Florence, 24/06/2020

We thank the reviewer for the very useful comments. In the following, we answer the specific comments (included in "boldface" for clarity) and, whenever required, we describe the related changes that will be implemented in the revised manuscript. Page and line numbers indicated refer to the original version of the paper published on AMTD.

1) Title of the paper: the current title is far too general, and it should be focused on the actual work presented on the paper. I strongly recommend clearly indicating the only species analyzed "atmospheric ozone profile", the only measurements used "Meteosat Third General and EUMETSAT Polar System Second Generation" and that is a sensitivity study based on simulations. I disagree indicating "Full Exploitation" since it is not the only way to exploit these measurements and "Copernicus Atmospheric Sentinel Level 2 Products" since only one species and 2 missions are concerned. By the way, Sentinel 4 and Sentinel 5 are the denomination of the UVNS sensors and not the ones operating in the infrared. The correct denotation is IRS onboard MTG-S and IASI-NG onboard EPS-SG. The whole manuscript should be corrected in that sense.

We agree that the title is too general so we propose a new one that is focused on the actual work: "***Application of the Complete Data Fusion to the ozone profiles measured by the Copernicus Atmospheric Sentinel missions: a feasibility study***" This new title aims also to address another possible source of misunderstanding that probably originates some of the referee comments. This study is focused on the CDF performances and the particular case and species addressed can be considered examples. In other words, the primary objective of this study is to show that, if the considered L2 products are profiles retrieved with optimal estimation techniques and if all the needed quantities are available, then the CDF is, in general, a valuable alternative.

The fact that the L2 products have been simulated in realistic conditions is an important supplementary benefit of the article that allows to quantitatively estimate the benefits of the CDF application, but in any case, is of secondary importance with respect to the feasibility study in general.

2) Introduction and scientific context of this work: this is one the major missing aspects in the paper. The instruction and the other section very rarely cite nor mention other related works on the subject different from previous works of the authors themselves. Although they do exist relatively abundantly, the authors do not mention any other approaches (except for the plain average) to use synergistically different measurements of the same or different sensors to derive ozone profiles. The authors should thoroughly provide a full exhaustive list of approaches of synergism of several measurements to derive ozone profiles and compare them (at least conceptually) to the proposed fusion approach.

Two new paragraphs have been added to the introduction mentioning and conceptually comparing different approaches for the synergistic use of the same or different sensors to derive ozone profiles and 5 new related references have been added to the bibliography.

3a) Explanation of the results of the approach: only the equations of the fusion approach are transcribed in the paper, without any physical explanations, and the obtained results are very superficially described, with very little explanation for understanding them. Although this is the heart of the paper, the reader cannot understand why the fusion approach allows an enhancement of vertical resolution or sensitivity to ozone from the synergism of several hundreds of profiles. This aspect should be thoroughly explained in physical terms and illustrations should be given.

Even if the presentation of the CDF is not the focus of the paper the fusion equations have been rearranged to be more readable and some comments have been added.

Regarding the presentation of the results, the section 3.2. "Single grid-box analysis (0.5ºx0.625º)" has been extended and a new figure has been added to study in finer detail the enhancement of vertical resolution. More explanations have been also added to section 3.3. "Statistical analysis for a large domain".

3b) Concerning the retrieval errors, many explanations are missing, and this should be at least compared to the error of the average within the main text, in each occasion. Another missing aspect is how the authors represent systematic errors and co-locations errors. This is important to know and be explained in this current paper, since the persistance of such errors partly differentiates the current approach from the arithmetic average.

One entire section of the paper (2.4. "Arithmetical average and biases") and one section of the supplementary material ("Fusion of 1000 pixels in coincidence") have been dedicated to explain why the authors do not consider the error of the average in the remaining part of the paper. Regarding the systematic errors, the authors agree that this is an important point that has to be deeply studied in a dedicated work, possibly dealing with the application of CDF to real data. However, this paper is based on synthetic data and the systematic errors have not been considered. This is explained in the introduction (L56-61) where it is also mentioned how systematic errors can be treated in the CDF context. The interpolation error is not applied here but all the details concerning it and the coincidence errors can be found in two dedicated papers reported in the bibliography (Ceccherini et al. 2018 and 2019).

4a) Datasets used in the paper: the explanation of the missions and simulated datasets is extremely brief. The reader does not understand what the differences between sensors are and why they provide certain DOFs or spatial coverage. The performance of the "S4:TIR" sensor does not seem to correspond to its instrumental characteristics. This sensor, which in reality is called IRS onboard MTG-S and not S4:TIR, has similar instrumental characteristics as IASI, with even a coarser spectral resolution and similar radiometric noise in the ozone band around 10 microns. The total column DOF for an ozone retrieval from IASI is typically 3 and at most 3.5. The current paper shows DOFs for "S4:TIR" greater than 5, which is not possible in my understanding. Unless thoroughly explained, justified and compared to IASI, all results considering "S4:TIR" simulated L2 products should be done again with proper instrumental characteristics.

A new paragraph has been added at the end of section 2.2 L2 Product Simulation with a more detailed explanation of the missions and of the simulated datasets, with some more references and with explicit mention of relevant details.

Regarding the DOFs of TIR:S4 (i.e., IRS/MTG-S), the paper by (Crevoiser et al., 2014) shows that the number of DOSF is 4.7, when the IASI-NG instrument configuration is IRS2b, which was used for this study. Moreover, for generation of the synthetic retrieval products we used climatological a priori information rather than state of the art ozone information with the result of applying a weaker constraint that further increases the DOFs.

Ref:: Crevoisier, C., Clerbaux, C., Guidard, V., Phulpin, T., Armante, R., Barret, B., Camy-Peyret, C., Chaboureau, J.-P., Coheur, P.-F., Crépeau, L., Dufour, G., Labonnote, L., Lavanant, L., Hadji-Lazaro, J., Herbin, H., Jacquinet-Husson, N., Payan, S., Péquignot, E., Pierangelo, C., Sellitto, P., and Stubenrauch, C.: Towards IASI-New Generation (IASI-NG): impact of improved spectral resolution and radiometric noise on the retrieval of thermodynamic, chemistry and climate variables, Atmos. Meas. Tech., 7, 4367–4385, https://doi.org/10.5194/amt-7-4367-2014, 2014.

4b) Moreover, the description of the atmospheric scenario should provide within this paper in much more detail (e.g. the resolution and particularities of pseudo-reality, sources, variability, etc).

The section 2.1. "Atmospheric scenario and ozone climatology" has been extended and in particular the quantities taken from these external databases have been put in relation with the other quantities defined in the equations that describe the simulation process. The details on the atmospheric scenario goes beyond the scope of this paper and can be found in the cited references.

5) Abstract: only tens or hundreds of measurements fall within tens of kilometers if satellite observations are finely resolved. This was not the case before Sentinel 5P for ozone retrievals and it is not the case for IASI-NG either.

"in the near future" added at the end of the sentence.

6) L23: One can also average or do the median of the datasets.

The mention of the use of simple averages has been added to the new abstract.

7) L45-51: This is not true. By reducing the horizontal resolution, we lose natural variability within the grid cell.

The premise "whenever the user does not need the full spatial and temporal resolution" was originally introduced to prevent this objection. The information content intended here is the one represented by the DOFs, but since at L45 they have not yet been defined in the paper we preferred this more general formulation. Nevertheless, since the sentence is not strictly necessary we removed it.

8) Sections should be number in order to cite them.

Section numbering added.

9) L105-108: this sentence is not very clear. Please reformulate it clearly defining S and Stotal

The definitions of S and Stotal have been separated in two distinct paragraphs.

10) L114: "The above formulation was used to simulate ozone profiles in the two spectral bands (UV and TIR) for both S4 and S5" please reformulate. Ozone profile are retrieved using measurements from a spectral band.

In this study ozone profiles have not been retrieved using simulated spectral measurements. Added "In this study".

11) L120: However, these are not true retrieval from an iterative numerical procedure. How this formulation compares to true retrievals as those from true measurements? This aspect should be clarified and illustrated. Are S and Stotal consistent with those from true full retrievals?

This method uses a linear approximation (Eq. (1)) of the relationship between the retrieved profile and the true profile in the optimal estimation method (see Rodgers, 2000). Therefore, the validity of this method is within the correctness of this approximation. A sentence clarifying this aspect is added.

12) L132: why 5%? This should be justified.

The value of 5% is a reasonable value. However, the study presented in (Ceccherini et al. 2019) clearly show that even if the coincidence error is strictly needed for the correct behaviour of the CDF product, this is not strongly dependent on its exact amount. Therefore, this choice does not affect the conclusions of the study. A paragraph has been added in the paper.

13) L139: the notion of "good" or bad is subjective. This cannot be expressed in such a why, but in objective terms (reduction of errors, bias, sensitivity, etc). Please, reformulate.

Removed "good".

14) L173: when embedded in the text, please use the word Figure and not Fig.

OK.

15) Error of fuse profile: it is not as low as 1 over the square root of the number of measurements as it would be for random errors and the arithmetic average but only around -30%. Comment thoroughly and explain.

Explanation added in section 2.3 right after eqs 6. This aspect is also studied in the supplementary material commenting Figure S1, right panel.

16) Fig.3: It should be clearly written in the caption of the figure and the text that the AK of the "fuse profile" comes from the fusion of 118 profiles and the other AK are for single measurements.

Sentence added in the caption.

17) Fig. 3: A full description of the 4 instruments and their characteristics should be given.

We have expanded the instrument description, see also answer to point 4, above.

19) Fig. 3: Why only 4 curves are displayed instead of 5 AKs (4 sensors + fusion result)?

As explained in the beginning of the section, no S5-UV1 products falls in the considered cell so only 4 curves are displayed. This choice also aims to simplify as far as possible the understanding of the discussion and the figures. The discussion of a case with 4+1 curves has been added to the supplementary material in the section "Single grid-box analysis (1ºx1º)".

20a) Fig. 3: Results not explained: DOF of 9.5 how do you explain this in physical terms? Where does it come from?

The CDF acts removing the a priori information of the L2 fusing measurements and adding an a priori information (independent of that of the input measurements) in the fusion process. This characteristic of the CDF allows to increase the relative weight of the information coming from the measurements with respect to the information coming from the a priori, as a consequence we obtain an increase of the number of DOFs. This mechanism does not occur in the arithmetic average.

The section 3.2. "Single grid-box analysis (0.5ºx0.625º)" has been extended and a new figure has been added to study in finer detail the enhancement of vertical resolution observing the behaviour of the AK matrices rows.

20b) Having only a few S5-derived profiles and more than a hundred of S4-derived, what is the influence of having an asymmetric number of profiles from one or the other instruments?

This is really an interesting point that cannot be explained in few words. A section has been added to the supplementary material entitled "Single grid-box analysis (1ºx1º)" that accounts for both this aspect and the effect of enlarging the coincidence cell size. This analysis has not been included in the paper since we believe that it is too detailed for a feasibility study based on simulated products. On the other hand, we agree with the reviewer that these aspects are crucial and worth a throughout investigation in the case of the application of the CDF to real measurements.

21) Large domain section: the title indicates 0.5◦ x 0.625◦ which is not large.

The indication of the cell size in the title was misleading so it has been removed.

22) L193-L195: The sentence is not clear. Please reformulate. What is fusion grid-boxes?

Sentence modified. See also previous point (the misleading indication of grid box size in the title could also contribute to the confusion).

23) L200: The explanation of variables should also provided in the caption of table 2

Caption modified.

24) Please use Table and not Tab. Same Figure and not Fig. in the caption.

OK

25) L195-198: What happened to S5-TIR? Why it is not here?

L195-198 do not list all the types of product; S5-TIR is simply not mentioned here because not relevant in the discussion.

26) Table 2: this nomenclature is not clear S4:TIR+UV1_S5:TIR+UV1. Please reformulate here and elsewhere. What is the meaning of ":" and "_" in a name. They should be avoided in the names and the true names for the TIR sensors should be used.

The nomenclature of table 2 is the same used in the figures. The use of the real name of the sensors is too long for the legends in the figures, therefore, we prefer to maintain that nomenclature that, although not precise, is coherently used throughout the text. On the other hand, that nomenclature is explained in the "Description" column of table 2 itself.

27) Figure 4: the fraction of clear sky measurements seems very reduced. Although they exist in reality, no measurements with a small cloud fraction are considered? This should be clearly stated.

The section 3.1. "Fusion in realistic spatial and temporal resolution conditions: the L2 Datasets" has been expanded to explain better these aspects.

28) Why S5UV1 are only available over Northern Africa? Why we do not have S5 pixels near Greenland? This should be explained thoroughly in this paper since is a major dataset of the paper and not referred to previous papers.

The section 3.1. "Fusion in realistic spatial and temporal resolution conditions: the L2 Datasets" has been expanded to explain better these aspects. In particular the products considered in this paper have been simulated considering the daylight, clear sky pixels belonging to one particular hour and one particular orbit that do not refer to clear-sky daylight pixels in Northern Africa and Greenland.

29) L217 and L230: Why do you justify again avoiding the use of averaging without any quantified and clear statement. This method of arithmetic average should be explicitly included in the comparison every time and compared in terms of error and performance.

The justification for not using the arithmetic average is explained both in formulas (section 2.4) both with a numerical example (supplementary material, section entitled "Fusion of 1000 pixels in coincidence").

30) L220: What is the meaning of pure number? Without units?

Yes, because it is a relative quantifier expressed by the ratio of two quantities with the same units.

31) Fig.5 shows that SF DOF is at most 1.9 and this seems to be the case for Fig. 3. This means that the example of Fig.3 is not a typical case with SF DOF around 1.5 but the maximum performance of the fusion. The choice of this example should be reviewed, and typical case should be taken, but not the best.

The SF DOF of the case represented in Figure3 is 1.77 and not the maximum value 1.9. Therefore, this case does not correspond to the best performance, and it was chosen because it represents well the phenomena that are discussed in the paper. The value of 1.77 has not to be compared with the global average of SF DOF but with the SF DOFs obtained in cells with a similar number of fused L2 products. The case of Figure 5 has been chosen also because it is relatively easy to be explained. A more complicated case has been added in the supplementary material section entitled "Single grid-box analysis (1^ox1^o)". On the other hand, figure 3 and 4 give an overall picture about fusion performances; a new discussion and a new figure in section 3.2 and the new final paragraph of section 3.3 give an insight on the less than one values of SF AK in the troposphere and in the middle upper atmosphere that is also the origin of the cluster of green points in Figure 5. We think that this new scenario represents a good compromise in which all the relevant aspects are addressed.

32) I do not understand why the best performance is found for the use of the two S4 products, since the performance of the TIR sensor of S4 is not the best, as compared to that on S5.

The performance here is evaluated in relative terms (all SFs are ratios) so the better performance of CDF does not necessarily take place when the quality of the L2 product is higher. On the contrary the best performances of the fusion in terms of Synergy Factors are obtained when many products with comparable quality are fused. See also the new section entitled "Single grid-box analysis (1^ox1^o)" in the supplementary material for a case dealing with L2 products with different quality.

33) Fig.5 caption: it should be explained that is the product of combining a large number of measurements. I recommend not to use N but Number in the axis label.

Figure 5 represent the SF DOFs for all the 1979 fused products. The term N has been used consistently throughout the text, figures and tables of the article to indicate the number of cells or products so we prefer to maintain the original nomenclature in this figure.

34) Figure 6: What is the link between SF for AK and for Error? It seems that large SF AK correspond to smaller SF Error and vice versa (looking at green and purple dots). This should be explained and clearly quantified.

We do not think that this phenomenon can be generalized: for example, this does not seem to happen at lower altitudes. As shown in the paper the behaviour of the AK matrices is not intuitive and deserves a detailed study to be explained (see point 31, above and new section 3.2 in the paper).

35) Table 3: it seems strange that no results of the 1x1 ◦ cells are provided in the text. This should be commented and at least provided in terms of a table and compared to the smaller cells.

Since the 1x1 case does not introduce significant new features with respect to the smaller cell, we decided to briefly cite it in the text and to document it in the supplementary material.

36) L293-295: I do not think that this is not true. The fusing products come from L2 products; they are intrinsically dependent in a priori information of these products.

The CDF removes (by means of AK matrices) the original a priori information from L2 products before to combine them so the fused product is effectively independent from these a priori. In other words, the a priori information of the fused product can be chosen independently from the ones of the L2 products.