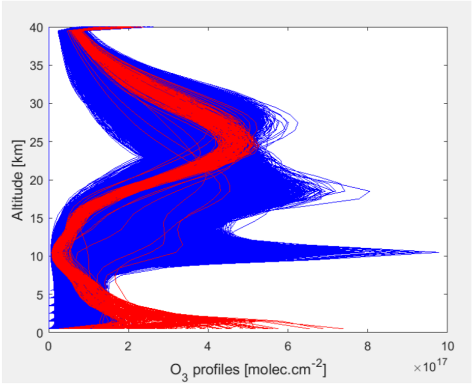

Figure: Example of IASI vertical profiles for one day. The red profiles correspond to C-shape profiles.

7. P5 L 9 " Differences between IASI-A and IASI-B. Is the plot A-B or B-A ? More generally for all comparisons in this paper most are ambiguous on this simple point. Every instance in the text and captions should be made explicit.

The relative difference between IASI-A and IASI-B is calculated as: 100 x (IASIA – IASIB) / IASIA. We made it clearer in the manuscript as follows:

"*Before validating IASI-A and IASI-B O_3 products, we assess the consistency between both instruments over the common period May 2013 – July 2017. For the intercomparison exercise, we first calculate the daily IASI-A and IASI-B averages over a 1°x1° grid. Then for each 1°x1° grid cell, we calculate the relative difference as 100x[(IASI-A – IASI-B)/IASI-B]. Finally we calculate the monthly averaged data from the daily gridded differences. A statistical analysis of IASI-A and IASI-B TOCs and TROPO O_3 columns is performed with respect to time and latitude.*"

The formulae of the difference calculation is also indicated in the revised captions.

Following a comment from Referee #2, we have moved the dataset characteristics and validation method in two sections:

- **2. IASI measurements and datasets used for the validation**: this section describes the IASI *O_3* retrievals as well as the independent measurements used for the validation
- **3. Comparison methodology**: this section includes the formulae of difference calculation as well as the different comparison methods used in the present work.

8. P6 L2 'excellent agreement' How is excellent agreement defined? Is there a reference for comparison of spectra?

We changed 'excellent agreement' to 'better quality' and we rephrased the sentence as follows:
"*As a result, since October 2015, the IASI-A and IASI-B spectra are of better quality/stability (Buffet et al., 2016; Jacquette et al., 2016).*"

Buffet, L., Villaret, C., Jacquette, E., Vandermarcq, O., Astruc, P., and Anstötz, S. :Status of IASI instruments onboard Metop-A and Metop-B satellites, 4th IASI International Conference, Antibes Juan-Les-Pins, France, 11-15 April 2016, https://iasi.cnes.fr/sites/default/files/drupal/201612/default/bpc_iasi-conference4-1_02_instruments_buffet.pdf, 2016.

Jacquette, E., Maraldi, C., Standfuss, C., Coppens, D., Delatte, B., Baqué, C., Calvel, J.-C., Buffet, L., Vandermarcq, O. : IASI performance assessment after permanent cube corner compensation device stop, 4th IASI International Conference, Antibes Juan-Les-Pins, France, 11-15 April 2016, https://iasi.cnes.fr/sites/default/files/drupal/201612/default/bpc_iasi-conference4-s1-08_jacquette.pdf, 2016.

9. P7 L21-27 Here are given possible sources of differences at high latitudes between IASI & GOME data. They are apparently (per reference) the same or very similar to Boynard 2016. These do not help the reader know if and/or how FORLI v20151001 is an improvement. For instance given GOME data quality and known issues with UV/Vis retrievals from GOME, what changes in v20151001 make it an improvement over v20140922? More specifically i) 'limited information content. . .low surface temperatures'. Is this due to lower spectral SNR? Or something else? ii) misrepresentation of surface emissivity is vague. Are these values have large errors? Or have large biases or both? Is there a reference? iii) if the temperature profiles are known to be less reliable this implies variability not necessarily bias (but this list is describing a GOME/IASI bias. Further what magnitude temperature error delivers the O3 bias seen?

In Boynard et al. (2016) we identified a systematic bias between the IASI data and observations obtained using the UV spectral range. The amplitude of the bias varies in latitude, and is more pronounced at higher latitudes. In v20151001 we fixed this issue by modifying the following :

- the loop-up tables (LUT) have been recalculated to cover a larger spectral range (960-1105 cm^{-1} versus 960–1075 cm^{-1}), with correcting numerical implementation in FORLI v20151001, especially with regard to the LUTs at higher altitude.
- the HITRAN spectroscopy database was updated from HITRAN 2012 to HITRAN 2004

But these changes are not fully efficient at high latitudes, and as the reviewer noticed, the exact cause of the discrepancy left is still under investigation. To answer the comments in more details, we added the following arguments to the revised manuscript:

"*Despite the global improvement of ~5 % with the new IASI TOC product with respect to the previous IASI TOC product (v20140922), large discrepancies are still observed at high latitudes and are partly explained by:*

i) *the low spectral signal to noise ratio due to very low surface temperature in this region leading to limited information content in the IASI observations in these regions;*

ii) *a misrepresentation of the wavenumber-dependent surface emissivity, which is a critical input parameter to describe the surface, especially above continental surfaces (Hurtmans et al., 2012). FORLI uses the emissivity climatology built by Zhou et al. (2011) providing weekly emissivity values on a 0.5°x0.5° latitude/longitude grid for all 8461 IASI spectral channels. However, Zhou et al. climatology can have missing values. In such cases, the MODIS climatology built by Wan (2006), which provides values for only 12 channels in the IASI spectral range is used instead. Furthermore, in case of no correspondence between the IASI pixel and either climatologies, the reference emissivity used for the Zhou climatology (Zhou et al., 2011) is used, which can significantly impact the retrievals, in particular in arid or semi-arid regions where variations in emissivity are large both on spectral and spatial scales (Capelle et al., 2012) but also in ice region since the reference emissivity does not necessarily reflect the actual snow or sea ice coverage;*

iii) *the temperature profiles used in FORLI-O_3 that are less reliable at high latitudes and over elevated terrain (August et al., 2012). As shown in Boynard et al. (2009), the errors introduced by the uncertainties of 2 K on the temperature profile can reach up to 10 % of total error on the retrieved vertical profile, with the error due to the temperature uncertainty on the TOCs being much lower.*

Errors on thermal contrast can also have an impact on the retrievals.

iv) the errors associated with TOC retrievals in the UV-vis spectral range increasing at high solar zenith angles in these regions, mostly because of the larger sensitivity of the retrieval to the a priori O_3 profile shape (Lerot et al., 2014).

In the section below, a detailed analysis of the larger bias found in the Antarctic region is undertaken for individual ground-based Brewer and Dobson station to try to understand the larger bias (see next section)."

10. P8 L1 Is there some explanation what the physical basis is for the rejection of data over deserts and Antarctica?

Data characterized by a C-shape profile, which is not realistic, are generally located over desert and Antarctica. A possible explanation of this issue is a misrepresentation of emissivity above sand and ice surfaces. More explanations on the impact of misrepresented emissivity on ozone retrievals are given in the previous answer and were added to the revised manuscript.

11. P11 L18 Its not completely clear are all comparisons of the upper most partial column to 10hPa despite the statement to the contrary on L17 (25-3hPa)?

As ozonesonde profiles sonde generally burst around 30-35 km, the middle stratosphere upper was limited to 10 hPa. We made it clearer in the revised manuscript (new comparison methodology section) as follows:

3. Comparison methodology

"*The IASI-A and IASI-B O_3 products are assessed in terms of TOCs and partial ozone columns. The validation exercise is performed using the same partial columns as those used in Wespes et al. (2016) since these columns contain around one piece of information, have maximum sensitivity approximately in the middle of each of the layers, and reproduce the well-known cycles related to chemical and dynamical processes characterizing these layers: surface-300 hPa (TROPO), 300-150 hPa (UTLS), 150-25 hPa (LMS) and 25-3 hPa (MS). On average, these pressure columns correspond to the following altitude columns: surface-8 km, 8-15 km, 15-22 km 22-40 km, respectively. Note, however, that for the comparison between IASI and ozonesonde data, the MS is limited to the column 25-10 hPa as sonde generally burst around 30-35 km (see Section 3.2 below). For the assessment of IASI vertical profiles, we refer to Keppens et al. (2018, this issue).*"

12. P11 L22 What is the source of the extremely low O3 in the UTL? Is it low in the sonde, IASI or both? How much data is lost?

The extremely low O_3 in the UTLS (<1 DU) concerns the sonde smoothed with the IASI averaging kernels. Up to 8 % of the data in the tropical UTLS are lost. We added this information to the revised manuscript as follows:

"*To avoid unrealistic statistics skewed by extremely unrealistic low values in the UTLS O_3 columns found in the smoothed ozonesonde data, we filter out extreme outliers exceeding 200 % relative differences with IASI (which can be up to ~8 % of the data in the tropical UTLS).*"

13. P12 L8 It is very reasonable to follow that the high variability in the UTLS give large SD but less so to see what the a priori has to contribute given DOFS 1.

We changed to sentence to:

" *The standard deviation is maximum in the UTLS in all latitude bands (compared to the other partial columns) due to strong O_3 variability and large total retrieval error as shown in Wespes et al. (2016).*"

14. P12 L14 Its not clear what information is missing? The comparison is to 10hPa for both instruments – is that correct?

Yes that is correct. We made this point cleared in the revised manuscript as follows:

"*The correlation coefficient ranges between 0.6 (tropics and high latitudes) and 0.8 (mid-latitudes) for the LMS column while they are much lower for the MS column, which is explained by the low DOFS values ranging between 0.4 and 0.6 as indicated on the scatter plots. Note that the DOFS for the MS columns are lower than those calculated in Fig. 13 because they do not correspond to the full MS column calculated from IASI (25-3hPa i.e. ~25-40km) but to the MS columns truncated to match the maximum altitude (30-35km) of the sonde measurements.*"

15. P12 L29 Please clarify are these 40 pairs globally?
We rephrased to sentence to:
"*The long-term stability of IASI-A partial O_3 column vs ozonesonde measurements is assessed in Figure 16, which presents the monthly relative differences between IASI-A and ozonesonde for the TROPO, UTLS, LMS and MS O_3 partial columns for a total of 18 ozonesonde stations in the NH that cover eight years or longer (over 2008 – 2017). With more than 30 IASI-sonde pairs per month, the NH presents sufficient collocated data to assess a good statistical drift analysis on the contrary to the SH (only 8 ozonesonde stations).*"

16. P13 L21 Please provide a reference for the stability of the IASI radiances.
We removed this sentence.

17. P13 L31 use of the adjective generally is not helpful.
We removed the word 'generally'.

Technical:

1. P1 L19 what does ir "generally consistent" versus just "consistent"
We have changed 'generally consistent' to 'consistent'

2. P1 L23 "retrieved ones" would better be "retrieved TOC's"
As highlighted by Referee #2, Brewer/Dobson data are not retrieved products. We removed the word "retrieved" and changed "ones" to "TOCs".

3. P1 L24 "on the instruments" would better be "on the compared instruments"
We have made the change.

4. P1 L30 (up to 20% for the 150-25 hPa column) should be (up to 20% for the 150-25 hPa partial column)
Following a previous comment of the referee, this sentence has been rephrased.

5. P1 L30 "within" might better be "less then"
We have changed 'within' to 'less than'.

6. P2 L 4 "amount" should be "amounts"
This has been corrected.

7. P9 L2 'latitude belt' might be better worded 'latitude band'
This has been corrected.

8. P3 L11 "overestimates the ultraviolet (UV) Total Ozone Column (TOC)" Do you really

mean to differentiate the uv TOC from some other spectral region?
Only comparisons between IASI and UV-vis Total Ozone Column have been performed so far that is why we only refer to bias between IASI and UV-vis data. However this does not mean we differentiate the UV TOC from other spectral region such as TIR. We added this sentence:
"*It is worth mentioning that Boynard et al. (2016) did not perform any comparison with measurements in other spectral ranges than UV.*"

9. P5 L19 'posteriori' should be 'after'
We have made the change as follows: '*after September 2015*'

10. P5 L 31 remove second 'between April. . .'
We have made the change.

11. P6 L1 'proved' might better be worded 'shown'. Fruthermore is there a reference for this statement/conclusion?
This has been corrected. We added references for this statement.

12. P6 L13 'being preferentially be used' might better be worded 'are recommended'.
This sentence has been changed to:
"*Even if for the period April – September 2015, IASI-B O_3 products are better recommended for a high quality use, it is worth noting that the IASI-A instrumental issue only affects the TOC by 0.4 % and the tropospheric ozone by 10 %, which are much lower than the TOC and tropospheric total retrieval errors estimated to 2 % and 20 % on average, respectively, justifying the potential use of the IASI-A data over that period if it is required. In the validation exercise presented in the next section, the period April-September 2015 is included.*"

13. P12 L20 From plot 14 I read +1.5 & – 3.5% difference (not +- 3.5) and SD of maximum 14.6% (not 14)
We changed the sentence to:
"*The comparison is good for all latitudes, with IASI-A O_3 products underestimating the TROPO O_3 abundance in the mid-latitudes and tropics by ~1.6-3.5 DU (7.1-14.3 %) and overestimating the TROPO O_3 abundance in the high latitudes by ~1.5 DU (4–7 %).*"

14. P12 L 23 Dobson units (sp)
This has been corrected.

15. P12 L28 . . .O3 partial columns. . .
The word "partial' has been added.

16. P13 L20 . . .proven to be very. . .
This sentence has been removed.

17. P14 L7 'better from October' might be better worded 'better after October'
Since the change in the IASI-A viewing angle was corrected in September, we have changed 'better from October' to 'better after September'

18. P33 caption, use of the term 'sub' column should be removed (in all cases) and partial be used for consistency
We have changed 'sub-column' to 'partial column'