Answers to reviewer #1 :

1. The new SOFRID product, v3.5, shows well visible stripes, especially in the southern hemisphere. Even if the authors mentioned that these discontinuities are consistent with their retrieval errors, these unphysical discontinuities would bring some difficulties to compare and evaluate models for example. The authors show how important it is to compare raw and smoothed data to satellite observation validation but it would be the same for model comparison and these stripes will compromise the comparison. The authors should mention these difficulties in using their product for model comparison and provide some possible ways to overcome these artifacts.

Please see our reply to reviewer #2 to a similar comment (comment #2).

2. The authors state, especially in the conclusion, that the improvement of their new approach (dynamical a priori) is mainly due to the consideration of the tropopause height in the choice of the a priori. But they do not demonstrate the impact of taking an a priori profile on this basis, as they do not consider this selection independently from the latitude and season selection. Intuitively, we would expect that a better consideration of the tropopause would help resolving the biases in the UTLS, but no specific improvement are shown between v1.6 and v3.5. I would suspect that the bias correction is more related to the latitude/season selection as the v1.6 was based on a NH a priori only. The authors should better discuss this in the paper and show how the tropopause selection impact their retrieval if they think it is a key point.

The same point regarding the relative importance of the tropopause-, latitude- and month-dependence of the a priori has ben raised by reviewer #2. The important point is that tropopause and « latitude/season» are strongly correlated and it is therefore not possible to fully adress this question. The Sofieva et al. (2014) climatology provides an information about the intra-seasonal tropopause related variability of the O3-profile on top of the larger seasonal variability. We have performed a sensitivity test to separate the impact of the intra-seasonal tropopause-dependence from the seasonal variability. The results are given and discussed in our answer to reviewer #2 (comment #3).
Concerning specifically this comment of reviewer #1: it is true that the UTLS bias has not been corrected by the dynamical a priori which is an important result: our papers further demonstrates that this bias is not or little related to the a priori as highlighted in the paper. Nevertheless, the UTLS variability has been largely improved with v3.5 (see Taylor diagram Fig. 5a) showing the advantage of a climatological a priori for UTLS retrievals. As discussed in the answer to reviewer # 2, part of this improvement is due to the intra-seasonal variability of the a priori.

Moreover, the authors state that this dynamical approach based on the tropopause selection is presented for the first time. It is not completely true. The authors should refer to different publications using the KOPRAFIT-O3 algorithm they mentioned in the paper in which the selection of the a priori (and regularization) is based on the tropopause height (Dufour et al., ACP, 2015, 2018 and Eremenko et al., JQSRT, 2019).

We acknowledge the work of Sellito et al. (2013) (reply to public comment by Pasquale Sellito in this discussion), Dufour et al. (2015) and Eremenko (2019) with KOPRAFIT-O3 in the manuscript. The approach by Sellito (2013) and Dufour (2015) are similar and rather basic using only 2 and 3 different a priori profiles for tropical and mid latitudes and for high, mid and tropical latitudes respectively. Eremenko et al. (2019) developped a more sophisticated method that has only been tested on synthetic data and not on real satellite observatioons. We have therefore added the following statements:

- In section 2.2: «… In a first attempt to take this tropopause effect into account for satellite data, Sellito et al. (2013) have implemented 2 a priori profiles in the KOPRAFIT-O3 retrieval algorithm to basically discriminate tropical (tropopause higher than 14 km) from other latitudes. Dufour et al. (2015) have slightly improved the approach with a set of 3 a priori profiles for high latitudes (tropopause lower than 10km), mid-latitudes (tropopause between 10 and 14 km) and the tropics (tropopause higher than 14 km). Eremenko et al. (2019) have tested a set of N profiles for retrievals on a synthetic database.»
- In the conclusion: «… Other satellite O3 retrievals use a priori profiles from climatologies but they are chosen based on geographical and temporal criteria only (Bowman et al., 2006; Liu et al., 2010). Dufour et al. (2015) use 3 different a priori profiles picked up according to 3 broad tropopause height classes to represent high, mid and tropical latitudes...»

In the present paper we use hundreds of a priori profiles based on latitude (10° bins), month and tropopause (1 km bins). It is therefore really the first time that such a *comprehensive* approach is used for real data retrievals. Therefore:

- in the abstract we keep our statement just adding the word *comprehensive* to differentiate our approach from Dufour et al. (2015) : «For the first time we have implemented a *comprehensive* dynamical a priori profile for spaceborne O3 retrievals which takes the pixel location, time and tropopause height into account for SOFRID-O3 v3.5 retrievals.»
- similarly, in the conclusion we have added *in such a comprehensive way* : « To our knowledge it is the first time that the tropopause height is used *in such a comprehensive way* for the choice of the a priori for spaceborne O3 retrievals.»

3. Concerning the comparison with FORLI, as the authors use information from literature, the sampled pixels are likely different between the two algorithms and it can have a possible impact on the statistics, in particular if the cloud mask considered by the two algorithms is different. No information concerning the cloud filtering is mentioned for both SOFRID and FORLI. This should be added and discuss.

The same issue was raised by reviewer #2.
We have added missing information concerning our cloud mask and we have made sensitivity tests with different values of our data filters such as cloud mask, cost function and DFS. In each case, the general statistics on which the FORLI-SOFRID comparisons are based are not altered.
See our detailed reply to reviewer #2 comment #4 « *Section 5, p.13, L9-10* ».

4. The quality of presentation of the results is sometimes poor and not suitable for publication. Figure and Table captions miss a lot of information such as units. A lot of typos remain. The authors should have read carefully their manuscript. In section 5 the authors referred to FORLI 16 and 18 on the Figures they comments but the Figures available online only show information named FORLI in green, whereas in the submitted paper for the quick review, both were present. Please, be consistent between the figure and the text.

We have completed the captions of the Figures and Tables with missing information such as units. We have also improved other presentation details mentioned by the reviewer. We have removed remaining mention to FORLI 16 and 18 as we have just kept comparisons with the latest FORLI-O3 version.

Specific comments:
- p1, line 18: should we read theoretical or theoretically?

Theoretical as the information is theoretically computed.
- P2, line 31: troposphEric
OK
- P3, line24: quantify "weakly contaminated"
We have added our AVHRR cloud fraction cover threshold in the manuscript : 25 %.
- P4, line 1 and 20: Is there a difference for ozone between HITRAN2004 and HITRAN2008?
A major update concerning line positions, intensities and lower state energy has been made for the three main O3 isotopologues in HITRAN2008 according to Rothman et al. (2009). Nevertheless, this update marginally concerns the 9.6 microns absorption band used for IASI O3 retrievals. We therefore think that it does not affect significantly SOFRID O3 retrievals.
- P5, line 11: emmissivity should be emissivity
OK
- P5, Equation 1 : G is not define in the text.
We have added its definition : « G is the gain matrix that represents the sensitivity of the retrieval to the measurement. »
- P5, line 22, change "devided" to "divided"
OK
- P6, line 13: change "chinese" to "Chinese"
OK
- P7, line 20: could you please clarify if the monthly mean is computed with at least 3 profiles by stations or by latitude bands? If it is by latitude bands, is it sufficiently representative?
In fact we use at least 4 profiles (more than or > 3) per latitude band. We have corrected in the paper :
« … at least 4 coincident profiles within this latitude band »
We have performed tests with higher and lower numbers of sondes required per month and latitude band. We found that 4 was a good compromise between a representative monthly O3 in a 30° band and two many lacking months in the time serie. Requiring more profiles mostly impacted the time series in the southern tropical band which is rather poorly sampled. We discuss this issue «only two stations (La Reunion and Nairobi) provide data regularly (30-50 profiles per year) over the period» and its implications (bias variability over the period) in the paper.
- P7, line 21: Jcost is not defined
We have added its definition « retrieval cost function ».
- P8, line 8: change "interanual" to "interannual"
OK
- p8, line 13: documentS
OK
- Table 2: please specify the units.
OK
- P10, line 16: mid anD high latitudes
OK
- P10, line 19: improvEment
OK
- P10, line 27 and p11, line26: change "dissapears" to "disappears"
ok
- P12, line 19: missing units
OK
- P13, lines 10-13: the discussion is not clear. The authors state first the reason of the differences is the noise level and then it is not clear. They should provide the noise levels for the different algorithms to elaborate their hypothesis.

We have added a ref to Dufour et al. (2012) where the noise levels are given for both algorithm. This is enough for the qualitative argument concerning the DFS given here.
« This probably results from the retrieval noise level which is lower for FORLI than for SOFRID (Dufour et al., 2012) »
- *P13, lines 6-34: the discussion is not consistent with the figure (no FORLI 16 and 18 display in the Figures).*
OK
- *P14, line 5: change "positives" to "positive"*
OK
- *P15, line 12: drifTs*
OK
- *Figure 1: what are the error units? The way the authors present the plots with +/- for different variables is not very conventional. They should explain more precisely in the caption how to read the figure (this is also the case for Figs. 13 and 14*).
The units have been added. The errors are given as negative values with « -ERROR » indicated on the y axis. This way of presenting allows to plot two variables at the same time.
We have added (-) in the caption for things to be clearer.
- *Figure 2: units are missing*
Units « Dobson Units (DU) » has been added.
- *Figure 6: explain RS and SmRS, please*
The captions have been updated.

Answers to reviewer #2 :

1/ My major comment is related to the validation methodology used by the authors; Ido have doubts about the fact that ozonesonde data are used both for building apriori (single and variable) and for the validation. It is commonly accepted that one specific dataset or instrument cannot be used both for the apriori used for the retrieval and for the validation of the corresponding retrieved product, for evident reasons. Even if the IASI period validated here (2008-2017) is different from that used to build the apriori (1980-2006 for V3.5 and 2008-2009 for V1.6, hence, the WOUDC measurements used to generate the V1.6 single apriori are included in the validation dataset), I'm wandering to what extent it might affect the results.

We do not agree with the reviewer about the fact that the use of O3 profiles from the WOUDC ozonesonde dataset for both the a priori and the validation could cause a problem in the validation methodology. Moreover we are not aware of publications which would have highlighted this issue.
We have mostly two objections about the reviewer statement #1:
- First and most importantly, the a priori for an OEM algorithm is not equivalent to a training dataset for an AI (Artificial Intelligence) or a NN (Neural Network) retrieval algorithm or to the ensemble used to constrain a model to provide analyses with a data assimilation systems. For an AI or NN algorithm, the retrieved quantities are strongly bounded within the variability of their training datasets and for an assimilation system the analyses will likely provide better comparisons if compared with assimilated data. For an OEM algorithm, the a priori is built from an ensemble of data (mixture of ozone sondes and satellite datasets in our case) to provide the best knowledge of the **average** state and of its **variability**. The ozone sonde profiles from the ensemble are not used individually to train or constrain the algorithm. Therefore we can consider that our a priori data and each sonde profiles are completely independent.
- Second, the O3 sonde instruments are state of the art calibrated and validated in situ instruments. They provide O3 concentrations as close as possible to the « truth ». Each sonde can reasonably be considered as independent from each other. They are very different from remote sensing measurements with limited vertical sensitivity and likely systematic biases. Using observations from a satellite instrument to build an a priori and validate the instrument using this a priori with the same data would raise issues related to the reviewer concerns.

The a priori contribution contained in the retrieved product would tend to improve the comparison. That a priori contribution can be easily calculated and should be discussed in the validation section. Please discuss that point.

The a priori is used to complete the information provided by the instrument in the part of the state vector space (O3 profile in our case) where the instrument is not providing information (see Rodgers 2000 for instance). Understandably, this information has to be as accurate as possible. I will rather return the question:
why should we use a single a priori when there are comprehensive and state of the art climatologies available which give better results ?
See reply to next comment which raises the same concern. The quote of Rodgers (2000) (who theorised the OEM for atmospheric soundings) advocates for a climatological latitude and time dependent a priori.
The a priori contribution is theoretically calculated and provided in the paper as the smoothing error based on the retrieval equation (Eq. 1 and Fig. 1 in the manuscript).
Furtjhermore, the paper gives a rather complete evaluation of the a priori «contribution» on the retrieval comparing retrievals from V1.6 and V3.5. The conclusion is that using a good a priori significantly improves rather than « *tends to improve* » the retrievals based on comprehensive comparisons with 10 years of global ozone soundings.

2/ Section 3.3: Even if not necessary for a pure validation exercise, the comparison with raw vs smoothed data is interesting as it allows a better evaluation of the O3 variability captured by the instrument. However, one should note that when considering variable apriori (according to season, location and tropopause height), a part of the expected variability is injected by default in the retrieved product through the a priori contribution, making the comparison with raw data wrongly improved when using variable vs single apriori. In addition, the presence of visible stripes (Section 2.4) due to the use of variable apriori that depend on location may constitute an issue for further comparison study, e.g. with CTM. This is exactly why, one can usually prefer using a single vs variable apriori profiles; it gives a homogeneous retrieval at the global scale and the retrieved variability is not distorted by that of the variable apriori. Hence, the true capability of the pair instrument/algorithm to capture the O3 variability is better infer when using a single apriori profile. That point should be clearly discussed in Sections 3.3 and 4.

Here again we do not agree with the reviewer. A single a priori has long been used in SOFRID for the reasons mentioned by the reviewer and also because it is easier to build and easier to use. These are probably good reasons. Nevertheless, since that time some thorough O3 climatologies based on ozonesonde and satellite data such as Sofieva et al. (2014) have been made available and are used for other sensors (TES/OMI) retrievals as mentioned in the manuscript. They are the «most satisfactory» choice of a priori according to the textbook « inverse methods for atmospheric soundings » of Rodgers (2000). See for instance p166:

« The most satisfactory source of a priori information is from independent high spatial resolution measurements […] as may be obtaind from radiososnde measurements. Such data is often available as climatologies partitioned by, for example, latitude and date».

We fully acknowledge the stripes visible in the global distributions with V3.5. As mentioned in the manuscript they are due to the smoothing error with which they are in good quantitative agreement (< 5DU). If we look at tropospheric O3 in the SH mid-latitudes, we indeed see some marked stripes of about 2.5-5 DU which are of course errors. If we now look at the differences between SOFRID v1.6 and raw sondes we have an average bias of ~ 30% (Fig. 9e) while this difference drops to ~ 0% with V3.5 (fig. 11e). In the first case we have a smooth distribution with large but «invisible» biases and in the second case we have some visible effect of the a priori but low biases. We think that the retrievals are better in the second case. We have chosen to show the striped distributions to clearly document that issue for SOFRID data user.
There is also the improvement concerning the NH seasonal variability which is significant and very satisfactory. We think that these are not «wrong» improvements but just clear and documented ones. SOFRID is not IASI but a comprehensive system based on RTTOV and 1DVar algorithms with IASI radiances, ECMWF auxiliary data and a priori data as input. The validation exercice we have performed gives an evalutation of the whole system. For model comparisons there is no real issue because (i) the problem is already clearly acknowledged in the paper in case of comparisons with raw model data and (ii) as the IASI retrievals are the validating datasets (contrarily to here where they are the validated datasets) the modeled profiles should be smoothed by the retrieval AvKs which will take the issue of the variable a priori into account. To make things clearer, we added the following comment at the end of section 2.4 :

« Such stripes may appear as a problem for the use of SOFRID v3.5 data for model validation. They are a minor problem for two main reasons. First, as is demonstrated in next section, the use of a dynamical a priori largely improves the retrieved O3 profiles. Second, when model profiles are compared to SOFRID retrievals the impact of the a priori profile is taken into account by using Eq. 1 such as in Barret et al. (2016)»

3/ Through Section 4, the authors insists on the fact that "the improvement of SOFRID accuracy . . .is the clearest advantage of using a dynamical apriori profiles". Given that several sources of improvement are taken into account: dependence on tropopause height, latitude and month, how can the authors be able to dissociate between their respective effects? Please, provide sensitivity tests or clarify that point? Comments 2/ and 3/ highlight the limitations in using variable apriori and evaluating the V3.5 product. The authors should better discuss those issues through the manuscript in order to get a better feeling for the real advantage of using variable apriori (in terms of both location,season and dynamical tropopause).

In our introduction we define « a dynamical a priori profile for spaceborne O3 retrievals which takes the pixel location, time and tropopause height into account » and not only the tropopause. Therefore the statement mentioned by the reviewer about SOFRID improvements is correct and the important point is that the improvements are significant using such a « dynamical » a priori.
The validation has been performed in 30° latitude bands with monthly means. Therefore an important part of the answer is clearly in the paper. Indeed, the large improvement in the seasonal variability in the NH mid (and high) latitudes results from the monthly a priori. As these seasonal variabilities are latitude dependent it also highlights the importance of a latitude dependent a priori.
As the tropopause height is largely month- and latitude-dependent it is not possible and it would be artificial to fully « dissociate » the impact of the three parameters on the SOFRID improvements: a climatological a priori is implicitly tropopause dependent. Nevertheless, it is possible to assess the difference between a fully tropopause dependent a priori and a climatological a priori with an implicit tropopause dependence. This has been achieved with a sensitivity test with a single a priori (the one corresponding to the highest occurrence from Sofieva et al. (2014)) for each month and each 10° latitude bands therefore removing the intra-seasonal tropopause variability from the a priori choice.
The results are similar to those of the v3.5 highlighting that the improvement are little dependent on the intra-seasonal variability of the a priori profile. Nevertheless, v3.5 is better concerning the TOC variability in the 30-60°N band which is the most significant region in terms of sonde sampling and in the 60-90°S band. In the UTLS v3.5 is also better in terms of variability and correlation coefficients in most latitude bands.
Therefore, we have changed our manuscript in order to document the fact that the largest part of the improvement is due to the use of a climatological a priori dependent on month and latitude. This is of general interest for other scientists which are working on O3 retrievals and could use simpler climatologies. A new section (4.2 Impact of the intra-seasonal tropopoause dependence of the a priori profile on SOFRID improvements) including a figure with a Taylor diagramm presenting TOC and UTLS columns (Figure 6) has been added to illustrate this point.

We have also added a sentence in our conclusion:

«A sensitivity test demonstrated that these SOFRID improvements are dominated by the seasonal- and latitude- dependence of the a priori.»

4/ Regarding the comparison with FORLI, the authors are very negative through the manuscript and the critics are most of the time out of context.

We agree and we have improved the manuscript being more positive with FORLI. Nevertheless, we would like to drow the reviewer attention to the fact that our initial statements were based on results published by the FORLI team.

For instance:
- In the abstract: "(iii) in the N.H., no significant temporal drift is detected in SOFRID contrarily to FORLI (∼8%)"

This statement is based on Boynard et al. (2018) :
- in the text: «Based on the drift value with the 2σ standard deviation and the value (indicated on each plot), the derived **drifts** […] are **statistically significant** for the TROPO [...] columns (−8.6±3.4 % decade−1…)».
- in the conclusion : « A **significant negative drift** of −8.6 ± 3.4 % decade−1 is also found in the IASI-A to ozonesonde TROPO O3 column comparison for the Northern Hemisphere. »

Nevertheless, we have used « jump » instead of « drift » in the abstract.

« in the northern hemisphere, the 2010 **jump** detected in FORLI TOCs is not present in SOFRID ».

- Introduction, L21: "They both document a problem (drift or jump) . . ."

The full statement is «They both document a problem (drift or jump) in the O3 retrievals around year 2011 but this **does not hinder** the fact that TOC are decreasing according to Wespes et al. (2017). »
This is rather positive acknowledging the possibility to use the data for trends analysis as done in Wespes et al. (2017).

Following the reviewer recommendation about the use of « jump » rather than « drift », we have modified the statement as follows «…They both document a **jump** in the O3 retrievals in 2010 which does not hinder … »

- Section 5, p.14, L.7-9: "the SOFRID NH tropospheric drifts discussed in section 4.3 are smaller and opposite in sign to the significant -8.6±3.4%/dec drift between FORLI and smoothed sonde data in the NH troposphere presented in B18."
That comparison of the "drift" calculated from SOFRID vs FORLI does not make sense. Indeed, the authors have to make a clear distinction between a "drift" that usually refers to an instrumental drift in validation studies, and a "jump" (or sudden discontinuity) as observed in the FORLI dataset, which induces an artificial drift, in order to avoid any confusion. It has already been clearly explained and discussed in Boynard et al. (2018) and in Wespes et al. (2018; 2019): the drift strongly decreases (< 1DU/dec on average) after the jump and it becomes even non-significant for most of the stations over the periods before or after the jump, separately. The discontinuity is strongly suspected to result from updates in level-2 temperature data from Eumetsat, which occur at the same date of the detected jump and which are used as inputs into FORLI. Hence, it is obvious that "No significant change occuring around 2010 is detectable for SOFRID v1.6 (Fig. 8(h)) and v3.5 (Fig. 10(h)) NH time series", given that SOFRID uses L2 from ECMWF, not from EUMETSAT. It should be clarified through the manuscript.

As mentioned above, we based our comments and our mention of a **drift** on the recent papers concerning FORLI-O3 (Boynard et al., 2016, 2018 and Wespes et al. 2018). The two validation papers present the same 2010 «jump» even though it has not been clearly documented in Boynard et al. (2016). We understand the reviewer argument concerning the difference between a «jump» and a « drift » in FORLI O3 data. Nevertheless, it was not clear in the validation papers of the FORLI team (Boynard et al. , 2016 and 2018) and posterior publications. On the contrary:

- In the statement cited above from *Boynard et al. (2018)*, the words **statistically significant drift** are used.
- In *Wespes et al. (2018)* we read : « Note, however, that a **drift** in the NH middle-low troposphere (MLT) O3 over the whole IASI dataset is reported in Keppens et al. (2018) and Boynard et al. (2018) from comparison with O3 sondes. »

- in *Keppens et al. (2018)* : « Looking at latitude-resolved drift studies for the Ozone_cci IASI-A nadir ozone profiles (not shown), a **significant decadal negative drift** of the order of 25 % or higher can be observed in the Antarctic UTLS and the northern hemispheric troposphere. »

Concerning the cause of this jump, the reviewer mention «...i*s obvious that "No significant change [...] given that SOFRID uses L2 from ECMWF, not from EUMETSAT. It should be clarified through the manuscript. »*
We do not agree. The reason for the «jump» is not hypothesised as resulting from EUMETSAT L2 discontinuity in Boynard et al. (2018) and Wespes et al. (2018). More specifically :

- in the AMT Discussion of *Boynard et al. (2018),* Reviewer #2 stated «Unfortunately the **significant drift** in the troposphere is barely explained and addressed». The authors replied « … a few more years are needed to confirm the observed **negative drifts** and evaluate it on the longer term… », statement which can be found in both the text and the conclusion of the final version of the paper.
- in *Boynard et al. (2018),* the EUMETSAT L2 discontinuity is mentioned «It is worth mentioning that the EUMETSAT dataset is not homogenous, as it has been processed using different versions of the IASI Level 2 Product Processing Facility between 2008 (v4.2) and 2016 (v6.2)», but it is not clearly mentioned as an explanation for the TROPO-O3 drift/jump.
- in *Wespes et al. (2018) :* « This **drift** (∼ 2.8 DU decade−1 in the NH) is shown in Boynard et al. (2018) to result from a discontinuity (called "jump"by the author) in September 2010 in the IASI O3 time series, for **reasons that are unclear** at present.»

Therefore, in these publications, no evidence (based on the EUMETSAT L2 products for instance) is given to explain the drift/jump.

It is in the two latest publications, *Keppens et al. (2018)* which is a general paper dealing with nadir ozone products and *Wespes et al. (2019)* dealing with Antarctic stratospheric O3, that the hypothesis of a causal link between EUMETSAT L2 and FORLI-O3 discontinuity is mentioned:
- *Keppens et al. (2018)*: «Part of the **overall negative tropospheric drift** of the FORLI v20151001 IASI retrievals **could, however, be due** to a change in the processing of the IASI L2 processor (e.g. temperature profile) at EUMETSAT that changed to version 5.0.6 in September 2010.»
- *Wespes et al. (2019)*: «The discontinuity is **suspected** to result from updates in level2 temperature data from Eumetsat that are used as inputs into FORLI (see Hurtmans et al., 2012). Hence, the apparent drift reported by Boynard et al. (2018) **likely** results from the jump rather than from a progressive "instrumental" drift.»

In these latest publication the words «**could be**», «**suspected**» and «**likely**» clearly mean that no formal evidence has been found to date.

Based on this review of FORLI litterature, we find that the possible explanation of the difference in calculated drifts for both algorithms with the EUMETSAT L2 products was not completely «**obvious**» for us at the time of writing our manuscript.

Nevertheless, we agree with Wespes et al. (2018) that a «jump» occuring in FORLI TROPO-O3 in September 2010 is responsible for most of the 8% «drift».
It is also noteworthy that the authors already discussed (i) the «jumpy» nature of the drift (ii) the role of this jump in the SOFRID FORLI difference concerning the NH tropospheric drift and (iii) even the potential link with EUMETSAT temperature at the end of the SOFRID-FORLI section a couple of lines after the statement cited by reviewer #2:
«These authors attribute their NH tropospheric significant drift to an abrupt change or jump detected in 2010 in FORLI [...] The difference could be linked to the use of EUMETSAT L2

products and of ECMWF analyses for FORLI and SOFRID retrievals respectively. As mentioned previously refering to B18, EUMETSAT L2 product are not homogeneous over the 2008-2016 period and a version change could result in the jump discussed in B18.»

In order to make things clearer we have modified this part taking the reviewer statements into account:

«Nevethelesss, the NH tropospheric drift from FORLI is attributed to an abrupt change or jump detected in 2010 (Boynard et al., 2018, Wespes et al. 2018). The drift strongly decreases after the jump and it becomes even non-significant for most of the stations over the periods before or after the jump, separately (Wespes et al. 2018). The discontinuity is suspected to result from updates in level-2 temperature data from EUMETSAT used as inputs into FORLI (Wespes et al., 2019). The absence of jump and the small drift in SOFRID v1.6 and v3.5 NH tropospheric data is therefore probably linked to the use of temperature profiles from ECMWF analyses instead of EUMETSAT L2 products.»

Finally, in the conclusion, we have modified the text to mention a jump rather than a drift in FORLI data:

« The difference with FORLI which is impacted by a significant TOC jump in 2010 (Boynard et al. 2018, Wespes et al. 2018) is likely linked to the use of different temperature profiles for the radiative transfer calculations (ECMWF analyses for SOFRID and EUMETSAT L2 for FORLI). »

- Section 4.3, p.12, L.6-7: It has also to be clearly noted that Gaudel et al. (2018) study suffers from a lack of consistency between a series of parameters, such as the calculation of the tropopause, making the comparison not quantitative.

In Gaudel et al. (2018) the tropopause is calculated according to the 2°K lapse rate from WMO for all of the satellite product. The different groups may have used different met-analyses (e.g. NCEP fo OMI-MLS, GOME and OMI, ECMWF for SOFRID, IASI-L2 for FORLI) but these resulted in rather little differences in tropopause height that cannot explain the significant differences in trends documented in this publication. In order to evidence the little impact of tropopause calculation in TOC trends and to compare our results with Boynard et al. (2018), we have computed the trends of the difference between sondes and SOFRID for both TOC and Surface-300 hPa columns. While the difference in column values is important (tropopause ~ 250-100 hPa versus 300 hPa), there are no significant changes (less than 0.2%) in the trends of the difference for the whole NH (see Fig 9 and 11 versus Fig. 16).

- Section 5, p.12, L.32-33: First of all, on the contrary to what is stated in Section 3.4, three indicators (not only two) were calculated in Boynard et al. (2016, 2018), the fourth one (ratio of std) being rarely calculated in validation studies.

We have corrected in the manuscript.

That last one that makes possible to draw Taylor diagram is indeed interesting as it allows evaluating the representation of the retrieved variability. It could indeed be investigated for the validation of future FORLI products. Nevertheless, I am surprised that the authors did not perform their own analysis using the FORLI dataset that is publicly available on the french Ether/Aeris platform. It would have prevented possible inconsistencies between the SOFRID and the IASI datasets, the validation methodologies. . .

As FORLI O3 data have been validated, we did not mean to re-validate them but to take advantage of the corresponding publications to check the consistency between datasets.

For instance, in:
- Section 5, p.13, L9-10: One source of difference between FORLI and SOFRID could be the series of quality flags that have been applied on the datasets to select the best observations in terms of spectral fit and cloudy scenes. Are the flags comparable between the FORLI and the SOFRID datasets? Please comment.

For SOFRID we filter the data with 3 quality flags. The first one described in the retrieval section (section 2) concerns clouds: we exclude cloud scenes based on the AVHRR cloud fraction cover. We now give the threshold which was missing:

« Pixels with AVHRR-derived fractional cloud cover larger than 25% are excluded»

Then we have three more data filters described in the validation section (Section 3): two based on the quality of retrieval (cost function Jcost > 0.0 for correct convergence and Jcost < 1.0 to elliminate worst fits), one concerning the information content (DFS > 2.0).

In order to show that the threshold values are not impacting substantially the SOFRID-FORLI comparisons we have performed sensitivity tests with different values. For Jcost we have used a trheshold value of 0.15 instead of 1.0 and the number of selected pixels decreased by 6%. Concerning the cloud filter, we have performed a test with 13% which is the value used in Boynard et al. (2018) instead of 25%. The number of pixels decreased by 5%. For the DFS we have made the comparisons with a trheshold of 1.75 instead of 2.00 (which is the trheshold used in Boynard et al. (2018)) and the number of pixels increased by 2%.
In each case, the general statistics changes are negligibles as can be seen in Figure 1 where the biases and RMSDs for v3.5 with the standard quality flags and with the modified trhesholds are presented.

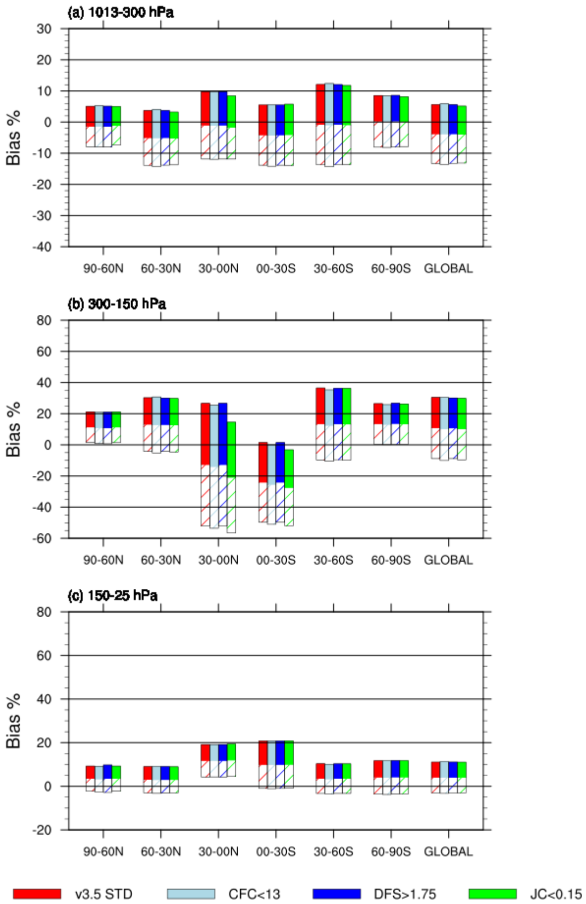

Figure 1 : Biases and RMSDs of the differences between IASI retrievals and sonde data for the standard v3.5 (red), and v3.5 with modified quality flags: AVHRR cloud fraction cover (light blue), DFS (blue) and Jcost (green) (similar to Figure 14 in the paper).

The same is true for the correlation coefficients (r2) and slopes of the linear regressions (b) as shown in Fig. 2.

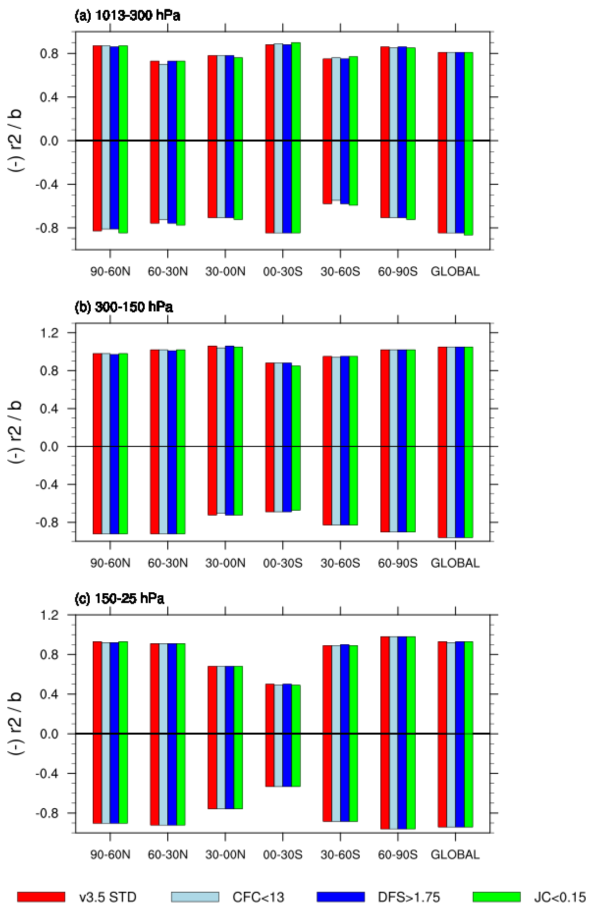

Figure 2. Slopes of the linear regression (positive values) and (-) r2 correlation coefficients (negative values) between IASI retrievals and sonde data (same as Fig. 13 in the paper).

The comparison of the two IASI-O3 retrievals presented in the paper is therefore robust and not highly dependent on the thresholds used to filter the data.

We have modified the text in the section corresponding to the SOFRID-FORLI comparisons accordingly adding the following text:

«Another limitation is that FORLI and SOFRID use their own quality flags to filter the data. In order to document the impact of the pixel selection on SOFRID validation we have performed the comparison with sonde data using modified quality flags. The cloud filtering trheshold is the clearest source of difference between the pixel selection of both algorithms. We have therefore lowered the upper limit of the AVHRR cloud fraction cover to 13% which is the trheshold used by Boynard et al. (2016, 2018) resulting in a loss of 5% of the treated pixels. The Jcost threshold has been decreased from 1.0 to 0.15 with a 6% decrease of the selected retrieved profiles. Finally the DFS lower value has been set to 1.75 increasing the number of selected retrievals by 2%. These threshold modifications resulted in negligible changes of the general statistics (bias, RMSD, R) for the 3 atmospheric layers (troposphere, UTLS and stratosphere) and the different latitude bands that are presented in this section. These statistics, based on large numbers of data are therefore not hindered by pixel selection differences.»

This is why taking data directly from literature for a quantitative comparison might be inappropriate and mislead the comparison. That issue/limitation in the comparison between SOFRID and FORLI should be clearly highlighted and discussed by the authors. I would strongly recommend the authors to better put the FORLI-SOFRID comparison into context with the reasons mentioned here above (i.e. jump in contrast with real drift, use of different quality flags, possible inconsistency between validation methodology. . .) through the manuscript.

We agree with the reviewer that the SOFRID-FORLI comparison has important limitations because it is based on published results. Some limitations were already highlighted in the manuscript such as the fact that we could only compare results after smoothing of the sonde data. The jump issue has been largely discussed and amended throughout the manuscript as described above. The issue concerning data filtering is also now largely discussed in the manuscript with evidence given by sensitivity tests on the « quality flags ». Nevertheless, it is important to realize that the quality flag issue is inherent to retrieval algorithm comparisons even with dedicated studies where the data are not taken from the literrature.

- P.6, L.6-7: Why the behavior of TOC errors is similar to that of DFS while one can read above that the dominant source is the smoothing error? Please explain.

The a priori variability is larger for the TOC in the tropics that at mid and high latitudes because of the higher tropopause height resulting in larger smoothing errors even with higher DFS. We have added the following explanation :

« This is due to the fact that the tropopause height is higher in the tropics resulting in a larger a priori variability. The impact of the increased variability exceeds the one of the increased information content resulting in a larger smoothing error »

- P.10, L.2-3: Why does the smoothing of sonde profiles not improve the bias in UTLS while the DFS is < 1? Please explain.

The application of Equ. 1 to the sonde profiles is supposed to correct biases linked to the a priori profiles independently from the DFS value. The fact that (i) the bias is present for v1.6 and v3.5 for which the a priori are different (ii) the application of Equ. 1 does not change significantly the bias, indicate that this particular bias (unlike the TOC biases) is not related to the a priori.

Therefore, this UTLS bias in IASI O_3, already identified for the three retrieval algorithms (with and without smoothing) by Dufour et al. (2012) remains an issue.

- Regarding the figures 12-14, one could think that the authors make their own analysis from the FORLI datasets, while the values are taken from previous validation papers. This should be clearly mentioned in the figure captions to avoid misunderstandings.

We have added the ref to Boynard et al. (2018) in the captions.

Technical comments and typos:

- P.2, l.22: The jump is detected in year 2010, not 2011.

OK

- P.2, L.30: tropospheric -> tropospheric

OK

- *P.3, L.7: methodology -> methodology*
OK
- *P.4, L.33: “The use OF a . . .”*
OK
- *P.5, L.8: atmospheric -> atmospheric*
OK
- *P.6, L.1: Th -> The*
OK
- *P.7, L.20: one reference is missing here.*
We have added the ref Havemann (2020) for the convergence criteria of the NWP-SAF 1D-Var algorithm.
- *P.7, L.9: below -> above*
balloons with O3 sondes often explode below 40 kms.
- *P.7, L.21: elliminate -> eliminate*
OK
- *P.8, L.23: variance -> ratio of the variance (?)*
« ... is proportional to the variance of the experiment. Both RMSDs and standard deviations are normalised by the standard deviation of the reference... »
The second sentence implied that, after normalisation « the radial distance from the origin » is proportional to the ratio of the variance of the experiment to the variance of the reference.
We have changed the sentence to be clearer :
« We have normalised both RMSDs and standard deviations by the standard deviation of the reference to display the results from multiple experiments on a single diagram (see Taylor (2001) for details). »
- *Table 2: Units are missing*
We have added the units (%).
- *P.9, L.21: tropospehric -> tropospheric*
OK
- *P.9, L.27: UT -> UTLS*
OK
- *Fig.6 and 7: The legend is not clear. I guess RS means Raw Sondes and SmRS means Smoothed Sondes. Hence, SmRS should be SmS (?). Please correct or clarify in the caption.*
The caption has been clarified.
- *Error in the caption of Fig.9: “Same as Fig.9” -> “Same as Fig.8”*
OK
- *Fig.8: The color legend should be indicated in the top panels.*
The line colors are documented in the captions in order to avoid problems of legend superimposed on the lines.
- *P.12, L.6: Which version of SOFRID are you referring to?*
SOFRID v1.5 was used in Gaudel et al. (2018). We have added the version.
- *Fig.12 to 14 do not seem in correct order. Please consider this: Fig.14 -> Fig.12, Fig.12 -> Fig.13, Fig.13 -> Fig.14*
We have reordered the citation to the figures in the text.
- *P.13, L.1: delete “(b)” in the sentence. I don’t see that in Fig.13.*
We have modified the caption adding the ref to « b ».

A tropopause-based a priori for IASI-SOFRID Ozone retrievals: improvements and validation

Brice Barret[1], Emanuele Emili[2], and Eric Le Flochmoen[1]

[1]Laboratoire d'Aérologie/OMP, Université de Toulouse, Toulouse, France.
[2]CECI, Université de Toulouse, CERFACS, CNRS, Toulouse, France.

Correspondence: B. Barret
(brice.barret@aero.obs-mip.fr)

Abstract.

The Metop/IASI instruments provide data for operational meteorology and document atmospheric composition since 2007. IASI Ozone (O_3) data have been used extensively to characterize the seasonal and interannual variabilities and the evolution of tropospheric O_3 at the global scale. The SOFRID (SOftware for a Fast Retrieval of IASI Data) is a fast retrieval algorithm that provides IASI O_3 profiles for the whole IASI period. Up to now SOFRID O_3 retrievals (v1.5 and 1.6) were performed with a single a priori profile which resulted in important biases and probably a too low variability. For the first time we have implemented a **comprehensive** dynamical a priori profile for spaceborne O_3 retrievals which takes the pixel location, time and tropopause height into account for SOFRID-O3 v3.5 retrievals. In the present study we validate SOFRID-O3 v1.6 and v3.5 with ECC ozonesonde profiles from the global WOUDC database for the 2008-2017 period. Our validation is based on a thorough statistical analysis using Taylor diagrams. Furthermore we compare our retrievals with ozonesonde profiles both smoothed by the IASI averaging kernels and raw. This methodology is essential to evaluate the inherent usefulness of the retrievals to assess O_3 variability and trends. The use of a dynamical a priori largely improves the retrievals concerning two main aspects: (i) it corrects high biases for low-tropospheric O_3 regions such as the southern hemisphere (ii) it increases the retrieved O_3 variability leading to a better agreement with ozonesonde data. Concerning UTLS and stratospheric O_3 the improvements are less important and the biases are very similar for both versions. The SOFRID Tropospheric Ozone Columns (TOC) display no significant drifts ($< 2.5\%$) for the northern hemisphere and significant negative ones (9.5% for v1.6 and 4.3% for v3.5) for the southern hemisphere . We have compared our validation results to those of the FORLI retrieval software from the litterature for smoothed ozonesonde data only. This comparison highlights three main differences: (i) FORLI retrievals contain more theoretical information about tropospheric O_3 than SOFRID (ii) RMSDs are smaller and correlation coefficients are higher for SOFRID than for FORLI (iii) in the northern hemisphere, **the 2010 jump detected in FORLI TOCs is not present in SOFRID**.

1 Introduction

Ozone (O_3) in the stratosphere protects life from solar UV radiation. Close to the surface, O_3 is an oxidative pollutant harmfull for human health through irritation of respiratory tract (Brunekreef and Holgate, 2002) and for vegetation through deposition on leafs that leads to the reduction of plant growth (Ainsworth et al., 2012). Tropospheric O_3 is also a powerful greenhouse gaz which increase during the 20-th century has significantly contributed to global warming (Shindell et al., 2006). The radiative forcing of O_3 is particularly important in the tropical Upper Troposphere-Lower Stratosphere (UTLS) (Chen et al., 2007).

It is therefore important to document the evolution of O_3 in these different layers independently. There are clear evidences from satellite databases that upper stratospheric O_3 has increased since 1997 following the ban of CFC's by the Montreal protocol (Ball et al., 2018). Nevertheless, the total column O_3 is stable since 1998. According to Ball et al. (2018) this contradiction is due to the fact that lower stratospheric O_3 is declining and compensates both stratospheric and tropospheric O_3 increase. Based on OMI/MLS Tropospheric Ozone Columns (TOC) they state that TOC is globally increasing. OMI/MLS data for the 2005-2016 period are indeed documenting global positive TOC trends with particularly large increases over Asia (Ziemke et al., 2019). Based on 10 years of retrievals with the Fast Optimal Retrievals on Layers for IASI O_3 (FORLI-O3) software, Wespes et al. (2018) document a decrease in tropospheric O_3 levels in the Northern Hemisphere (NH). Another IASI tropospheric O_3 product (KOPRAFIT-O3) displays a TOC decrease over continental China (Dufour et al., 2018). In their exhaustive work about TOC evolution, Gaudel et al. (2018) clearly highlight the contradiction between global increase (OMI/MLS and other UV-Vis products) on the one hand and global decrease (IASI) on the other hand. They also show that the different satellite products agree on a TOC increase over Asia. Among the two global IASI TOC datasets used in Gaudel et al. (2018), FORLI-O3 is indicating a significant global decrease and O_3 retrievals with the Software for a Fast Retrievals of IASI Data (SOFRID) a slightly weaker and less significant one. Two versions of FORLI-O3 have been validated by Boynard et al. (2016) (v20141022) and Boynard et al. (2018) (v20151001). They both document a **jump** in the O_3 retrievals in **2010** but this does not hinder the fact that TOC are decreasing according to Wespes et al. (2017). It has to be noted that both validation studies compare IASI retrievals to ozonesonde profiles smoothed by the retrieval averaging kernels. Such a comparison enables the detection of abnormal biases, variability or drifts in the retrievals but do not document the ability of FORLI-O3 to reproduce real O_3 levels and variabilities. SOFRID-O3 has only been validated at the beginning of the IASI period on a very short time period (Barret et al., 2011) and on a longer time period together with FORLI-O3 and KOPRAFIT-O3 (Dufour et al., 2012). Furthermore, EUMETSAT L2 atmospheric temperature products retrieved from IASI and used for FORLI (v20141022 and v20151001) and for SOFRID-O3 v1.5 retrievals are not stable in time (Boynard et al., 2018). Therefore we have reprocessed the whole IASI database using ECMWF operational analyses for temperature and humidity to produce SOFRID-O3 v1.6. SOFRID-O3 has been shown to overestimate low tropospheric ozone over the southern hemisphere (SH) (Dufour et al., 2012; Emili et al., 2014, 2019). Emili et al. (2014) have hypothetized that this overestimation was due to the use of a single a priori profile biased towards northern hemisphere (NH) mid-latitudes O_3. In order to verify this hypothesis and to improve our O_3 retrievals, we have developed a new version of SOFRID-O3 (v3.5), with a dynamical a priori profile based on a global O_3 climatology (Sofieva et al., 2014).

The aim of the present paper is to validate both SOFRID-O3 latest products (v1.6 and v3.5) on the whole IASI period (2008-2017) in order to infer their ability to reproduce tropospheric O_3 levels and variability on seasonal to decadal time scales. The validation is based on O_3 profiles from ozonesonde retrieved from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) database. In section 2 we describe the characteristics and differences of SOFRID-O3 v1.6 and v3.5 retrievals. Section 3 is dedicated to the description of the validation methodology based on comparisons between smoothed and raw ozonesonde data and we provide our validation results in section 4. Based on Boynard et al. (2018) we also compare our results to FORLI-O3 (section 5) before concluding in section 6.

2 IASI SOFRID-O3 Retrievals

IASI is a spaceborne thermal infrared nadir spectrometer. IASI has a moderate spectral resolution combined with a high signal to noise ratio and a 12 km footprint at nadir (Clerbaux et al., 2009). Thanks to its large accross track scanning ($\sim$ 2200 km), IASI revisits each scene twice daily around 9:30 solar time in the morning and in the evening. Three IASI instruments have been launched on the Metop meteorological platforms (Metop-A in 2008, Metop-B in 2012 and Metop-C in 2018). Here we present results based on O_3 retrievals from 10 years of Metop-A/IASI data. We will present results based on the morning overpass data only as they are known to provide more information that nighttime data. Furthermore, it facilitates the comparison to other validation studies (Boynard et al., 2018) also based on morning data.

The SOFRID software first described in Barret et al. (2011) is based on the RTTOV (Radiative Transfer for TOVS) operational radiative transfert code (Saunders et al., 1999; Matricardi et al., 2004) combined with the 1D-Var software (Pavelin et al., 2008) both developed within the framework of EUMETSAT NWP-SAF. The O_3 profiles are retrieved from the 980-1100 cm^{-1} spectral window encompassing the 9.6 μm O_3 absorption band. Only cloud free or weakly contaminated pixels are processed. **Pixels with AVHRR-derived fractional cloud cover larger than 25% are excluded**. We also use a test based on brightness temperatures at 11μm and 12μm when AVHRR cloud cover is not available as described in Barret et al. (2011). The two SOFRID-O3 versions that are validated and compared in the present paper have significant differences that are described below.

2.1 Single a priori profile: V1.6

SOFRID-O3 V1.6 is almost similar to V1.5 described in Barret et al. (2011). It is based on RTTOV V9.3 (Saunders et al., 1999). In RTTOV, the optical depths are expressed as a linear combination of profile dependent predictors that are functions of temperature, absorber amount, pressure and viewing angle. In RTTOV V9.3, the regression coefficients are derived from computations with the LBL Radiative Transfer Model V11.6 (LBLRTM Clough et al. (2005)) on 43 atmospheric levels using the HITRAN2004 spectrocopic database (Rothman and Jacquemart, 2005). The single difference is that V1.6 uses temperature and humidity profiles from ECMWF operational analyses for the RTTOV simulations and V1.5 was using IASI L2 products delivered by EUMETSAT. The change has been operated for availability problems and mostly because the EUMETSAT L2 products are not homogeneous over the whole 2008-2017 period which could result in retrieval inconsistencies (Boynard et al., 2018). We use 6 hourly ECMWF analyses which are provided on 91 (resp. 137) vertical levels until (resp. after) 24 June 2013 from the ground up to 0.02 hPa on a 0.25°x0.25° horizontal grid. The ECMWF temperature and humidity profiles are interpolated to the time and location of the target IASI pixel with a 3D-linear interpolation scheme.

O_3 concentrations are retrieved on the 43 RTTOV levels with the NWPSAF 1D-Var algorithm (Pavelin et al., 2008) based on the Optimal Estimation Method (Rodgers, 2000). The OEM is a Bayesian method where the incomplete information provided by the measurement is complemented by a priori information which is supposed to represent the best knowledge of the state vector at the moment of the measurement. In our case the state vector is the O_3 profile. For both V1.5 and V1.6 we use a single O_3 a priori profile which is based on two years (2008-2009) of WOUDC and MOZAIC-IAGOS profiles completed to the top of the RTTOV V9.3 model (0.1 hPa) by MLS averaged profiles (see Barret et al. (2011) for details).

2.2 Dynamical a priori profile: V3.5

As V1.6, SOFRID-O3 V3.5 uses interpolated temperature and humidity profiles from ECMWF analyses. It is based on the more recent RTTOV V11.1 (Hocking et al., 2015) which regression coefficients are derived from LBLRTM V12.2 computations on 101 vertical levels with the HITRAN2008 spectroscopic database (Rothman et al., 2009). The second and more important one is that it uses dynamical a priori profiles from TpO_3, the O_3 profile tropopause based climatology of Sofieva et al. (2014). This climatology is based on ozone profiles resulting from merging ozonesonde data in the troposphere and SAGE-II V6.2 data (Wang et al., 2006) in the stratosphere. The ozonesonde profiles (36000) extracted from the Binary Data Base of Profiles (BDBP) come from 136 stations for the period 1980 to 2006 (Hassler et al., 2008). For each merged ozonesonde-SAGE-II profile, the tropopause was computed according to the World Meteorological Organization (WMO) definition of the lapse-rate tropopause (WMO, 1957). For each month, the ozone profiles are gathered according to 10° latitude bins, 1 km tropopause intervals and the corresponding averaged profiles together with their 1σ variabilities are computed and provided. Variable a priori profiles have already been used for satellite sensor retrievals. For instance, TES O_3 retrievals used monthly mean profiles from the MOZART CTM averaged over a 10°latitude x60° longitude grid (Bowman et al., 2006). OMI O_3 a priori profiles are based on a monthly and latitude dependent ozone profile climatology (McPeters et al., 2007) derived from ozonesonde and satellite data (Liu et al., 2010). Nevertheless, the use of an a priori simply based on the geographical location of the satellite pixel does not allow taking the atmospheric dynamics into account. For instance, at a mid-latitude location, the O_3 profile can be typical of mid-latitudes one day and polar (low tropopause) or tropical (high tropopause) a few days later depending on the global atmospheric dynamics (position of the polar or subtropical jets, anticyclones). The use of a tropopause dependent climatology allows us to take the atmospheric dynamics into account and provides a more accurate a priori O_3 profile. This technique was once used for O_3 total column retrievals from FTIR spectra at the Jungfraujoch station (De Maziere et al., 1999). It was shown that the retrieved O_3 columns were largely improved when the tropopause was taken into account in the choice of the a priori. **In a first attempt to take the tropopause into account for satellite retrievals, Sellitto et al. (2013) have implemented 2 a priori profiles in the KOPRAFIT-O3 retrieval algorithm to basically discriminate the tropics (tropopause higher than 14 km) from other latitudes. Dufour et al. (2015) have slightly improved the approach with a set of 3 a priori profiles for high latitudes (tropopause lower than 10km), mid-latitudes (tropopause between 10 and 14 km) and the tropics (tropopause higher than 14 km). Eremenko et al. (2019) have tested a set of N profiles for retrievals on a synthetic database.** In SOFRID-O3 V3.5, we compute the tropopause using the WMO lapse rate definition from the ECMWF interpolated temperature profiles. The a priori profile is then picked up from the TpO_3 climatology according to month, latitude and tropopause height.

2.3 Information content and retrieval error

A remote sensing instrument is not equally sensitive to the different atmopsheric layers. Its vertical sensitivity depends on its instrumental characteristics and on local parameters. In the case of a thermal infrared nadir sounder such as IASI, surface parameters such as surface emissivity, surface temperature, thermal contrast between the surface and the first atmospheric layer are key parameters to determine the vertical sensitivity, especially in the lower troposphere (Barret et al., 2005; Boynard et al., 2016). The vertical sensitivity of a remote sensing instrument is characterised by the so-called Averaging Kernel (AK) matrix. For each retrieval layer, the retrieved quantity is the result of the convolution of the whole real profile by the corresponding averaging kernel (row of the AK matrix) plus a contribution from the a priori profile ($\mathbf{x}_a$) and a noise (ϵ) contribution (see Eq. 1).

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{G}(\boldsymbol{\epsilon}) \quad (1)$$

In an ideal case, the AK matrix ($\mathbf{A}$) would be the identity matrix ($\mathbf{I}$) and real ($\mathbf{x}$) and retrieved ($\hat{\mathbf{x}}$) profiles would be identical modulo the noise (ϵ) contribution. **$\mathbf{G}$ is the gain matrix that represents the sensitivity of the retrieval to the measurement**. In a real case, the AKs are bell shaped functions which peak at an altitude that could be different from the nominal altitude and which width gives an indication of the retrieval vertical resolution.

The Degree of Freedom for Signal (DFS) of a retrieval describing the number of independent pieces of information provided by the measurement is the trace of the AK matrix (Rodgers, 2000). We have divided the atmosphere in 5 layers which are described in Table 1. The Troposphere 2 layer has been selected for comparison with Boynard et al. (2018) who did not compute a tropopause based TOC for their validation (see section 5). The DFS corresponding to these different layers is displayed in Figure 1 for V1.6 and V3.5 averaged over the validation dataset. The total DFS ranges from 2.4 to 3.3 for v3.5 and is about 0.2 lower for v1.6. The DFS for the troposphere (WMO lapse rate), UTLS and stratosphere are almost identicals for both versions. The tropospheric DFS is the lowest (0.3-0.5) at high latitudes where surface temperature, thermal contrast and tropopause height are the lowest and the highest in the tropics (about 1.5) where surface temperature and tropopause height are the highest. At mid-latitudes the tropospheric DFS is about 0.6. Therefore, except in the tropics, SOFRID retrievals provide less than one independent piece of information in the troposphere. In the UTLS (resp. stratosphere) the DFS range from 0.7 to 1 (resp. from 0.9 to 1.5) which means that SOFRID provides around one independent piece of information in these layers.

The retrieval error is the sum of the measurement and smoothing errors (Rodgers, 2000). Uncertainties in auxiliary parameters (Temperature and humidity profiles, surface properties...) are also responsible for errors. Coheur et al. (2005) and Barret et al. (2005) have shown that in the case of O_3 and CO retrievals from thermal infrared satellite sensors the dominant source of errors was the smoothing error. The retrieval error for SOFRID-O3 v1.6 and v3.5 are displayed in Fig. 1. V1.6 displays slightly larger errors than v3.5 but the same behaviour. For the Total and stratospheric columns, the errors decrease from high latitudes (9-12 DU) to the tropics (6-8 DU). The behaviour of UTLS errors is similar with lower values (4 to 6 DU). For the TOC, errors are larger in the tropics (5 DU) than at middle and high latitudes (4 DU). **This is due to the fact that the tropopause height is higher in the tropics resulting in a larger a priori variability. The impact of the increased variability exceeds the one of the increased information content resulting in a larger smoothing error.**

2.4 Global distributions of tropospheric ozone columns

The global distributions of TOC from SOFRID v1.6 and v3.5 for July and December 2017 are displayed on Fig. 2. The global TOC structures are similar for both versions. They both clearly show the highest TOC over the NH mid-latitudes in summer with a large export region over the north Pacific off the Chinese coast and the summertime TOC maximum over the Eastern Mediterranean already documented with the GOME-2 sensor (Richards et al., 2013). The tropical Wave-one pattern (Thompson et al., 2003; Sauvage et al., 2006) with the highest TOC over the tropical Atlantic and the lowest one over the South Pacific Convergence Zone (SPCZ) is also noticeable for both versions. Sauvage et al. (2006) have shown that the tropical Atlantic maximum was mostly a result of African and South American Lighning NOx (LiNOx) emissions. High TOC are also detected during austral summer over southern Africa and the southern Indian Ocean towards Australia. According to Zhang et al. (2012), these high TOC are mostly caused by LiNOx emissions from central Africa with a yearly maximum in May. The clearest difference between both versions is that v3.5 produces lower TOC than v1.6 in the low tropospheric O_3 regions. This is clear over the Inter Tropical Convergence Zone (ITCZ) and the SPCZ, over the SH for both seasons and over the NH mid latitudes in winter. We will show in the validation part of the paper that this is an important improvement of the SOFRID-O3 retrievals. The agreement is better in regions of high TOC such as NH mid latitudes in summer or the tropical Atlantic.

The use of a dynamical a priori is responsible for visible stripes along the 10° latitude bands. These stripes are generally indicating a discontinuity of 2.5 to 5 DU between two adjacent latitude bands with different a priori profiles. They are clearly caused by the impact of the a priori on the retrieval which is taken into account in the retrieval error (see Equ. 1). The latitudinal discontinuities are therefore consistent with our retrieval errors (4-5 DU) from Fig. 1. **Such stripes may appear as a problem for the use of SOFRID v3.5 data for model validation. They are a minor problem for two main reasons. First, as is demonstrated in section 4, the use of a dynamical a priori largely improves the retrieved O_3 profiles. Second, when model profiles are compared to SOFRID retrievals the impact of the a priori profile is taken into account by using Eq. 1 such as in Barret et al. (2016).**

3 Validation methodology

3.1 Ozonesonde data

Ozonesonde data come from the WOUDC database (hhttps://www.woudc.org/). For consistency purposes we have chosen to use data from ECC sondes only. For the 10 years IASI period (2008-2017), valid comparisons were effective for about 12000 ozonesonde profiles among the 16000 downloaded. A map with the number of sondes used for the validation at each station over the 2008-2017 period is displayed in Fig. 3. Most (∼7000) of the validation sondes were launched in the NH mid-latitudes with 15 stations providing more than 1 profile per month on average (more than 120 profiles for 10 years) mostly in Western Europe and Northern America. For all other 30° latitude bands, the number of validation profiles range from 800 to 1200 with only 3 to 4 stations providing more than 120 profiles. The balloons that carry the ozonesondes **often** explode below 40 km. In order to complete the ozonesonde profiles in the upper stratosphere and mesosphere, we have used MLS data averaged on 10 days on a 10°x10° grid (see Barret et al. (2011) for details).

3.2 Coincidence criteria

The spatiotemporal coincidence criteria are ±1°latitude, ±1°longitude and ±12 hours. They are similar to those used in Barret et al. (2011), Boynard et al. (2016) (50 km ±10h), Boynard et al. (2018) (100 km, ±6h), Dufour et al. (2012) (110 km, ±7h). As we compare sondes with IASI morning data only and that most of the sonde launches are performed in the morning, using 6 or 12h coincidence does not introduce significant differences. We have computed statistics for 9 latitude bands which are the whole globe, the two hemispheres and six 30° wide latitude bands. For each band, the monthly mean is computed if there are at least **4 coincident profiles within this latitude band**. Pixels are selected according to 3 quality criteria. We first keep pixels for which convergence is achieved meaning a positive retrieval cost function (Jcost) output from the 1DVar **based on the value of Jcost, of its normalised gradient and on the evolution of Jcost between the two last iterations (Havemann , 2020)**. We have also set an upper limit (1.0) for Jcost in order to eliminate pixels with poor quality fits. Thirdly, only pixels with a total DFS > 2.0 are selected. Using these criteria we have kept about 9.0E5 pixels out of 1.1E6.

3.3 Comparison with raw and smoothed data

To compare remote sensed to in-situ or modeled profiles it is important to apply Eq. 1 to the in-situ or simulated profile (Rodgers, 2000; Barret et al., 2002). This procedure allows us to check the quality of the retrieval taking its degraded vertical resolution and sensitivity into account.

Nevertheless, in a validation objective it is also necessary to compare the retrieved profiles to raw (not smoothed by the AKs) in-situ profiles in order to perform a fully informative validation. This is of particular importance when the satellite data are used for issues such as the ozone seasonal to interannual variabilities (Wespes et al., 2017; Peiro et al., 2018) or to document the long term tropospheric ozone tendencies (Gaudel et al., 2018; Wespes et al., 2018; Dufour et al., 2018). Indeed, the application of Equation 1 implies the mixing of information between the different layers. Therefore, the variabilities and the drifts computed from raw and smoothed sonde data may be different and need to be documented. Raw ozone sonde data have been compared to IASI retrievals in few studies at the beginning of the IASI period (Barret et al., 2011; Dufour et al., 2012) but have been disregarded in more recent validation work (Boynard et al., 2016, 2018). The importance of raw data validation regarding seasonal and interannual variabilities and trends analyses will be highlighted in details in section 4.

3.4 Taylor diagram

In order to validate remote sensing with reference in-situ observations we need to determine how well they are able to reproduce the same behaviour. There are four statistical indicators that have to be computed: (i) the absolute difference or bias which documents the accuracy, (ii) the root mean square of the differences (RMSD) which tells wether the bias is significant or not, (iii) the correlation coefficient (R) that document the consistency and phase of the variabilities of both datasets and (iv) the ratio of the standard deviations of both datasets which documents the goodness of the amplitude of the retrieval variability. In the case of IASI O_3, the first **three** indicators are frequently computed (Boynard et al., 2016, 2018; Barret et al., 2011) but the last one is rarely compared (Dufour et al., 2012) which makes most validation exercices incomplete.

Based on the relationship between correlation coefficients, RMSDs and variances of the reference (validating) and test (validated) datasets, Taylor has developed the Taylor diagram initially for climate models evaluation (Taylor, 2001). It displays all of these parameters (except the biases) in a more convenient and synthetic way than tables with numbers. Each experiment or observation to be validated correspond to a point placed within a quarter circle. The reference is located in the middle of the X-axis (see Fig. 4, 5). The correlation coefficient between the reference and test dataset is given by the azimuthal position of the point. The RMSD is proportional to the distance between the test and the reference point. Finally, the radial distance from the origin is proportional to the variance of the experiment. We have normalised both RMSDs and standard deviations by the standard deviation of the reference to display the results from multiple experiments on a single diagram (see Taylor (2001) for details).

4 Validation results

4.1 General statistics for tropospheric, UTLS and stratospheric partial columns

For the different latitude bands, the statistics from the comparisons between ozonesondes and SOFRID data are presented in Table 2 for the biases and corresponding RMSDs. Taylor diagrams are displayed in Fig. 4 for the TOC and lower tropospheric columns and in Fig. 5 for the UTLS and stratospheric columns.

Concerning the troposphere, the comparison between SOFRID and raw sonde clearly shows the improvement from v1.6 to v3.5 (Fig. 4(a)). The v3.5 displays a larger variability in better agreement with the raw sonde data with a ratio between SOFRID and sonde variances ranging from 0.62 to 1.01. For v1.6 this ratio ranges from 0.15 to 0.45. The RMSDs of the SOFRID versus raw sonde data are lower and the correlation coefficients larger for V3.5 than for V1.6. Tropospheric biases are smaller than 10% with the noticeable exception of mid and high latitudes of the SH for v1.6 and raw sonde data with significant biases of 29 and 55% respectively (Table 2). This problem of SOFRID v1.6 retrievals in the SH had already been diagnosed by Dufour et al. (2012) and by Emili et al. (2014). The use of a dynamical a priori in v3.5 allows us to reduce these large biases to almost zero.

As expected, when the sonde profiles are smoothed with SOFRID AKs (Figure 4(b) and (d)) the agreement between sonde data and SOFRID retrievals is better. The retrieval variabilities are closer to the sonde variabilities, the RMSDs are smaller and the correlation coefficients are higher. It is also noticeable that differences between both retrieval versions is less important and that the improvement of v3.5 relative to v1.6 is less evident. Furthermore, the large v1.6 biases in the SH troposphere at mid and high latitudes is reduced below 10 % when the impact of the a priori is taken into account with Equ. 1, hiding the problem.

The lower tropospheric retrieved columns agree less with raw sonde data with degraded correlation coefficients and larger RMSDs (Figure 4(c)) compared to the TOCs. For raw sonde data comparisons, the lower tropospheric variability is better for V3.5 than for V1.6. When the sondes are smoothed, the statistics are much better and similar to the TOC results (Figure 4(d)). The added value of lower tropospheric columns relative to TOCs is therefore not obvious for SOFRID-O3.

In the UTLS, both v1.6 and v3.5 are in good agreement with raw sonde data (Figure 5(a)) and the differences between both versions are much lower than for the tropospheric columns. Correlation coefficients range from 0.67 to 0.93 and the ratios between retrieved and raw sonde variances range from 0.5 to 1.0 at mid and high latitudes. For the northern and southern tropical latitudes the correlation coefficients range from 0.6 to 0.75 and the variance ratios are between 1.6 and 2.1 highlighting a too high variability retrieved in the tropical UTLS. In the UTLS, biases are positive (5 to 18%) at high and mid latitudes and negative (-3 to -21%) at tropical latitudes and are not significant because of large RMSDs.

In the stratosphere, the agreement between raw sonde data and SOFRID retrievals is very good for the 2 versions and in all latitude bands with correlation coefficients in the 0.75-0.98 range and variance ratios in the 0.56-0.96 range except in the tropical bands where the retrieved variances are much lower than the ozonesonde variances (Figure 4(c)). Stratospheric columns from v3.5 are in slightly better agreement (higher r2, lower RMSDs) with ozonesonde data than v1.6. Large positive biases (10-14 %) are found at tropical latitudes for both v1.6 and v3.5 (Table 2).

Both in the UTLS and the stratosphere, the agreement is only slightly improved (larger correlation coefficients and lower RMSDs) when the sonde profiles are smoothed by the AKs (Figure 5(b) and (d)). Smoothing of the sonde profiles do not significantly modify the UTLS and stratospheric biases. In particular, the tropical UTLS large negative biases are still present when the AKs are applied to the sonde data. The small differences between v1.6 and v3.5 on the one hand and between raw and smoothed sonde data on the other hand highlight the larger sensitivity of IASI to the UTLS and the stratosphere than to the troposphere as already discussed in Barret et al. (2011) and Dufour et al. (2012) for SOFRID v1.5.

4.2 Impact of the intra-seasonal tropopause dependence of the a priori profile on SOFRID improvements

The Sofieva et al. (2014) climatology is tropopause dependent in two different ways. First, it implicitly documents the seasonal and latitudinal relationship between the tropopause and the O_3 profiles with the classification of the O_3 profiles by month and 10° latitude band. Second it documents explicitly the intra- seasonal tropopause-O_3 profiles relationship with various tropopause dependent profiles for each month and latitude band. In order to determine the impact of the intra-seasonal tropopause dependence on the SOFRID retrievals we have performed retrievals using the profile with the highest occurence for each month and each 10° latitude band as a priori. This version with a single monthly a priori profile in eache 10° latitude band is v3.3.

The comparisons between v3.5 and v3.3 are presented in Taylor diagrams for the tropospheric and the UTLS columns in Fig. 6. In the stratosphere the changes (not shown) are negligible. Concerning the TOC, the improvements are negligible except in the 60-30°N and 60-90°S bands where v3.5 better reproduces the variability of the TOC. In the UTLS v3.5 gives slightly better results in terms of correlation coefficients and variability relative to v3.3 in all the latitude bands except in the 60-90°S one. Nevertheless, the improvements from v3.3 to v3.5 is minor compared to the overall improvement from v1.6 to v3.5 (Fig. 4 and 5). We can therefore conclude that the seasonal (monthly) and latitude dependence of the O_3 profile a priori are responsible for most of SOFRID improvements.

4.3 Vertical profiles

After comparing partial columns, it is interesting to look at complete profiles to get a better insight about the discrepancies between IASI retrievals and sonde data. The annual average profiles for V1.6 and V3.5 are displayed in Fig. 7 and Fig. 8 resp. for the different latitude bands.

In the NH, v1.6 and v3.5 show similar behaviours with a large upper tropospheric positive bias at mid and high latitudes and a large oscillation from a negative bias at 250 hPa to a large positive bias at 100 hPa in the tropics. These profile features are responsible for the positive (resp. negative) biases for the mid and high latitudes (resp. tropics) UTLS columns and for the positive biases for the tropical stratospheric columns (see Table 2). In the SH the large tropospheric positive biases of SOFRID relative to raw sondes (below 300 hPa in the high and mid-latitudes and below 500 hPa in the tropics) present in v1.6 almost dissapear in v3.5. The improvement of SOFRID accuracy in the SH extra tropical troposphere is the clearest advantage of using a dynamical a priori profiles. In the SH tropics, the TOC difference between v1.6 and 3.5 is not so clear (see Table 2) because the positive bias in the lower troposphere is compensated by a larger negative bias in the upper troposphere in v1.6. As already discussed from column comparisons, it is also noticeable from profile comparisons (Fig. 7 and 8) that the agreement between SOFRID retrievals and smoothed sonde profiles is better than with raw sondes. An important exception is the large UTLS oscillations in both the NH and SH tropics and for both v1.6 and 3.5. Therefore, unlike expected, this important discrepancy between retrievals and sonde data does not result from the use of a single a priori profile too far from the real profile. The differences between v3.5 and v1.6 are largely reduced when sondes are smoothed. For instance the large tropospheric biases for v1.6 in the SH disappears when the smoothing is applied to the sonde profiles.

For all latitude bands RMSD profiles display the largest values around the tropopause (below 60% in the extra tropics and up to 100% in the NH tropics) as is expected because it is the altitude range with the largest relative variability. RMS from differences between retrievals and smoothed data are generally much lower than with raw data. This is also expected since the smoothing error is the largest source of error in IASI retrievals (see Barret et al. (2011); Dufour et al. (2012)). RMS of the differences with smoothed sondes in the troposphere are somewhat larger for v3.5 than v1.6 especially in the SH. This is an indication of the increased sensitivity and decreased smoothing of v3.5. This is also evident in the Taylor diagrams which show that tropospheric variabilities are larger and in better agreement with sonde data (raw and smoothed) for v3.5 (see Fig. 4).

4.4 Time series of tropospheric columns

As tropospheric O_3 trends assessment is a major issue and one of the main topic of the TOAR (Tropospheric Ozone Assessment Report)/IGAC international initiative (Gaudel et al., 2018), we focus in this section on TOCs time series. Time series are also interesting to bring insight about the general statistics discussed in the previous sections and to identify possible drifts of the data.

The time series of IASI and sondes monthly TOCs are presented in Fig. 9 (resp. 10) for V1.6 and in Fig. 11 (resp. 12) for V3.5 for northern (resp. southern) hemisphere. We present both raw and smoothed sonde data to highlight the impact of smoothing upon the agreement between IASI and sondes. This impact is particularly obvious for SOFRID v1.6 at mid-latitudes. At northern mid-latitudes the bias between SOFRID v1.6 and raw sonde TOCs displays large seasonal variations from -(5-10)% in summer to 10-20% in winter resulting in a negligible 2±15% average bias (Table 2). When sonde data are smoothed by IASI AKs, the sonde variability is largely reduced. Bias is varying from 5% in winter to -5% in summer.

For southern mid-latitudes, as already highlighted by Dufour et al. (2012) and Emili et al. (2014) SOFRID TOCs are significantly biased high (29±22%) relative to raw sonde data (Table 2). This was explained by the fact that the single a priori used in v1.6 is biased towards northern mid-latitude O_3 (Emili et al., 2014). When the sonde data are smoothed by IASI AKs, the agreement is much better and the bias becomes unsignificant (5±9%) as a result of taking the a priori contribution into account (Equ. 1). The largest significant bias (56±25%) is found in the SH high latitudes for v1.6 TOCs (Table 2) with large seasonal variations from 20% in winter to 120% in summer. The large bias variabilities at mid- and especially high latitudes of the SH result from the very low seasonal variability of the retrieved columns (see Fig. 4(a)).

For V3.5, the use of a dynamical a priori profile clearly improves the retrievals at mid-latitudes. At northern mid-latitudes the seasonal bias variation is reduced to -10-0% and the average bias remains small (-6±14%). When smoothing is applied, the seasonal variability almost disappears and the bias is only -3±9%. At southern mid-latitudes, the agreement is very good and very similar for raw and smoothed sonde data with no real seasonal signature detectable and an avergae bias close to 0%.

At tropical latitudes, the situation is quite different. First, the seasonal variability is not so notable and regular and the difference between raw and smoothed sondes is lower than at mid-latitudes. Furthermore, the behaviour of v1.6 and v3.5 are close even though v3.5 is in better agreement with sonde data (see section 4.1). In the southern tropics there is a noticeable variation of bias between 2011-2014 with large negative biases of -10% and -15% and 2008-2010 and 2015-2017 with biases of 0 and -5% for v1.6 and v3.5 respectively. As such a bias variation is not detected for other latitude bands, we assume that it may be linked to a gap in sonde data for the 2011-2014 period. A closer look to SH tropics ECC sonde data show that only two stations (La Reunion and Nairobi) provide data regularly (30-50 profiles per year) over the period. For the Pago-Pago Pacific station data are available only from 2014 to 2016 and since 2012 for Irene in South Africa. For the Natal Atlantic station more than 25 profiles are available during the 2008-2010 and 2014-2017 period and almost none during the 2011-2013 period.

One issue that was raised in TOAR (Gaudel et al., 2018) was the different trends computed from different satellite products. UV-Visible satellite sensors produce positive tropospheric O_3 burden trends in both hemisphere while trends from IASI products are negative. It has to be noted that in Gaudel et al. (2018), negative O_3 burden trends from SOFRID **v1.5** in the northern and southern hemisphere, and for the whole Earth are respectively 1/4, 1/2 and 1/3 smaller than FORLI's. The drifts computed from the SOFRID-sonde differences are displayed in Fig. 9 to 12.

At high northern latitudes, for both v1.6 and v3.5 the drifts are large (> 9 and > 5%.decade^{-1} for raw and smoothed data resp.) and significant at the 95% level. For mid and tropical latitudes, drifts are between 0.9 and -3.4 %.decade^{-1} but are not significant. The NH mid-latitude drift with raw sonde data is reduced from -3.1 with v1.6 to -0.4%.decade^{-1} with v3.5. For the whole NH, the drifts are not significant and decreases from -2.2 with v1.6 to 0.7%.decade^{-1} with v3.5 for raw sonde data. They are < 1.5±0.8%.decade^{-1} and hardly significant (p>0.10) for smoothed sonde data.

In the SH tropics, drifts are ∼-5 and ∼-3%.decade^{-1} for raw and smoothed sonde data resp. and only significant for v3.5 compared to raw data. These drifts are linked to the large negative biases of the 2011-2014 period resulting from misssing data (see above). For v1.6 a large but unsignificant drift (-8%) also occurs at high latitudes which is largely reduced for v3.5. For the whole SH we found a significant negative drift (relative to raw sonde data) of -9.5±4.7%.decade^{-1} for v1.6 which reduces to -4.3±1.4%.decade^{-1} and becomes unsignificant for v3.5.

5 Comparison with IASI-FORLI

Two versions of IASI O_3 retrievals with the FORLI software have been validated by Boynard et al. (2016) and Boynard et al. (2018) (B18). Part of their validation results are based on the same data as the present study, namely ECC ozone sondes from the WOUDC database between 2008 and 2014 for Boynard et al. (2016) and 2008 and 2016 for B18. As they document the latest FORLI version (v20151001) on a longer time period, we will focus our comparison with B18. They have used a comparable number (11600) of ozone sonde profiles than in the present study and their comparison methodology is close to the one we have used (spatio-temporal coincidence criteria set to 100 km and ± 6 h). We have collected the correlation coefficients (r2), the biases, the RMSDs, the DFS of the retrievals and the slopes of the linear fit between the smoothed sondes and retrievals from B18.

There are some limitations to the comparison between the validation of our SOFRID retrievals and the FORLI validation from B18. We are comparing with litterature results which are not providing the same information as we do. For instance B18 do not document the sonde and IASI variabilities and it is therefore not possible to draw their data in Taylor diagrams. B18 have also limited their comparisons to smoothed sonde data. **Another limitation is that FORLI and SOFRID use their own quality flags to filter the data. In order to document the impact of the pixel selection on SOFRID validation we have performed the comparison with sonde data using modified quality flags. The cloud filtering trheshold is the clearest source of difference between the pixel selection of both algorithms. We have therefore lowered the upper limit of the AVHRR cloud fraction cover to 13% which is the trheshold used by B18 resulting in a loss of 5% of the treated pixels. The Jcost threshold has been decreased from 1.0 to 0.15 with a 6% decrease of the selected retrieved profiles. Finally the DFS lower value has been set to 1.75 increasing the number of selected retrievals by 2%. These threshold modifications resulted in negligible changes of the general statistics (bias, RMSD, R) for the 3 atmospheric layers (troposphere, UTLS and stratosphere) and the different latitude bands that are presented in this section. These statistics, based on large**

numbers of data are therefore not hindered by pixel selection differences.

In Fig. 13 we have drawn DFS from SOFRID v1.6 and 3.5 and from FORLI for the layers selected by B18 (1013-300 hPa, 300-150 hPa and 150-25 hPa). Fig. 14 displays the correlation coefficients (r2) and the slopes (b) from linear relationships fitted between IASI retrievals and smoothed sonde data. Biases and RMSDs are shown for the three retrievals in Fig. 15. Finally, Fig. 16 documents the drifts between sondes and SOFRID retrievals for the whole NH for the Surface-300 hPa layer to be comparable to B18.

In the three atmospheric layers, the information content is larger with FORLI than with SOFRID v1.6 and v3.5 (Fig. 13). This is particularly visible for the mid-latitudes and tropics in the troposphere with 0.8 to 0.9 DFS for FORLI and only 0.4 to 0.6 for SOFRID. **This probably results from the retrieval noise level which is lower for FORLI than for SOFRID (Dufour et al., 2012)**. At high latitudes the DFS are low and closer for both algorithms and the increase from high to mid-latitudes is therefore much larger for FORLI than for SOFRID. As both algorithms use a single retrieval noise and a priori covariance matrix and similar surface and atmospheric temperatures, the reason for such a difference is unclear. In the UTLS and stratosphere the same increase of DFS from high latitudes to the tropics are visible for the three products. The difference in information content between retrievals is less pronounced in the UTLS and in the stratosphere than in the troposphere.

The RMSDs (see Fig. 15) are generally larger for FORLI than for SOFRID. In the troposphere, RMSDs reach 18% for FORLI and are below 10% for both SOFRID v1.6 and v3.5. In the UTLS, RMSDs are larger than in the other layers due to the lower absolute columns. For SOFRID UTLS RMSDs are in the range 10 to 30% and 20 to 45% for FORLI. For both SOFRID and FORLI the highest RMSDs are in the tropics where the 150-300 hPa columns are the lowest. In the stratosphere, FORLI's RMSDs are also systematically larger than SOFRID's. The differences are the largest at high latitudes with FORLI RMSDs 3 to 4 times larger than SOFRID's.

The r2 differences (Fig. 14) are partly related to the RMSDs differences. Generally, SOFRID has larger r2 than FORLI. As for the RMSDs, the differences between both algorithms are the largest at high latitudes (especially in the southern hemisphere where r2 < 0.4 for FORLI products) in the 3 layers. In the troposphere, the correlation coefficients are comparable for both algorithms in the tropical bands and SOFRID 3.5 gives higher r2 than SOFRID 1.6. The differences between retrieval versions are generally lower and can even be reversed in the UTLS and in the stratosphere.

The slopes of the linear fits between retrievals and sonde data provide complementary information to the r2 coefficients. A slope smaller than one indicates that the retrieved variability is too low compared to the reference data and conversly, a slope larger than one indicates an overestimation of the variability. In the troposphere, SOFRID v1.6 and 3.5 and FORLI have similar slopes except in the 60-90°S band where FORLI has a significantly lower slope than SOFRID (Fig. 14).

In the troposphere, FORLI products present systematic negative biases from 7 to 20% except in the polar regions. Concerning SOFRID, the tropospheric biases are within ±6% (comparable to TOCs biases in Table 2). The results are largely different when the raw sonde data are considered with very large biases in the southern hemisphere with SOFRID v1.6 as discussed in section 4.4. In the UTLS, SOFRID and FORLI biases are significantly positive except in the tropics and more specifically in the SH tropics where SOFRID columns are negatively biased by ∼20% as discusssed in section 4.1 (Table 2). In the stratosphere, both SOFRID and FORLI products are positively biased. The largest differences between both retrieval algorithms are found in the extra tropical southern latitudes with FORLI biases larger than SOFRID. In the 60-90°S latitude band FORLI biases reach about 40 % against about 5% for SOFRID.

In the perspective of a better quantification of tropospheric O_3 evolution and of the TOAR results (Gaudel et al., 2018), it is also important to compare the drifts between sonde and retrievals. B18 present and discuss the drift between FORLI and sonde data for different layers in the whole NH. The SOFRID NH tropospheric drifts discussed in section 4.4 are smaller and opposite in sign to the significant -8.6±3.4%.decade^{-1} drift between FORLI and smoothed sonde data in the NH troposphere presented in B18. As B18 computed a surface-300 hPa column instead of a tropospheric column, we have computed the drifts based on the same layer (see Fig. 16). Drifts for Surface-300hPa columns are slightly (0.1 to 0.4%) smaller than for TOCs and are not significant in both cases. The comparison of the NH drift with B18 is therefore not dependent on the tropospheric layer definition. For v1.6 and v3.5 compared with raw and smoothed sonde data, the surface-300 hPa column drifts range from -2.0 to 1.3%.decade^{-1} (see Fig. 16), values which are much smaller than in B18. **Nevethelesss, the NH tropospheric drift from FORLI is attributed to an abrupt change or jump detected in 2010 (Boynard et al., 2018; Wespes et al., 2018). Indeed, the drift strongly decreases after the jump and it becomes even non-significant for most of the stations over the periods before or after the jump, separately (Wespes et al., 2018). The discontinuity is suspected to result from updates in level-2 temperature data from EUMETSAT used as inputs into FORLI (Wespes et al., 2019). The absence of jump and the small drift in SOFRID v1.6 (Fig. 9(h)) and v3.5 (Fig. 11(h)) NH tropospheric data is therefore probably linked to the use of temperature profiles from ECMWF analyses instead of EUMETSAT L2 products.**

6 Conclusions

This study aimed at assessing the quality of two different versions of SOFRID-O3 at the global scale and over the 10 year IASI period using ozonesonde from the WOUDC. SOFRID-O3 v1.6 retrievals are based on a single a priori profile like most other global IASI O_3 retrievals (Barret et al., 2011; Dufour et al., 2012; Boynard et al., 2016, 2018). In V3.5 the a priori is dynamically selected from an O_3 profile climatology (Sofieva et al., 2014) based on latitude, season and the tropopause height. Other satellite O_3 retrievals use a priori profiles from climatologies but they are chosen based on geographical and temporal criteria only (Bowman et al., 2006; Liu et al., 2010). Dufour et al. (2015) use 3 different a priori profiles picked up according to 3 broad tropopause height classes to represent high, mid and tropical latitudes. To our knowledge it is the first time that the tropopause height is used **in such a comprehensive way** for the choice of the a priori for spaceborne O_3 retrievals.

The general statistics (Taylor diagrams) of the comparisons between ozonesonde and SOFRID have highlighted the large improvements brought by v3.5 especially in the troposphere. The use of a tropopause based a priori generally reduces the RMSDs and increases the r2 correlation coefficients and the amplitude of the retrieved variability. The high TOC biases of v1.6 relative to low O_3 is also corrected with v3.5. This is of particular importance in the SH extratropics where the very large biases almost dissapear. In the NH lower TOC are retrieved in winter leading to a better seasonal cycle. **A sensitivity test demonstrated that these SOFRID improvements are dominated by the seasonal- and latitude- dependence of the a priori.**

In the UTLS and stratosphere the improvements are less important. In particular both versions are impacted by positive biases for the UTLS (18% at NH mid-latitudes) and stratospheric (<7%) columns at extratropical latitudes that were already discussed in Dufour et al. (2012). In the tropics large profile oscillations around the tropopause result in negative biases in the UTLS (21% in the SH) and positive biases (< 14%) in the stratospheric columns.

Concerning the TOC drifts, we have shown that there were no significant differences between v1.6 ans v3.5. There are no significant drifts except at high northern latitudes (increase of 9-13%.dec) and at southern tropical latitudes (decrease of 4-5%.dec). For southern tropics, the apparent decrease is probably linked to a sampling weakness at different stations which makes the time serie inhomegeneous.

Our study have also demonstrated the importance of making comparisons with both raw and smoothed in-situ data. Comparing only with smoothed data could lead to the conclusion that the satellite data are better than they really are. For instance, the high bias for low TOC with the v1.6 is almost completly corrected when smoothing is applied. The real improvement of v3.5 relative to v1.6 is only sizeable when we compare SOFRID retrievals with raw sonde data.

Finally we have compared our validation results to the latest (v20151001) FORLI-O3 retrievals validation. The comparison had to be limited because the variability of FORLI-O3 retrievals and ozonesonde data were not provided in Boynard et al. (2018) which prevented us to draw Taylor diagrams. Furthermore, in Boynard et al. (2018) the FORLI-O3 are compared to smoothed sonde data only. FORLI produces larger RMSDs than SOFRID especially in the stratosphere at high latitudes. The correlation coefficients (r2) are consequently lower for FORLI columns than for SOFRID. Tropospheric biases are significantly larger for FORLI (7-20%) than for SOFRID (<6%). Finally, no significant tropospheric O_3 drift are detected for both versions of SOFRID-O3 in the NH. The difference with FORLI which is impacted by a significant TOC jump in 2010 (Boynard et al., 2018; Wespes et al., 2018) is likely linked to the use of different temperature profiles for the radiative transfer calculations (ECMWF analyses for SOFRID and EUMETSAT L2 for FORLI).

Data availability. The SOFRID-O_3 data are freely available at the IASI-SOFRID website (http://thredds.sedoo.fr/iasi-sofrid-o3-co/).

Author contributions. Brice Barret performed the validation of SOFRID-O_3 data and wrote the paper. Emanuele Emili initiated and contributed to the development of SOFRID-O_3 v3.5. Eric Le Flochmoen is in charge of the SOFRID retrieval operations.

Competing interests. no competing interests are present for the present publication

Acknowledgements. IASI L1c data have been downloaded from the Ether French atmospheric database (http://ether.ipsl.jussieu.fr). The research with IASI is conducted with some financial support from the CNES (TOSCA–IASI project). The ozonesonde data used in this study were provided by the World Ozone and Ultraviolet Data Centre (WOUDC) (http://www.woudc.org). The authors thank those responsible for the WOUDC measurements and archives for making the ozonesonde data available.

References

Ainsworth, E., Yendrek, C., Sitch, S., Collins, W., and Emberson, L.: The Effects of Tropospheric Ozone on Net Primary Productivity and Implications for Climate Change, Annual Review of Plant Biology, 63, 637–661, https://doi.org/10.1146/annurev-arplant-042110-103829, 2012.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stubi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P. F., Ziemke, J. R., Rozanov, E. V., De Maziere, M., and Demoulin, P.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmos. Chem. Phys., 18, 1379–1394, https://doi.org/10.5194/acp-18-1379-2018, 2018.

Barret, B., De Maziere, M., and Demoulin, P.: Retrieval and characterization of ozone profiles from solar infrared spectra at the Jungfraujoch, Journal of Geophysical Research-Atmospheres, 107, https://doi.org/10.1029/2001JD001298, 2002.

Barret, B., Turquety, S., Hurtmans, D., Clerbaux, C., Hadji-Lazaro, J., Bey, I., Auvray, M., and Coheur, P.: Global carbon monoxide vertical distributions from spaceborne high-resolution FTIR nadir measurements, Atmospheric Chemistry and Physics, 5, 2901–2914, 2005.

Barret, B., Le Flochmoen, E., Sauvage, B., Pavelin, E., Matricardi, M., and Cammas, J. P.: The detection of post-monsoon tropospheric ozone variability over south Asia using IASI data, Atmospheric Chemistry and Physics, 11, 9533–9548, https://doi.org/10.5194/acp-11-9533-2011, 2011.

Barret, B., Sauvage, B., Bennouna, Y., and Flochmoen, L.: Upper tropospheric CO and O3 budget during the Asian Summer Monsoon, Atmospheric Chemistry and Physics, 16, 9129–9147, https://doi.org/10.5194/acp-16-9129-2016, 2016.

Bowman, K. W., Rodgers, C. D., Kulawik, S. S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Lou, M., Eldering, A., Shephard, M., Worden, H., Lampel, M., Clough, S., Brown, P., Rinsland, C., Gunson, M., , and Beer, R.: Tropospheric emission spectrometer: Retrieval method and error analysis, IEEE T. Geosci. Remote, 44, 1297–1307, 2006.

Boynard, A., Hurtmans, D., Koukouli, M. E., Goutail, F., Bureau, J., Safieddine, S., Lerot, C., Hadji-Lazaro, J., Wespes, C., Pommereau, J. P., Pazmino, A., Zyrichidou, I., Balis, D., Barbe, A., Mikhailenko, S. N., Loyola, D., Valks, P., Van Roozendael, M., Coheur, P. F., and Clerbaux, C.: Seven years of IASI ozone retrievals from FORLI: validation with independent total column and vertical profile measurements, Atmospheric Measurement Techniques, 9, 4327–4353, https://doi.org/10.5194/amt-9-4327-2016, 2016.

Boynard, A., Hurtmans, D., Garane, K., Goutail, F., Hadji-Lazaro, J., Koukouli, M. E., Wespes, C., Vigouroux, C., Keppens, A., Pommereau, J. P., Pazmino, A., Balis, D., Loyola, D., Valks, P., Sussmann, R., Smale, D., Coheur, P. F., and Clerbaux, C.: Validation of the IASI FORLI/EUMETSAT ozone products using satellite (GOME-2), ground-based (Brewer-Dobson, SAOZ, FTIR) and ozonesonde measurements, Atmospheric Measurement Techniques, 11, 5125–5152, https://doi.org/10.5194/amt-11-5125-2018, <GotoISI>://WOS:000444187600001, 2018.

Brunekreef, B. and Holgate, S. T.: Air pollution and health, The Lancet, 360, 1233–1242, 2002.

Chen, W., Liao, H., and Seinfeld, J.: Future climate impacts of direct radiative forcing of anthropogenic aerosols, tropospheric ozone, and long-lived greenhouse gases, Journal of Geophysical Research-Atmospheres, 112, https://doi.org/10.1029/2006jd008051, 2007.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P. F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, Atmospheric Chemistry and Physics, 9, 6041–6054, 2009.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., and P.D., B.: Atmospheric radiative transfer modeling: A summary of the AER codes, J. Quant. Spectrosc. Radiat. Transfer, 91, 233–244, 2005.

Coheur, P., Barret, B., Turquety, S., Hurtmans, D., Hadji-Lazaro, J., and Clerbaux, C.: Retrieval and characterization of ozone vertical profiles from a thermal infrared nadir sounder, Journal of Geophysical Research, 110, https://doi.org/10.1029/2005JD005845, 2005.

De Maziere, M., Hennen, O., Van Roozendael, M., Demoulin, P., and De Backer, H.: Daily ozone vertical profile model built on geophysical grounds, for column retrieval from atmospheric high-resolution infrared spectra, Journal of Geophysical Research-Atmospheres, 104, 23 855–23 869, https://doi.org/10.1029/1999jd900347, 1999.

Dufour, G., Eremenko, M., Griesfeller, A., Barret, B., LeFlochmoen, E., Clerbaux, C., Hadji-Lazaro, J., Coheur, P. F., and Hurtmans, D.: Validation of three different scientific ozone products retrieved from IASI spectra using ozonesondes, Atmospheric Measurement Techniques, 5, 611–630, https://doi.org/10.5194/amt-5-611-2012, 2012.

Dufour, G., Eremenko, M., Cuesta, J., Doche, C., Foret, G., Beekmann, M., Cheiney, A., Wang, Y., Cai, Z., Liu, Y., Takigawa, M., Kanaya, Y., and Flaud, J. M.: Springtime daily variations in lower-tropospheric ozone over east Asia: the role of cyclonic activity and pollution as observed from space with IASI, Atmospheric Chemistry and Physics, 15, 10 839–10 856, https://doi.org/10.5194/acp-15-10839-2015, <GotoISI>://WOS:000362457400032, 2015.

Dufour, G., Eremenko, M., Beekmann, M., Cuesta, J., Foret, G., Fortems-Cheiney, A., Lachatre, M., Lin, W., Liu, Y., Xu, X., and Zhang, y.: Lower tropospheric ozone over the North China Plain: variability and trends revealed by IASI satellite observations for 2008?2016, Atmospheric Chemistry and Physics, 18, 16 439–16 459, https://doi.org/10.5194/acp-18-16439-2018, 2018.

Emili, E., Barret, B., Massart, S., Le Flochmoen, E., Piacentini, A., El Amraoui, L., Pannekoucke, O., and Cariolle, D.: Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical transport model, Atmospheric Chemistry and Physics, 14, 177–198, https://doi.org/10.5194/acp-14-177-2014, 2014.

Emili, E., Barret, B., Le Flochmoen, E., and Cariolle, D.: Comparison between the assimilation of IASI Level 2 retrievals and Level 1 radiances for ozone reanalyses, Atmos. Meas. Tech., 12, 3963?3984, https://doi.org/10.5194/amt-12-3963-2019, 2019.

Eremenko, M., Sgheri, L., Ridolfi, M., Cuesta, J., Costantino, L., Sellitto, P., and Dufour, G.: Tropospheric ozone retrieval from thermal infrared nadir satellite measurements: Towards more adaptability of the constraint using a self-adapting regularization, Journal of Quantitative Spectroscopy and Radiative Transfer, 238, https://doi.org/10.1016/j.jqsrt.2019.106577, <GotoISI>://WOS:000499767800014, 2019.

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P. F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados-Munoz, M. J., Hannigan, J. W., Hase, F., Hassler, B., Huang, G., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S., Latter, B., Leblanc, T., Le Flochmoen, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M., Trickl, T., Weatherhead, E., Wespes, C., Worden, H. M., Vigouroux, C., Xu, X., Zeng, G., and Ziemke, J.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, Elementa-Science of the Anthropocene, 6, https://doi.org/10.1525/elementa.291, <GotoISI>://WOS:000431754000001, 2018.

Hassler, B., Bodeker, G. E., and Dameris, M.: Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, Atmospheric Chemistry and Physics, 8, 5403–5421, https://doi.org/10.5194/acp-8-5403-2008, <GotoISI>://WOS:000259221400022, 2008.

Havemann, S., NWPSAF 1D-Var User Manual Software Version 1.2, Report NWPSAF-MO-UD-032, EUMETSAT Numerical Weather Prediction Satellite Applications Facility, Met Office, Exeter, U.K., www.nwpsaf.eu, 2020.

Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P., , and Vidot, J.: RTTOV v11 Users Guide, NWP SAF, 2015.

Liu, X., Bhartia, P. K., Chance, K., Spurr, R. J. D., , and Kurosu, T. P.: Ozone profile retrievals from the Ozone Monitoring Instrument, Atmospheric Chemistry and Physics, 10, 2521?2537, https://doi.org/10.5194/acp-10-2521-2010, 2010.

Matricardi, M., Chevallier, F., Kelly, G., , and Thepaut, J. N.: An improved general fast radiative transfer model for the assimilation of radiance observations, Quarterly Journal of the Royal Meteorological Society, 130, 153–173, 2004.

McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, Journal of Geophysical Research-Atmospheres, 112, https://doi.org/10.1029/2005jd006823, 2007.

Pavelin, E., English, S., and Eyre, J.: The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction, Q.J.R. Meteorol. Soc., 134, 737–749, 2008.

Peiro, H., Emili, E., Cariolle, D., Barret, B., and Le Flochmoen, E.: Tropospheric Ozone over the Tropical Pacific Ocean in response to ENSO : a contribution from the assimilation of IASI and MLS data, Atmospheric Chemistry and Physics, 18, 6939–6958, 2018.

Richards, N. A. D., Arnold, S. R., Chipperfield, M. P., Miles, G., Rap, A., Siddans, R., Monks, S. A., and Hollaway, M. J.: The Mediterranean summertime ozone maximum: global emission sensitivities and radiative impacts, Atmospheric Chemistry and Physics, 13, 2331–2345, https://doi.org/10.5194/acp-13-2331-2013, 2013.

Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics - Vol. 2, World Scientific, Singapore, New Jersey, London, Hong Kong, 238pp., 2000.

Rothman, L. S. and Jacquemart, D. Barbe, A. e. a.: The HITRAN 2004 molecular spectroscopic database, J. Quant. Spectrosc. Radiat. Transfer, 96, 139–204, 2005.

Rothman, L. S., Gordon, I. E., and Barbe, A. e. a.: The HITRAN 2008 molecular spectroscopic database, J. Quant. Spectrosc. Radiat. Transfer, 110, 533–572, https://doi.org/10.1016/j.jqsrt.2009.02.013, <GotoISI>://WOS:000266181300002, 2009.

Saunders, R., Matricardi, M., and Brunel, P.: An improved fast radiative transfer model for assimilation of satellite radiance observations, Quarterly Journal of the Royal Meteorological Society, 125, 1407–1425, 1999.

Sauvage, B., Thouret, V., Thompson, A. M., Witte, J. C., Cammas, J. P., Nedelec, P., and Athier, G.: Enhanced view of the "tropical Atlantic ozone paradox" and "zonal wave one" from the in situ MOZAIC and SHADOZ data, Journal of Geophysical Research-Atmospheres, 111, https://doi.org/10.1029/2005jd006241, 2006.

Sellitto, P., Dufour, G., Eremenko, M., Cuesta, J., Dauphin, P., Foret, G., Gaubert, B., Beekmann, M., Peuch, V. H., and Flaud, J. M.: Analysis of the potential of one possible instrumental configuration of the next generation of IASI instruments to monitor lower tropospheric ozone, Atmospheric Measurement Techniques, 6, 621–635, https://doi.org/10.5194/amt-6-621-2013, <GotoISI>://WOS: 000317011200009, 2013.

Shindell, D. T., Faluvegi, G., and Stevenson, D. S. e. a.: Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, Journal of Geophysical Research-Atmospheres, 111, https://doi.org/10.1029/2006jd007100, 2006.

Sofieva, V. F., Tamminen, J., Kyrola, E., Mielonen, T., Veeflkind, P., Hassler, B., and Bodeker, G. E.: A novel tropopause-related climatology of ozone profiles, Atmospheric Chemistry and Physics, 14, 283–299, https://doi.org/10.5194/acp-14-283-2014, <GotoISI>://WOS: 000329930600020, 2014.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research-Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000jd900719, <GotoISI>://WOS:000168189400001, 2001.

Thompson, A., Witte, J., and Oltmans, S. e. a.: Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998-2000 tropical ozone climatology – Tropospheric variability and the zonal wave-one, Atmos. Chem. Phys., 108, 8241, 2003.

Table 1. Atmospheric layers for comparisons

Layer	Lower boundary	Upper Boundary
Troposphere-1	Ground	Tropopause
Troposphere-2	Ground	300 hPa
Lower Troposphere	Ground	550 hPa
UTLS	300 hPa	150 hPa
Stratosphere	150 hPa	25 hPa

Wang, P. H., Cunnold, D. M., Trepte, C. R., Wang, H. J., Jing, P., Fishman, J., Brackett, V. G., Zawodney, J. M., and Bodeker, G. E.: Ozone variability in the midlatitude upper troposphere and lower stratosphere diagnosed from a monthly SAGE II climatology relative to the tropopause, Journal of Geophysical Research-Atmospheres, 111, https://doi.org/10.1029/2005jd006108, <GotoISI>://WOS:000241984500001, 2006.

Wespes, C., Hurtmans, D., Clerbaux, C., and Coheur, P. F.: O-3 variability in the troposphere as observed by IASI over 2008-2016: Contribution of atmospheric chemistry and dynamics, Journal of Geophysical Research-Atmospheres, 122, 2429–2451, https://doi.org/10.1002/2016jd025875, <GotoISI>://WOS:000396121200024, 2017.

Wespes, C., Hurtmans, D., Clerbaux, C., Boynard, A., and Coheur, P. F.: Decrease in tropospheric O-3 levels in the Northern Hemisphere observed by IASI, Atmospheric Chemistry and Physics, 18, 6867–6885, https://doi.org/10.5194/acp-18-6867-2018, <GotoISI>://WOS:000432392400005, 2018.

Wespes, C., Hurtmans, D., Chabrillat, S., Ronsmans, G., Clerbaux, C., and Coheur, P. F.: Is the recovery of stratospheric O-3 speeding up in the Southern Hemisphere? An evaluation from the first IASI decadal record (2008-2017), Atmospheric Chemistry and Physics, 19, 14 031–14 056, https://doi.org/10.5194/acp-19-14031-2019, <GotoISI>://WOS:000498797200003, 2019.

WMO: WMO: Meteorology - A three-dimensional science: Second session of the Commission for Aerology, WMO Bulletin, 4, 134–138, 1957.

Zhang, L., Jacob, D. J., Bowman, K. W., Logan, J. A., Turquety, S., Hudman, R. C., Li, Q. B., Beer, R., Worden, H. M., Worden, J. R., Rinsland, C. P., Kulawik, S. S., Lampel, M. C., Shephard, M. W., Fisher, B. M., Eldering, A., and Avery, M. A.: Ozone-CO correlations determined by the TES satellite instrument in continental outflow regions, Geophysical Research Letters, 33, https://doi.org/10.1029/2006gl026399, <GotoISI>://WOS:000240826900002, 2006.

Zhang, L., Li, Q. B., Murray, L. T., Luo, M., Liu, H., Jiang, J. H., Mao, Y., Chen, D., Gao, M., and Livesey, N.: A tropospheric ozone maximum over the equatorial Southern Indian Ocean, Atmospheric Chemistry and Physics, 12, 4279–4296, https://doi.org/10.5194/acp-12-4279-2012, 2012.

Ziemke, J., Oman, L., and Strode, S. e. a.: Trends in global tropospheric ozone inferred from a composite record of TOMS/OMI/MLS/OMPS satellite measurements and the MERRA-2 GMI simulation, Atmos. Chem. Phys., 19, 3257–3269, https://doi.org/10.5194/acp-19-3257-2019, 2019.

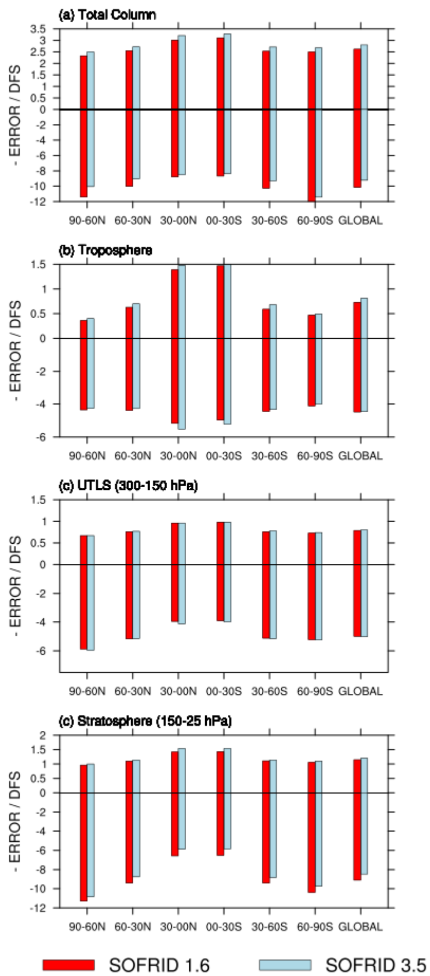

Figure 1. Degrees of Freedom for Signal (DFS) and (-) retrieval errors in Dobson Units (DU) for SOFRID-O3 V1.6 (red) and V3.5 (light blue) retrievals for (a) total column (b) Troposphere (c) UTLS (300-150 hPa) and (d) Stratosphere (150-25 hPa)

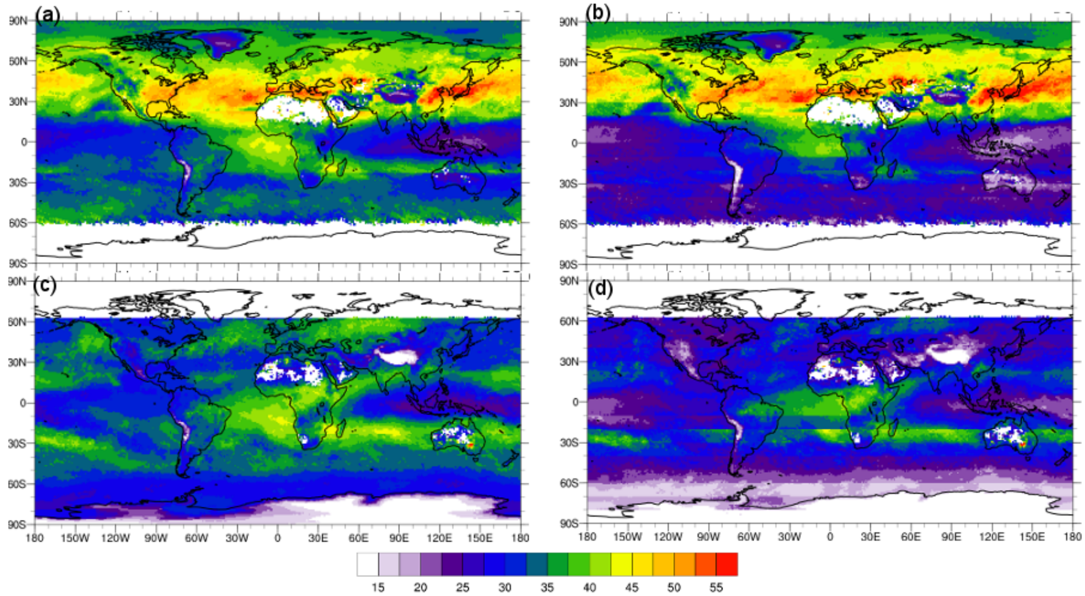

Figure 2. Tropospheric Ozone Column (TOC) distributions in Dobson Units (DU) for (a) July 2017 v1.6 (b) July 2017 v3.5 (c) December 2017 v1.6 and (d) December 2017 v3.5.

Table 2. Biases (%) between sondes and SOFRID retrievals with corresponding RMSDs (%). Values between brackets correspond to smoothed sonde data. Significant biases (Bias > RMSD) are in bold characters

Latitude band	SOFRID	Troposphere	UTLS	Stratosphere
90-60N	v1.6	6 ± 14 (0 ± 6)	6 ± 18 (10 ± 10)	7 ± 10 (4 ± 6)
	v3.5	-2 ± 14 (-1 ± 7)	10 ± 15 (11 ± 10)	1 ± 3 (3 ± 6)
60-30N	v1.6	2 ± 15 (0 ± 8)	18 ± 27 (13 ± 16)	2 ± 8 (4 ± 6)
	v3.5	-6 ± 14 (-3 ± 9)	17 ± 27 (13 ± 17)	1 ± 7 (3 ± 6)
30-00N	v1.6	2 ± 17 (4 ± 11)	-3 ± 30 (1 ± 37)	**14 ± 8** (**12 ± 7**)
	v3.5	-3 ± 16 (0 ± 14)	-12 ± 33 (-13 ± 39)	**12 ± 8** (**12 ± 7**)
00-30S	v1.6	-2 ± 14 (-2 ± 10)	-21 ± 27 (-16 ± 25)	**14 ± 10** (**10 ± 8**)
	v3.5	-8 ± 14 (-7 ± 12)	-21 ± 30 (-24 ± 25)	10 ± 11 (10 ± 11)
30-60S	v1.6	**29 ± 22** (5 ± 9)	11 ± 29 (13 ± 22)	1 ± 8 (4 ± 7)
	v3.5	1 ± 18 (1 ± 13)	10 ± 28 (13 ± 23)	3 ± 7 (4 ± 7)
60-90S	v1.6	**55 ± 25** (7 ± 6)	5 ± 22 (15 ± 13)	7 ± 12 (4 ± 7)
	v3.5	0 ± 16 (1 ± 9)	7 ± 19 (13 ± 13)	6 ± 11 (4 ± 8)

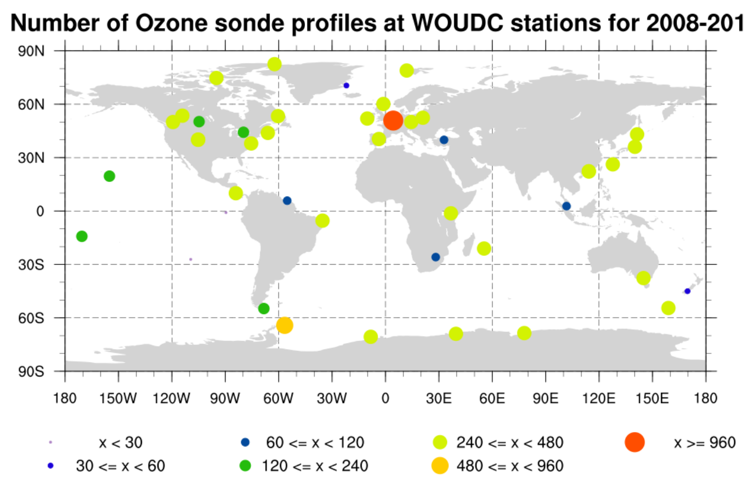

Figure 3. Maps of WOUDC stations with ECC O_3 sonde data during the 2008-2017 period. Colors and sizes of the markers indicate the number of valid sondes at each station.

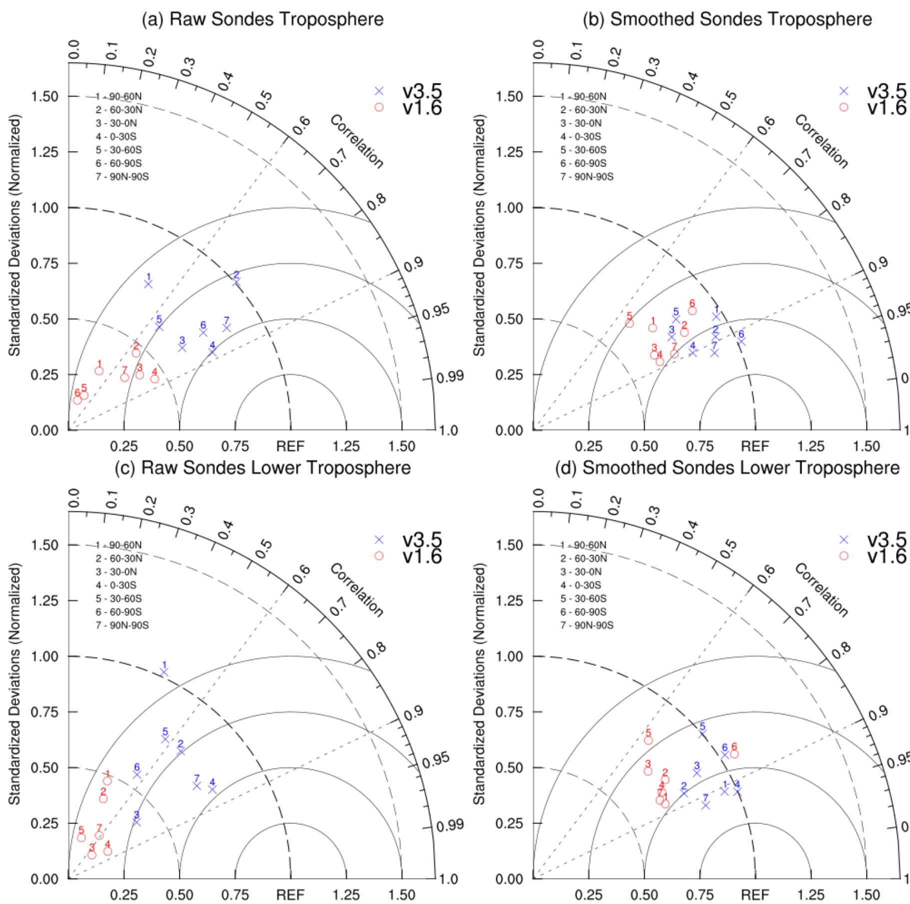

Figure 4. Taylor diagrams for (a) and (b) tropospheric columns and (c) and (d) lower tropospheric columns. (a) and (c) raw sonde data, (b) and (d) smoothed sonde data. Red circles (V1.6), Blue crosses (V3.5).

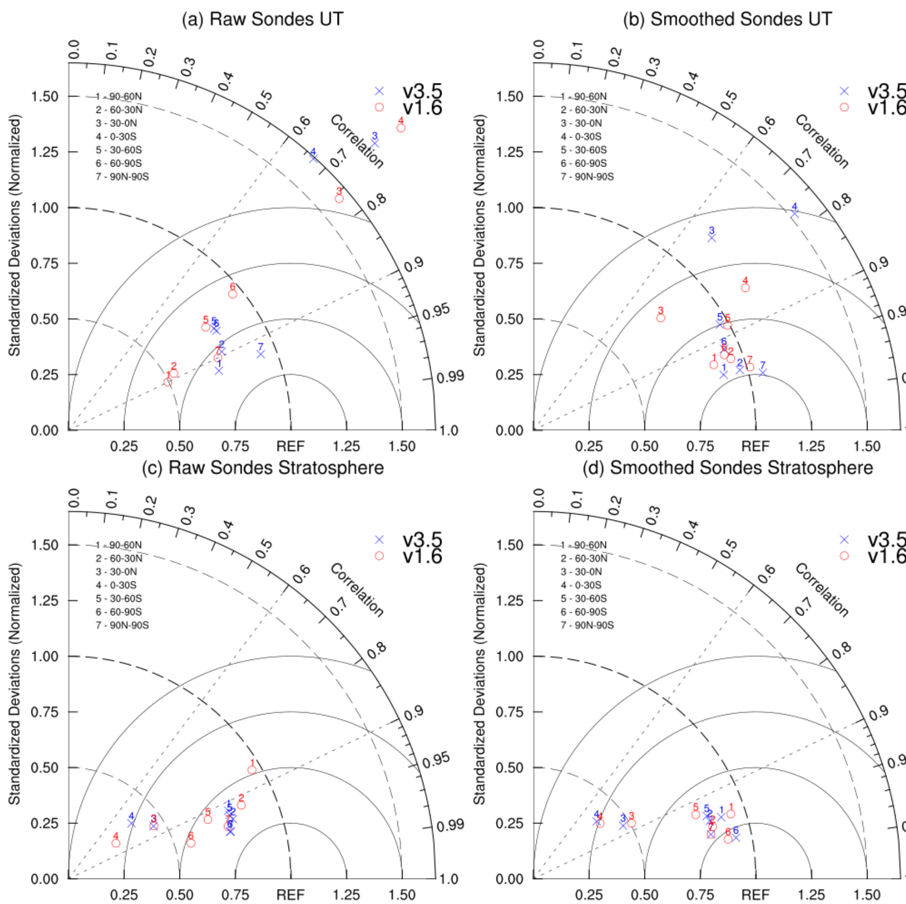

Figure 5. Taylor diagrams for (a) and (b) UTLS columns and (c) and (d) stratospheric columns. (a) and (c) raw sonde data, (b) and (d) smoothed sonde data. Red circles (V1.6), Blue crosses (V3.5).

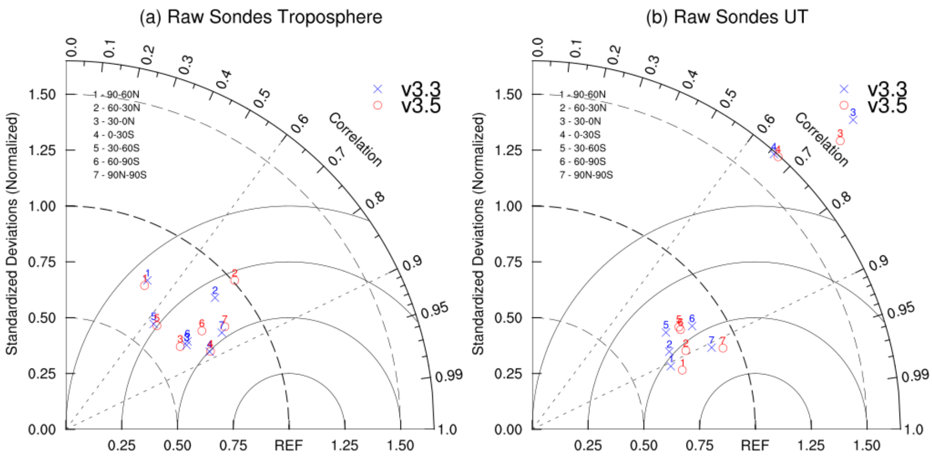

Figure 6. Taylor diagrams for (a) Tropospheric columns (b) UTLS columns for comparisons with raw sonde data. Red circles (V3.5), Blue crosses (V3.3).

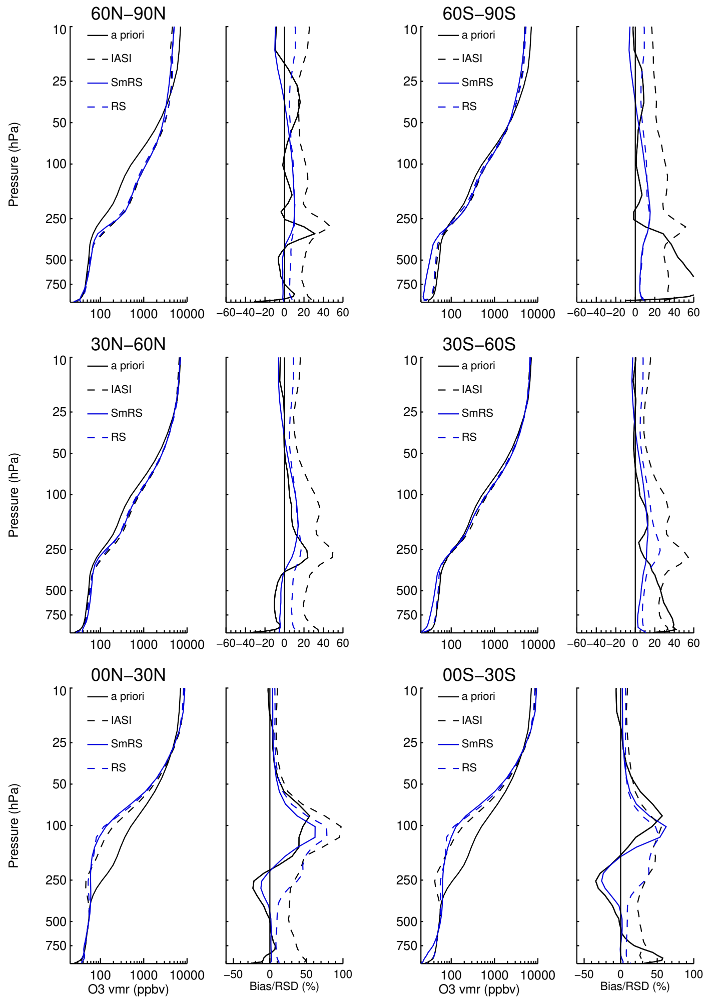

Figure 7. Profile comparisons between sonde and SOFRID-O3 v1.6 profiles (left panels) a priori (black solid lines), IASI (black dashed lines), smoothed (SmRS, blue solid lines) and raw (RS, blue dashed lines) sonde vertical profiles, (right panels) biases (solid lines) and RMSD (dashed lines) between IASI and raw (black lines) and smoothed (blue lines) sondes for the NH (left panels) and SH (right panels).

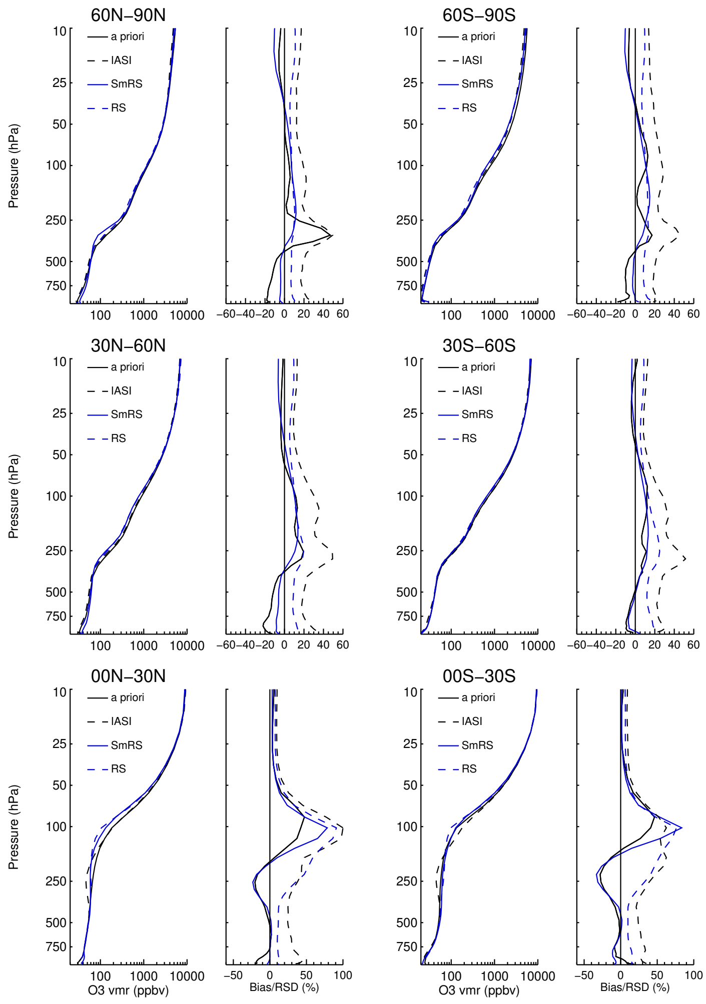

Figure 8. Same as 7 for SOFRID-O3 v3.5

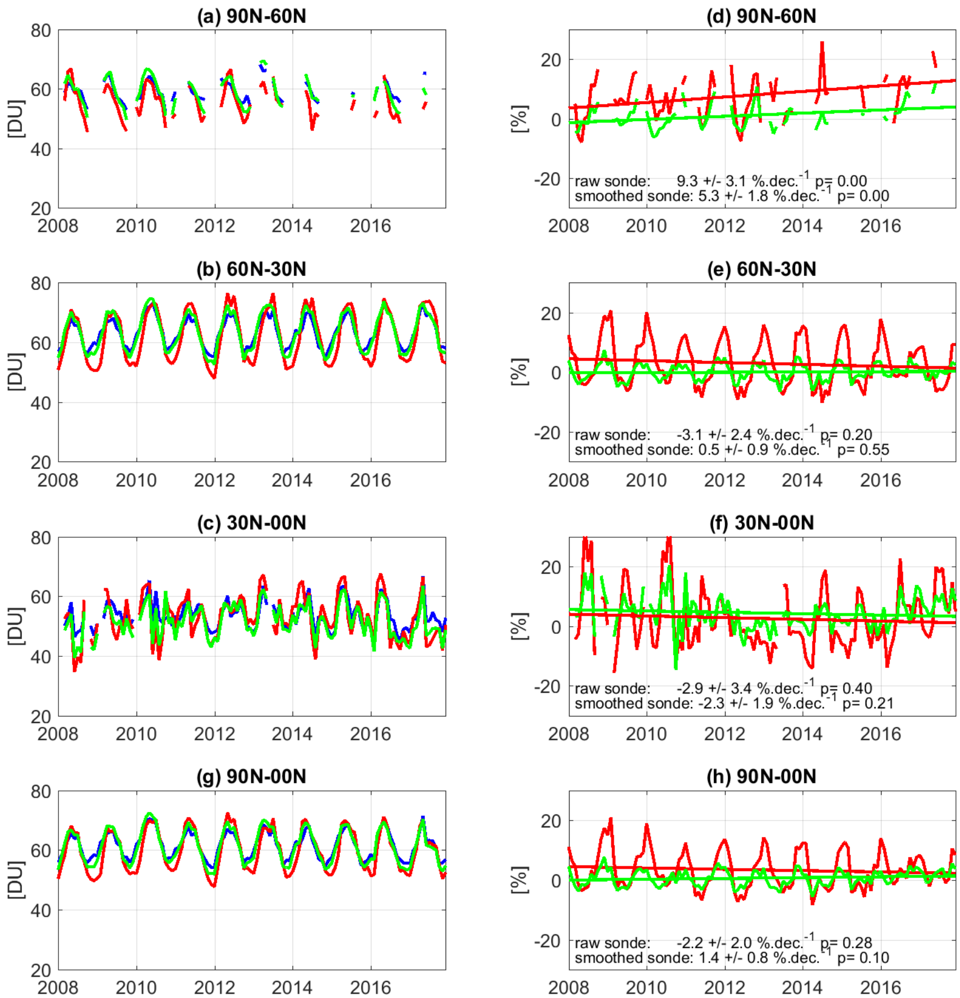

Figure 9. Time series of SOFRID-O3 v1.6 TOCs (DU) in the Northern Hemisphere for (a) 90-60°N (b) 60-30°N (c) 30-00°N (d) 90-00°N. Blue lines for IASI retrievals, red lines for raw sonde data and green lines for smoothed sonde data. Differences (%) between IASI and sonde data for (e) 90-60°N (f) 60-30°N (g) 30-00°N (h) 90-00°N. Red lines for raw sonde data and green lines for smoothed sonde data.

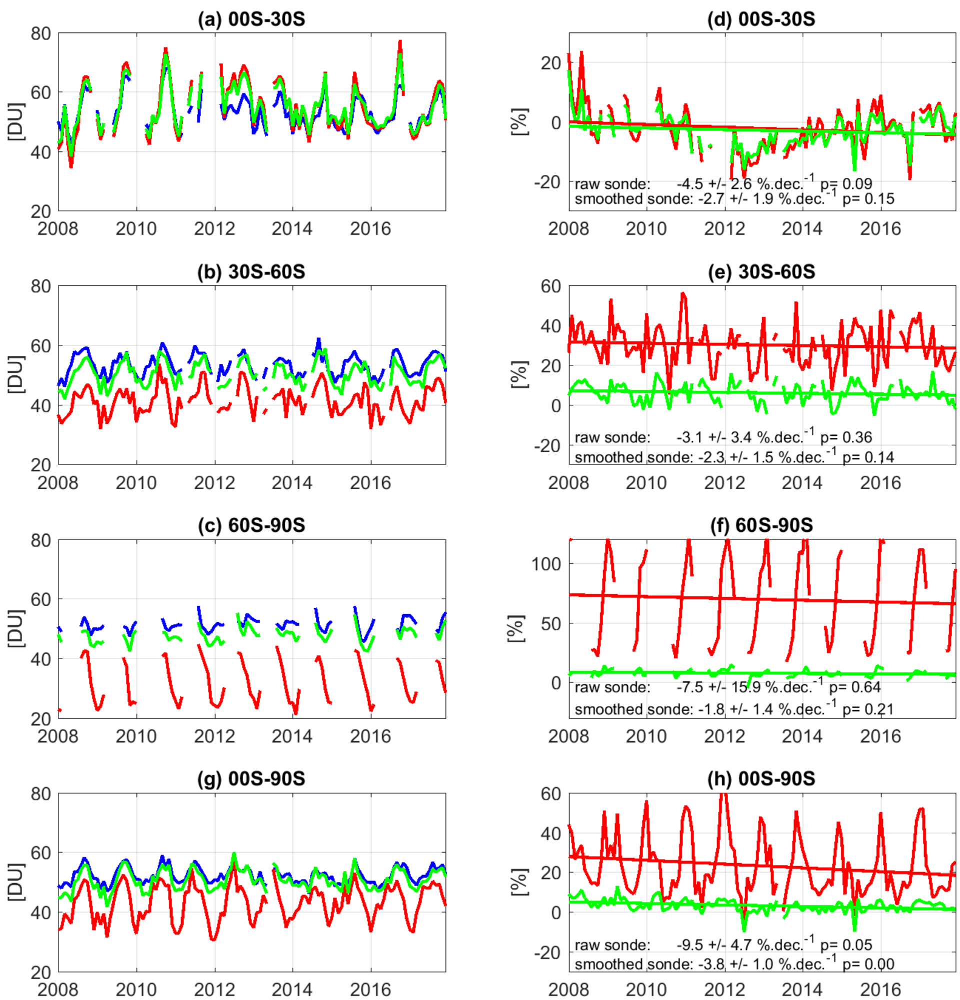

Figure 10. Same as figure 9 for SOFRID-O3 v1.6 in the Southern Hemisphere.

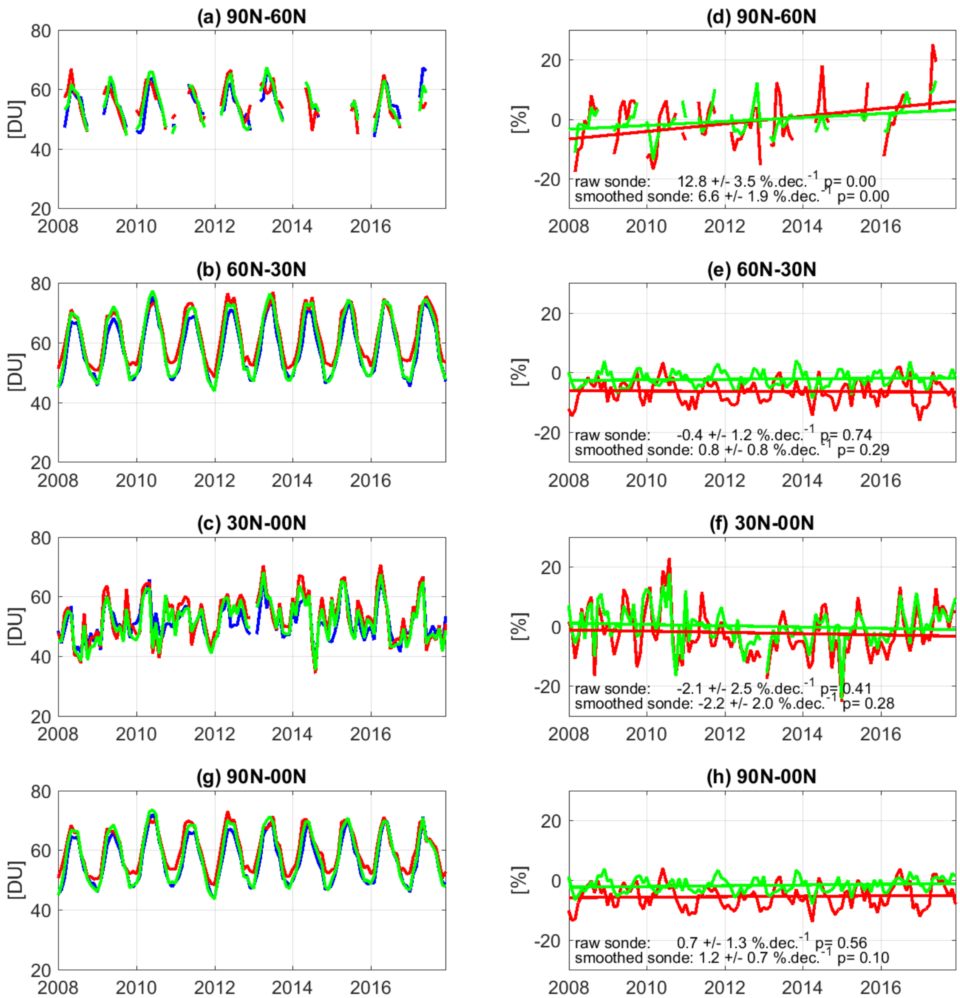

Figure 11. Same as figure 9 for SOFRID-O3 v3.5 in the Northern Hemisphere.

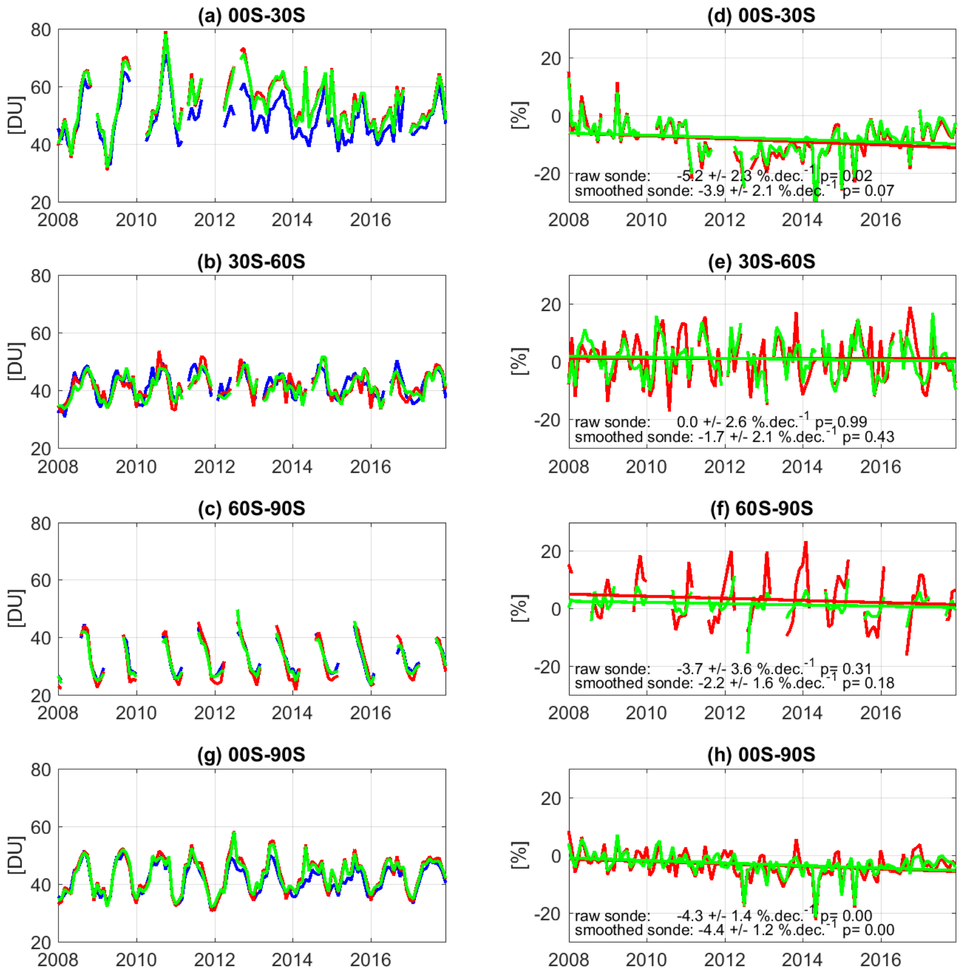

Figure 12. Same as figure 9 for SOFRID-O3 v3.5 in the Southern Hemisphere.

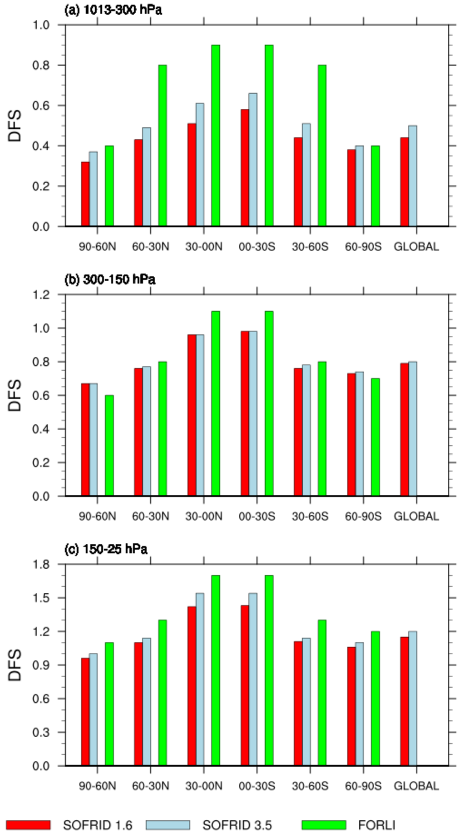

Figure 13. Degrees of Freedom for Signal (DFS) of IASI SOFRID-O3 v1.6 (red), SOFRID-O3 v3.5 (light blue) and FORLI-O3 (green) retrievals in the different latitude bands for the (top) 1013-300 hPa (middle) 300-150 hPa and (bottom) 150-25 hPa. FORLI data are taken from Boynard et al. (2018).

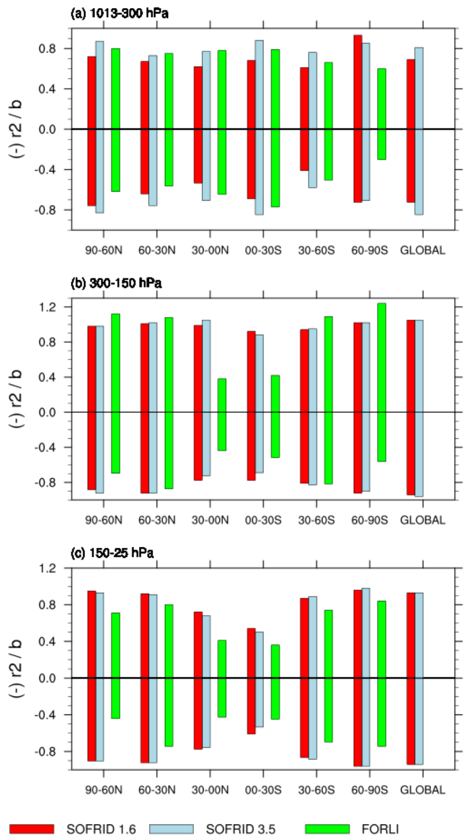

Figure 14. b: slopes of the linear regression (positive values) and (-) **r2**: correlation coefficients (negative values) between IASI retrievals and sonde data. (red) SOFRID-O3 v1.6, (light blue) SOFRID-O3 v3.5 and (green) FORLI-O3 (from Boynard et al. (2018))

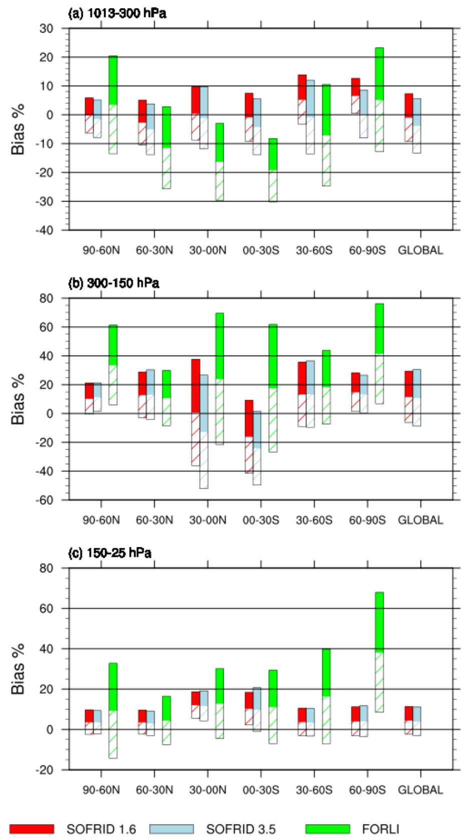

Figure 15. Biases and ±RMSDs (bars) of the differences between IASI retrievals and sonde data. (red) SOFRID-O3 v1.6, (light blue) SOFRID-O3 v3.5 and (green) FORLI-O3 (from Boynard et al. (2018))

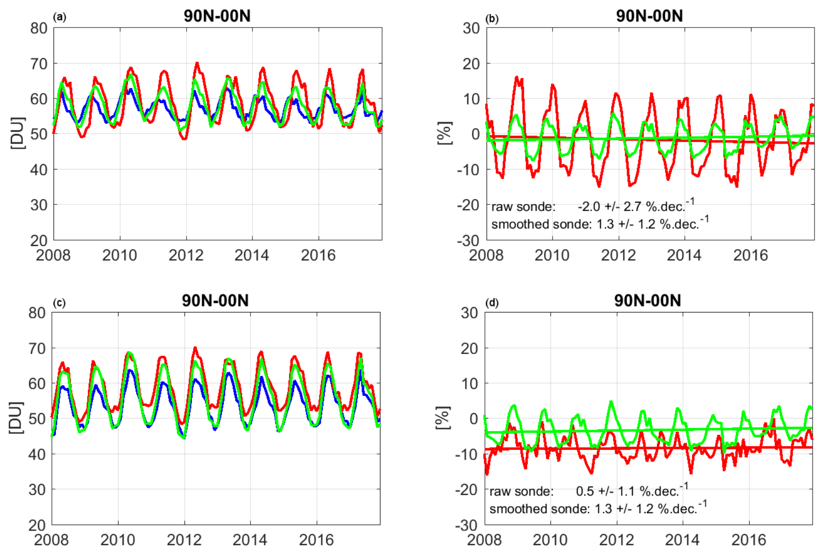

Figure 16. Time series of SOFRID-O3 (a) v1.6 and (c) v3.5 surface-300 hPa columns for the Northern Hemisphere (0-90°N). Blue lines for IASI retrievals, red lines for raw sonde data and green lines for smoothed sonde data. Differences between IASI and sonde data for (b) v1.6 and (d) v3.5. Red lines for raw sonde data and green lines for smoothed sonde data.