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.