Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-5-RC2, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "A tropopause-based a priori for IASI-SOFRID Ozone retrievals: improvements and validation" *by* Brice Barret et al.

Anonymous Referee #2

Received and published: 6 April 2020

The manuscript by Barret et al. presents the comparison and the validation of two versions of the SOFRID retrieval algorithm developed for IASI (SOFRID-O3 v1.6 and v3.5), which differ in apriori: single vs dynamical (month, latitude and tropopause height dependent). This study shows considerable work regarding the backprocessing of the whole IASI dataset with two recent versions of SOFRID, which use ECMWF operational analyses for temperature and humidity, in contrast with the previous versions. The comparison between products based on the use of a single vs variable apriori is particularly interesting, as the choice of the apriori remains an important source of discrepancy between retrieval datasets. The topic is suitable for ACP. However, I do have several key comments that should be addressed before publication:

Printer-friendly version

Discussion paper

General comments:

1/ My major comment is related to the validation methodology used by the authors; I do 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. 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.

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.

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).

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. For instance:

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

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

- 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$\pm$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.

- 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.

- 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. 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... 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.

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.

Minor comments:

- 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.

- 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.

- 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.

Technical comments and typos:

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

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

- P.3, L.7: methodology -> methodology

- P.4, L.33: “The use OF a . . .”

- P.5, L.8: atmospheric -> atmospheric

- P.6, L.1: Th -> The

- P.7, L.20: one reference is missing here.

- P.7, L.9: below -> above

- P.7, L.21: elliminate -> eliminate

- P.8, L.23: variance -> ratio of the variance (?)

- Table 2: Units are missing

- P.9, L.21: tropospehric -> tropospheric

- P.9, L.27: UT -> UTLS

- 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.

- Error in the caption of Fig.9: “Same as Fig.9” -> “Same as Fig.8”

- Fig.8: The color legend should be indicated in the top panels.

- P.12, L.6: Which version of SOFRID are you referring to?

- 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

- P.13, L.1: delete “(b)” in the sentence. I don’t see that in Fig.13.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-5, 2020.

Printer-friendly version

Discussion paper