Response to RC1

COMMENT #1: First, in the "Results and discussion", starting from line 186 and into the conclusions, the authors assume that a DFS increase automatically implies an increased vertical resolution of the retrieval. This is not true, as a DFS change can also be induced by a change of the weight of the measurement (or, equivalently, of the prior information) in the retrieval, as represented by the averaging kernel matrix' row sums, or by a change in the retrieval height offset (vertical shift of the measurement weight). Both are independent of the retrieval's vertical resolution, which e.g. can be determined from a kernel's FWHM. It is very important that the authors consider these alternatives based on a thorough discussion of (fused) averaging kernel matrix behaviour and either modify their discussion and conclusions accordingly or demonstrate that only the vertical resolution is impacted by the method.

To answer this comment, we take as reference the case presented in the paragraph entitled "**Single grid-box analysis (0,5ºx0,625º)**" rows 164-190, Figures 1, 2 and 3. In the right panel of Fig. R1 (new version of Fig. 3 in the original manuscript) the vertical resolution profile of the FUS product is compared with the vertical resolution profiles of the 118 L2 fusing products in the considered grid box cell, where the vertical resolution is calculated starting from AK matrices according the Full Width Half Height approach (Rodgers, 2000) and, in particular, using the algorithm defined in (Ridolfi and Sgheri, 2009). According to this algorithm the vertical resolution at a given vertical level is calculated as the ratio of the area under curve defined by the module of the AKM row whose diagonal value (i.e. the value that lies on the AKM diagonal) correspond to the considered level and the diagonal value of the same row; according to this particular formulation the presence of secondary lobes of the AKM row degrades the vertical resolution. In the figure, we can see that, even if the vertical resolution of the FUS product is almost everywhere improved with respect to the L2 fusing products, this does not happen in the range 30-45 km.

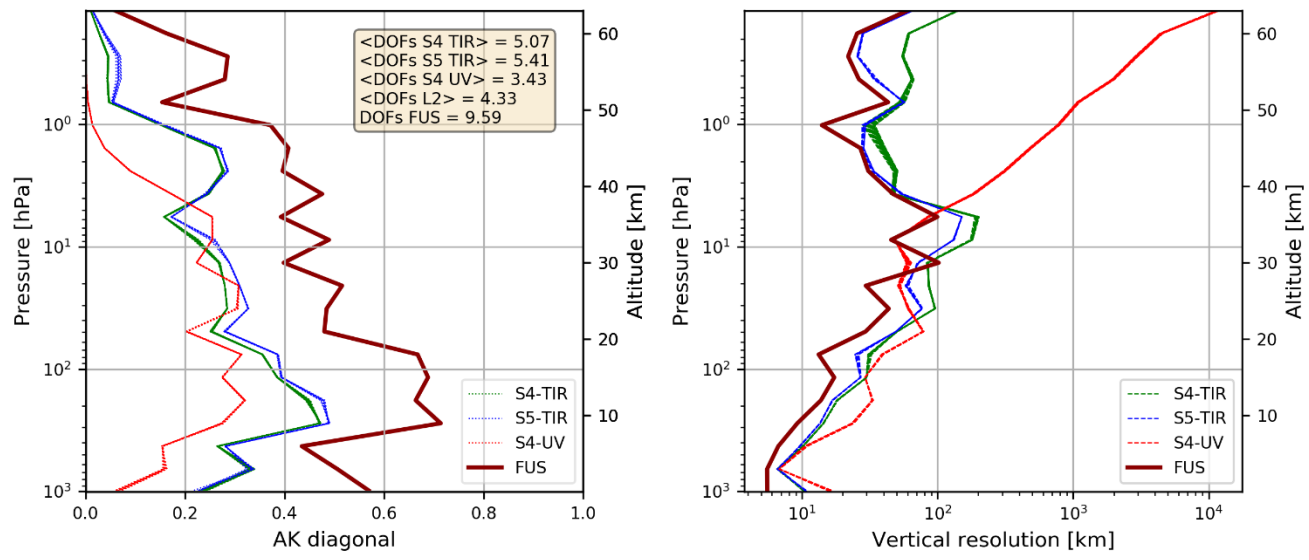

Fig. R1 (Fig3 new): Left panel: AKs diagonal of: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). In the text box, the average number of DOFs for each type of L2 product, the average number of DOFs for all L2 products and the number of DOFs of the FUS product are reported. Right panel: Vertical resolution (FWHH) profiles of: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line).

To better understand the effect of the fusion on the AK matrices is useful to analyse the behaviour of their individual rows. In Fig. R2 two rows are represented, one that refers to the troposphere (left panel, 6 km) one to the middle stratosphere (right panel, 39 km), where the indicated reference altitude is the one corresponding to the value of the row that lies on the diagonal of the AK matrix. The value of the vertical resolution at the considered altitude is reported in the legend while the diagonal

value of each row is evidenced in the graphs with cross (L2) or dot (FUS) markers. At lower altitudes (left panel), as suggested by the reviewer, the DOFs increase can be attributed to three distinct phenomena where the first is the shrinkage of the main FUS AK lobe and the consequent improvement (of more than 30%) of the vertical resolution with respect to L2 products. The second phenomenon is linked to the fact that while for the FUS product the maximum value of the AK row corresponds to its diagonal element, for the L2 products these maxima are shifted with respect to the reference altitudes. The last phenomenon is a stronger contribution of the measurements with respect to the a priori in the FUS product, where the latter effect can be evidenced considering the sum of all the elements of the rows that has 0.913 as maximum value for the L2 products and is 0.956 for the FUS product. In this particular case, all these three effects go in a direction that can be regarded as a benefit of CDF application. The results at higher altitudes (39 km, right panel) are primarily influenced by the shape of the AK rows that exhibit large secondary lobes degrading the vertical resolution. Both figures and their analysis will be reported in the reviewed paper.

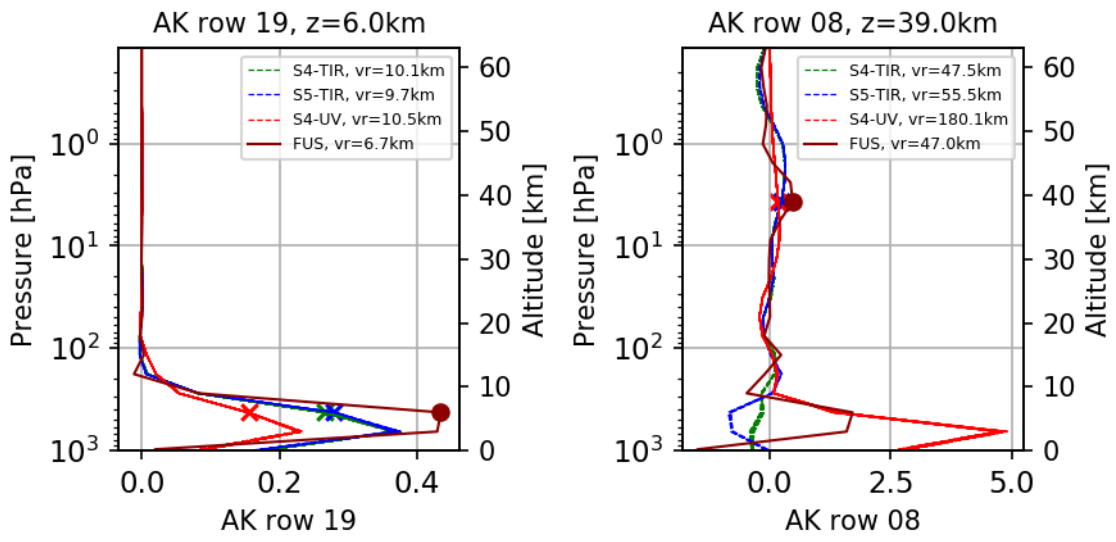

Fig. R2 (Fig.4 new): Rows of AK matrices at 6 km (Left panel) and at 39 km (Right panel) for: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). The value of the vertical resolution at the reference altitude is reported in the legend (the minimum of vertical resolution at the considered vertical level for each type of L2 product) while the diagonal value of each row is evidenced in the graphs with cross (L2) and dot markers (FUS).

COMMENT #2 The averaging of profiles as a simple combination technique is not mentioned in the abstract nor the introduction, while it is being discussed in a separate section eventually. Mentioning a discussion of the Complete Data Fusion method with respect to simple averaging from the beginning would improve readability.

Mentions will be added in abstract and introduction.

COMMENT #3 Stating "the quality of the products improving with larger grid boxes" in the abstract is misleading, as representativeness errors are also expected to increase with the grid box size. The latter is only briefly mentioned in the very last sentence before the conclusions. This important point deserves more discussion (especially regarding the differences in natural variability for different atmospheric molecules) and mentioning in the abstract.

The sentence "the quality of the products improving with larger grid boxes" will be deleted from the paper. The point of representativeness errors will be more mentioned and discussed even if a dedicated study is needed to analyse the problem.

COMMENT #4 Line 45: "whenever the user does not need the full spatial and temporal resolution" sounds misleading, as it seems that the Complete Data Fusion method can also be used to combine measurements from several satellites choosing e.g. the pixel size of one of the contributing instruments?

In the considered examples the coincidence grid cell is much larger than the footprint of the L2 fusing products that have homogeneous pixel sizes. On the other hand, when the CDF is applied to L2 products with very different horizontal footprint size, the largest pixel footprint can be chosen as coincidence grid cell, so we do not think that the cited sentence is misleading. Nevertheless, if the reviewer believes this may lead to misunderstandings, we will consider the option to modify or to remove the sentence.

COMMENT #5 Lines 75-77: The specific differences and usages of the ozone climatology and the atmospheric scenario might be far from clear for some readers. It would be very helpful to extend on this and relate this information with the quantities in the equations.

References to the equations added.

COMMENT #6 The sentence (paragraph) on line 120 is very unclear (or trivial) and fully without context.

Sentence removed.

COMMENT #7 Lines 132-133: "the diagonal elements of S_coinc are calculated as the 5% of profile estimates in the ozone climatology" is not clear at all.

"When CDF is applied to not perfectly coincident products, the diagonal elements of $\mathbf{S}_{coinc}$ are calculated as the square of the 5% of the a priori profile $\boldsymbol{x}_a$."

COMMENT #8 Lines 135-137: It is agreed that interpolation errors do not apply in this study, but as a demonstration "for a Full Exploitation of Copernicus Atmospheric Sentinel Level 2 Products" (title) the implications of the need for a preceding interpolation for upcoming reality should nevertheless be decently discussed.

The following discussion will be added to the reviewed paper. "The formulas of Eqs.(6) refer to the case of measurements made on the same vertical grid. In general, also an interpolation error may be needed considering that the retrievals of the products to be fused can be furnished on different vertical grids. In (Ceccherini et al. 2018) the general expressions of CDF in the case of the fusion of products characterized by different vertical grids is presented and discussed together with the expression of the interpolation error that depends on the involved grids and on the AK matrices of the fusing products. However, since the interpolation error does not apply to the present study (the L2 products have been simulated on the same vertical grid) it has not been considered in Eqs. (6) and in the following discussion."

COMMENT #9 - Line 153: The authors could refer to van Clarmann and Glatthor, 2019, for possibly improving their discussion on the averaging kernel matrix of an average product.

The reference will be added.

COMMENT #10 - Lines 158-161: This paragraph is way too short for a proper understanding of how the simulated retrievals are spatiotemporally allocated. How are instrument and orbit characteristics taken into account and applied to a past atmospheric scenario?

The L2 datasets have been generated according to the equations (1)-(5) described in paragraph 2.2. The details of the simulation process (how are instrument and orbit characteristics taken into account) can be explored in the technical note (Technical Note On L2 Data Simulations, 35 pp., https://www.spacehatch.eu/result/616192, 2017) considering that here we simulated all the

pixels corresponding to a clear sky line of sight in the atmospheric scenario without applying any additional selection criteria. In fact, in the AURORA project 4 months of data have been considered, but a subset of the clear sky pixels has been simulated to reduce the computational cost of the simulations. For this study, all the clear sky pixels in the considered hour of data have been simulated, without additional filtering, choosing the reference time interval so that S4-S5 coincidences occur; the spatial distribution of the simulated products is indirectly represented in the left panel of Fig.5 that reports the horizontal distribution of FUS types (see Tab. 2). This paragraph will be added in the reviewed paper.

COMMENT #11 - Figure 2 and comparisons in supplementary material: Although absolute differences have their benefits, (additionally) showing relative differences–here with respect to the known 'truth' – would be insightful in the tropospheric region, i.e. to see the Complete Data Fusion performance where absolute values are small. Next to the supplementary material, why not add the average in Figure 2 as well?

A supplementary panel will be added to Fig.2, showing relative differences (see Fig.R3).

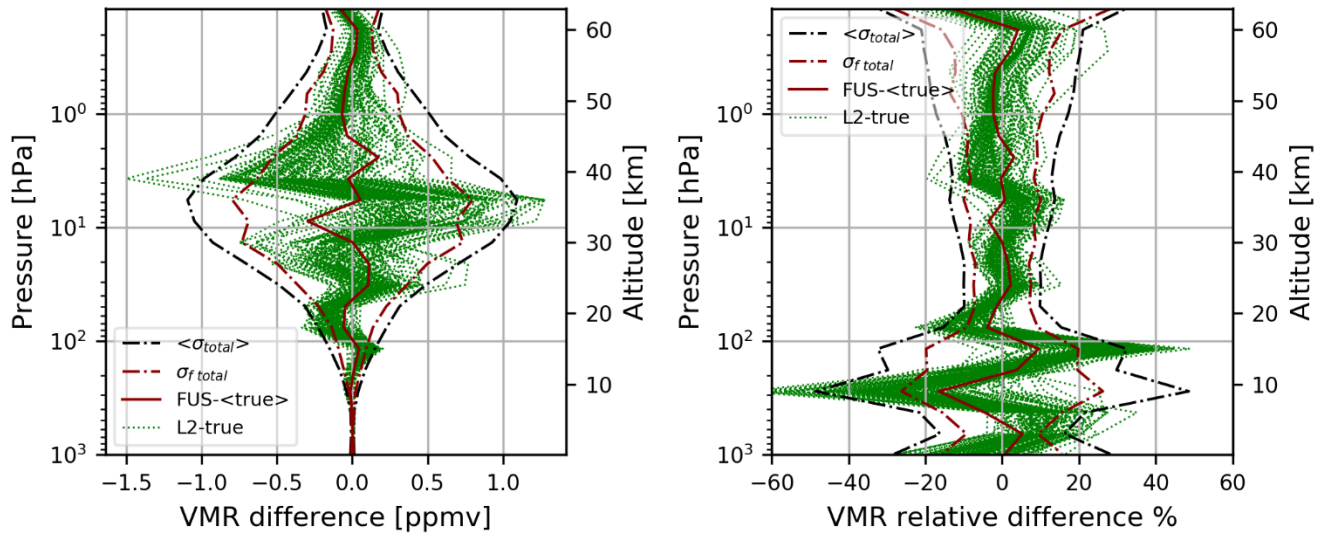

Fig.R3 (Fig2 new): (Left panel): absolute differences between L2 profiles and their true profiles (green lines), absolute difference between the fused profile and the average of the true profiles (dark red continuous line), the average of σ_{total} of L2 simulations (black dash-dotted lines), $\sigma_{f\ total}$ (dark red dash-dotted lines). (Right panel): relative percentage differences between L2 profiles and their true profiles (green lines), relative percentage difference between the fused profile and the average of the true profiles (dark red continuous line), the average of σ_{total} of L2 simulations normalized with respect to the true profile and expressed in percentage (black dash-dotted lines), $\sigma_{f\ total}$ normalized with respect to the true profile and expressed in percentage (dark red dash-dotted lines).

We prefer not to add the average in Fig.2 because too many lines can confuse the reader. On the other hand, we have motivated why we do not consider the average in the paragraph "Arithmetical average and biases" of the paper and in the paragraph entitled "Fusion of 1000 pixels in coincidence" in the supplementary material with particular reference to the right panel of **Fig. S1**.

COMMENT #12 The histogram in Figure 4 seems to go up to 160 only, while Table 2 mentions 165 as a maximum in its latest column. Or the values in the histogram above 160 too low to be observed, or is there a different reason for this discrepancy?

There was an error in the code generating the histogram so that the bars stopped at 160 even if there were data between 160 and 165. Graph corrected.

COMMENT #13 Lines 218 and following: Aires et al. (2012) have introduced a 'synergy factor' (not 'synergic factor' as stated in this work) for the errors only, while the authors have extended this concept to other quantities as well, including what they misleadingly call vertical resolution in Eq. (10) (see general comments). The authors should better explain the rationale behind the extension of the synergy factor to other quantities.

Corrected "synergic" with "synergy".

After the reference to Aires et al., the following paragraph has been added: "Although Aires introduces SF only for errors (Eq. (11)), we extended here his definition also for other quantities because they constitute a useful tool to synthetically represent the performances of fusion algorithms.

We modified the comment to Eq.10 as follows: "A value of *SF AK* larger than 1.0 at a specific vertical level (indicated by the index *l*) means that, at that level, the diagonal value of the AK matrix of the FUS product is higher than that of all the individual products. As we have seen commenting Fig.3 (new) and Fig.4 (new), the increase of the AK diagonal values at a specific level can happen for different reasons but all of them can be considered as an improvement in the product quality."

COMMENT #14 Lines 261-262: "This is likely caused by the coincidence errors that have to be added in the fusion process" is a too brief and unsatisfactory explanation. With full control over the simulation and its errors, this should be quantitatively examined. Moreover, only the middle to upper stratosphere is mentioned, while Figure 6 clearly has values below one in the lower troposphere as well.

This aspect has been studied in more detail and explained as follows.

Some SF AK values, both in the troposphere and in middle upper stratosphere, are smaller than one: in the troposphere, this happens in 20 cells out of 1939 while in the middle upper stratosphere this happens in almost 500 cells. In both cases, this is caused by two main simultaneous reasons: the first one (and the easier to explain) is the introduction of the coincidence VCM, which degrades the quality of the AKM. This effect is represented in Fig. R4 where the profiles of SF AK are represented for a single cell in which 56 L2 products have been fused, considering the fusion both with the coincidence error added and without. It can be noted that the introduction of the coincidence error provokes a SF AK values smaller than one in the troposphere.

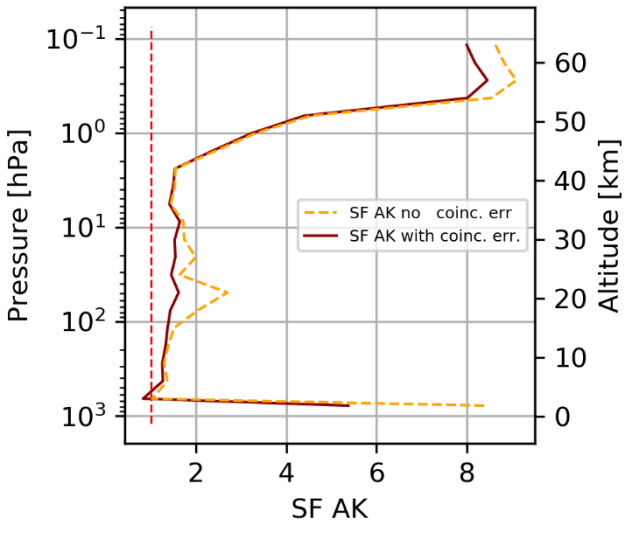

Fig. R4: profiles of SF AK for a single cell in which 56 L2 products have been fused, considering the fusion both with the coincidence error added (dark red line) and without (orange dashed line).

The second reason for these less than one SF AK values derives from the application of the CDF to cases in which one of the fusing products has the peak of the AKM row that is much larger and closer to the nominal vertical level than all the others. In Fig. R5, we represent the rows of the AK matrices referring to an altitude of 3 km for a cell in which two L2 products in perfect coincidence (S4-TIR and S4-UV) have been fused.

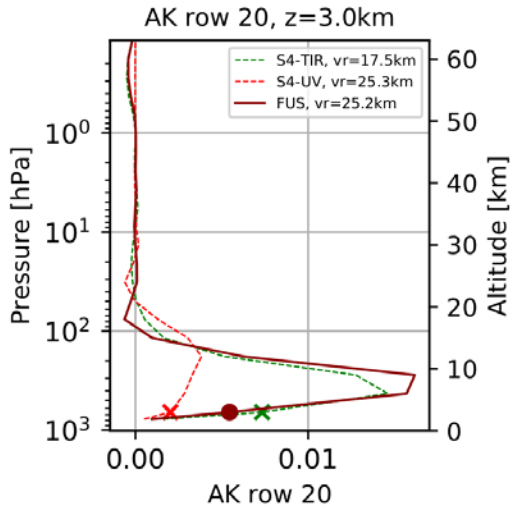

Fig. R5: rows of the AK matrices referring to an altitude of 3 km for a cell in which two L2 products have been fused; FUS product (dark red line), S4-TIR product (red dashed line), S4-UV product (green dashed line).

It can be noted that the peaks of the rows of the fusing products happen at 12 km (S4-UV, red line) and 6 km (S4-TIR, green line) while the peak of the FUS product row is located at 9 km. Even if the peak of the FUS row is the largest one, since it is more distant from 3 km with respect to the S4-TIR peak, the diagonal value of the row of the FUS product (dark red dot) is smaller than the one of the S4-TIR product (green cross) so that the SF AK is less than 1 at 3km. A summary of this discussion will be reported in the reviewed paper.

COMMENT #15 - Line 294: Level 3 products are often provided on monthly timescales. It would be insightful to include a note on the use and (representativeness) effects of Complete Data Fusion in a large temporal domain.

We have not yet investigated thoroughly the application of CDF to long time averages and we can deepen this aspect only to a limited extent. The application of CDF requires the introduction of a coincidence VC matrix that is needed to manage the variability of the fusing products and that consequently depends from the size in time and space of the considered coincidence grid cell. The fused product can be considered as an estimate of the average of the "true" quantity in the coincidence grid cell. When a monthly timescale or an entire latitude band are considered, since a big number of L2 products are fused together, the contribution of the a priori will be less and less important and the choice of the "right" amount (not too much and not too little) and the "shape" of the coincident error have to be studied in depth in a dedicated study. A sentence about this issue will be added in the conclusions of the reviewed paper.

COMMENT #16 - A single paragraph abstract would improve readability. Also, throughout the text and conclusions, very often very short paragraphs are used. Several of these could be combined for clarity.

In the reviewed paper, the abstract will be formatted as a single paragraph and the paragraphs combined throughout the text.

COMMENT #17 The authors seem to have somewhat exaggerated in their self-referencing: Ceccherini et al., 2003 and Ceccherini et al., 2010 do not seem to be required upon using Rodgers, 2000 already. Moreover, Ceccherini et al., 2014 is listed in the references, but not in the text. Finally, Kroon et al., 2011 is not required after Liu et al., 2010 (line 84).
Ceccherini et al., 2003, Ceccherini et al., 2010, Ceccherini et al., 2014 and Kroon et al., 2011 removed.

COMMENT #18 - Line 23: "and is therefore justifiable only as a temporary solution" is a user decision and irrelevant for this work.
Eliminated.

COMMENT #19 - - Line 24: It looks as if "while" is missing between "dataset" and "limiting"?
Corrected.

COMMENT #20 - - Lines 29-30: What errors (or uncertainties) are referred to here?
Total errors, one of the sentences will be removed.

COMMENT #21 - -Introduction, first sentence: The Copernicus program contains more than only the Sentinel missions, so the provided web link should be at the end of the sentence, possibly by the introduction of a second link for the Copernicus program.
Corrected.

COMMENT #22 - The program moreover is an initiative of the European Commission, not of the European Union.
Corrected.

COMMENT #23 -- Line 64: Remove "with each other".
Removed

COMMENT #24 -- Using sub-numbering (a, b, c...) in Eq. (6) would be helpful. Providing alpha_i and S-tilde_i (two last equations) before the four others with some additional clarification could help very much in understanding the Complete Data Fusion setup.
Equations reordered. Sub numbering added.

COMMENT #25 -- Figure 2 and comparisons in supplementary material: What the authors call "soliddot" and "dashed" lines actually both refer to "dash-dotted" lines as they are typically called.
Corrected.

COMMENT #26 -Line 192: What is the latitude-longitude range of the large domain?
The spatial distribution of the simulated products is represented in the left panel of Fig. 5 that reports the horizontal distribution of FUS types (see Tab.2).

COMMENT #27 - Repeating the single grid domain in the section title is misleading here.
Removed.

COMMENT #28 - - The averaging kernel matrix in the denominator of Eq. (9) should have index i,ll instead of f,ll.

Corrected.

COMMENT #29 - Line 297: The availability of the climatology and MERRA2 data should be mentioned as well.

Data availability added.

COMMENT #30 - Equations, figures, tables, and (lack of) section numbering do not (yet) follow AMT(D) guidelines.

Section numbering added. Figures and tables moved at the end of the paper. Captions formatted according to AMT(D) styles.

Response to RC2

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 (0.5^ox0.625^o)" 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 (1^ox1^o)".

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 (0.5^ox0.625^o)" 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 (1^ox1^o)" 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^{o}x1^{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.

~~The Complete Data Fusion for a Full Exploitation of Copernicus Atmospheric Sentinel Level 2 Products~~Application of the Complete Data Fusion to the ozone profiles measured by the Copernicus Atmospheric Sentinel missions: a feasibility study

Nicola Zoppetti[1], Simone Ceccherini[1], Bruno Carli[1], Samuele Del Bianco[1], Marco Gai[1], Cecilia Tirelli[1], Flavio Barbara[1], Rossana Dragani[2], Antti Arola[3], Jukka Kujanpää[4], Jacob C.A. van Peet[5,6], Ronald van der A[5] and Ugo Cortesi[1]

[1] Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy
[2] European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK
[3] Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, P.O.Box 1627, 70211 Kuopio, Finland
[4] Finnish Meteorological Institute, Space and Earth Observation Centre, P.O. Box 503, FI-00101 Helsinki, Finland
[5] Royal Netherlands Meteorological Institute, Utrechtseweg 297, 3731 GA De Bilt, The Netherlands
[6] Vrije Universiteit Amsterdam, Department of Earth Sciences, Amsterdam, The Netherlands

Correspondence to: Nicola Zoppetti (N.Zoppetti@ifac.cnr.it)

Abstract. The new platforms for Earth observation from space are characterized by measurements made with great spatial and temporal resolution. While this abundance of information makes it possible to detect and study localized phenomena, on the other hand it may be difficult to manage this large amount of data in the study of global and ~~large scale~~large-scale phenomena. A particularly significant example is the use by assimilation systems of ~~level~~ Level 2 products that represent gas profiles in the atmosphere. The models on which assimilation systems are based are discretized on spatial grids with horizontal dimensions of the order of tens of kilometres in which tens or hundreds of measurements may fall in the near future.

A simple procedure to overcome this problem is to extract a subset of the original measurements~~. However,~~ but this ~~procedure~~ involves a loss of information~~, and is therefore justifiable only as a temporary solution.~~ another option is the use of simple averages of the profiles but also this approach has some limitations that will be discussed in the paper. A more refined solution is to resort to the so-called fusion algorithms, capable of compressing the size of the dataset while limiting the information loss. A novel data fusion method, the Complete Data Fusion, was recently developed to merge a-posteriori a set of retrieved products in a single product. In the present paper, the Complete Data Fusion method is applied to ozone profile measurements simulated in the thermal infrared and ultraviolet bands, in a realistic scenario, according to the specifications of the Sentinel 4 and 5 missions of the Copernicus programme. Then the fused products are compared with the input profiles; comparisons show that the output products of data fusion have in general smaller total errors and higher information contents. ~~The most significant improvement is an increased vertical resolution together with a reduction of the errors.~~ The comparisons of the fused with the fusing products are presented both at single fusion grid-box scale and with a statistical analysis~~. T~~ and the grid box size impact was also evaluated, showing that the Complete Data Fusion method can be used with ~~a wide range of~~different grid-box size~~s, the quality of the products improving with larger grid boxes~~ even if this possibility is strictly connected to the natural variability of the considered atmospheric molecule.

1. Introduction

In the context of the Copernicus programme (https://www.copernicus.eu)~~(https://sentinel.esa.int/web/sentinel/missions)~~ of the European ~~Union~~Commission, the European Space Agency is responsible for the Space Component consisting of a novel set of Earth Observation (EO) satellite missions for environmental monitoring applications: the Sentinels

(https://sentinel.esa.int/web/sentinel/missions). Each mission focuses on a specific aspect of EO. In particular, the geostationary mission Sentinel-4 (S4) and the two Low Earth Orbit missions (Sentinel-5p and Sentinel 5 (S5)), referred to as the atmospheric Sentinels, are dedicated to monitoring air quality, stratospheric ozone, ultraviolet surface radiation and climate. The atmospheric Sentinels will provide an enormous amount of data with unprecedented accuracy and spatio-temporal resolution. In this scenario, a central challenge is to enable a generic data user (for example, an assimilation system) to exploit such a large amount of data.

A variety of approaches can serve the purpose to convey in a single product the information associated to remote sensing observations of the vertical distribution of a given atmospheric target from multiple independent sources. Despite the fact that methodologies to combine coincident measurements from vertical sounders developed to a relatively lesser extent compared to similar classes of algorithms applicable to imaging systems, they are of great and increasing importance to respond to the need for full exploitation of data from new satellites, such as the Copernicus Sentinels and contributing missions. Strategies for combined use of multiple atmospheric profile datasets encompass from a posteriori data fusion techniques to synergistic inversion processes (Aires et al., 2012 and references therein; Natraj et al., 2011; Cuesta et al., 2013; Cortesi et al., 2016; Sato et al., 2018), and, in broader terms, might include assimilation systems with their unique capability of gap filling by merging model and experimental data (Lahoz and Schneider, 2014).

These three different approaches differ in the accepted inputs and in the involved models. In the synergistic inversion the inputs consist in the radiance observations (Level 1 products) of all the involved measurements and the output profiles are obtained by a simultaneous retrieval of these observations. A posteriori fusion techniques consist in sophisticated averaging processes in which the inputs are profiles (L2 products) retrieved from the single measurements. The assimilation techniques, in their more general implementations, can accept as inputs both radiances and profiles and use the information of the measurements as inputs of an atmospheric model. Each of these strategies clearly implies different advantages and drawbacks, ultimately assessing the cost-to-benefit ratio that drives the selection of the option of choice for the specific case under investigation.

~~-~~In particular Data fusion algorithms, such as the Complete Data Fusion (CDF) (Ceccherini et al., 2015), can be particularly well suited to reduce the data volume that users need to access and handle while retaining the information content of the whole level 2 (L2) products. ~~In other words, whenever the user does not need the full spatial and temporal resolution, but wants to exploit without loss of information the global coverage of the observations, an algorithm such as CDF is particularly useful since it is able to reduce the data volume of the input products to that corresponding to the required space and time resolution, while using all the available observations.~~

The CDF input is any number of L2 profiles retrieved with the optimal estimation technique and characterized by their a-priori information, covariance matrix (CM) and averaging kernel (AK) matrix. The output of the CDF is a single product (also characterized by an a-priori, a CM and AK matrices) which collects all the available information content.

This work is based on the simulated data produced in the context of the Advanced Ultraviolet Radiation and Ozone Retrieval for Applications project (AURORA, Cortesi et al.,2018), funded by the European Commission in the framework of the Horizon 2020 programme. The project regards the sequential application of fusion and assimilation algorithms to ozone profiles simulated according to the specifications of the atmospheric Sentinels.

The use of synthetic data allows evaluating the performances of the algorithm also in terms of differences between the products of interest and a reference truth, represented by the atmospheric scenario used in the procedure to simulate the L2 products. On the other hand, the absence of systematic errors in the simulated measurements limits the study to ideal measurement conditions. However, the CDF algorithm intrinsically provides a mechanism to include different kinds of errors into the analysis. For instance, Ceccherini et al. (2018) discussed how interpolation and coincidence errors can be accounted for and Ceccherini et.al. (2019) explicitly introduces the treatment of systematic errors.

This work is divided in two parts. In the first part, we describe the datasets and methodologies (the L2 simulation procedure and the CDF) used in the present paper~~.~~ and ~~we~~ discuss the use of the profiles average as fusion technique. In the second part,

the quality of the fused products obtained from L2 profiles that are not perfectly co-located in space and in time ~~with each other~~ is analysed. To account for the geo-temporal differences in the L2 profiles, a coincidence error is added to the fused product error budget. The fused and standard L2 products are compared and assessed in terms of their information content, highlighting the better data quality provided by the fusion. Finally, we also show that the CDF can be applied with different coincidence grid-box sizes, allowing for different compression factors of the Level 2 input data volume.

The application of CDF to L2 products simulated with the characteristics expected from the atmospheric Sentinel 4 and 5 allows to establish the possible benefits in case of real Sentinel data.

2. Material and methods

2.1. Atmospheric scenario and ozone climatology

Two basic external sources have been used to generate the database of the standard L2 ozone products used in this work: the ozone climatology and the atmospheric scenario.

The ozone climatology was used as a priori information for both the simulation of L2 products and the calculation of the CDF. The atmospheric scenario represents the true state of the atmosphere and is used for both the simulation of L2 products and the quality assessment of the fused ones.

The ozone climatology was derived from McPeters and Labow (McPeters and Labow, 2012) and directly provides the a priori profile x_a used either in the simulation expressions (Eq (1)), as well as in the fusion (see Eq. (6)). The CM of the a priori profile $\mathbf{S}_a$ is obtained setting the diagonal terms equal to the square of the standard deviation of ~~the~~ McPeters and Labow climatology where this standard deviation is larger than 20% of the a priori profile and to the square of 20% of the a priori profile otherwise. The off-diagonal elements are calculated using a correlation length of 6 km. The correlation length is used to reduce oscillations in the simulated profiles and the value of 6 km is typically used for nadir ozone profile retrieval (Liu et al., 2010~~, Kroon et al., 2011,~~ Miles et al., 2015). The a priori CM is used in the simulation of the L2 products, and in particular in the expression of the AK matrix (Eq. (2)) and in the CMs of Eq. (4) and (5) of the next paragraph. The a priori CM $\mathbf{S}_a$ plays an important role also in the CDF equations (see Eq. (6)).

The atmospheric scenario is taken from the Modern Era-Retrospective analysis for Research and Applications version 2 (MERRA2) reanalysis (Gelaro et al., 2017). The MERRA2 data are provided by the Global Modelling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center. This reanalysis covers the recent time of remotely sensed data, from 1979 through the present. The atmospheric scenario is the source of true profile $\boldsymbol{x}_t$ used in Eq. (1) to synthetize the simulated L2 products and represents the main reference for the comparison of the quality of L2 and fused products.

2.2. L2 Product Simulation

The simulation algorithm has been originally formalized in the context of the AURORA project, aiming at an efficient computational process. The L2 retrieved state is simulated on a fixed vertical grid with a 3 km step, by the linear approximation given in Eq. (1):

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

where $\boldsymbol{x}_t$ is the true state of the atmosphere represented by the atmospheric scenarios, $\boldsymbol{x}_a$ is the a priori estimate of the state vector provided by the ozone climatology, $\boldsymbol{\delta}$ is the uncertainty in the retrieved value due to measurement noise, and $\mathbf{A} = \partial\hat{\boldsymbol{x}}/\partial\boldsymbol{x}_t$ is the AK matrix (Rodgers, 2000~~, Ceccherini et al., 2003; Ceccherini and Ridolfi, 2010~~) calculated according to Eq. (2):

$$\mathbf{A} = \left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} \quad (2)$$

In Eq. (2), **K** is the Jacobian matrix of the forward model, the superscript T represents the transpose operator, $\mathbf{S}_y$ is the CM of the observations and $\mathbf{S}_a$ is the CM of the a priori profile. The retrieval error $\boldsymbol{\delta}$ is calculated applying the gain matrix **G** (Rodgers, 2000) to an error $\boldsymbol{\varepsilon}$ on the observations randomly taken from a Gaussian distribution with average equal to zero and CM given by $\mathbf{S}_y$:

$$\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\varepsilon} = \left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}\boldsymbol{\varepsilon} \quad (3)$$

The CM **S** ~~and $\mathbf{S}_{total}$ CMs~~ associated to the retrieval error $\boldsymbol{\delta}$ (introduced in Eq. (3)) ~~and to the total error $\boldsymbol{\delta}_{total}$ (i.e. the difference between the simulated and the true profiles that is equal to the random $\boldsymbol{\delta}$ plus the so-called smoothing error caused by the limited vertical resolution, see also Eq. (7))~~ are is given by Eq. (4) (Rodgers, 2000~~, Ceccherini and Ridolfi, 2010~~), ~~and Eq. (5) (Rodgers, 2000), respectively~~:

~~and to the total error $\boldsymbol{\delta}_{total}$ (i.e. the difference between the simulated and the true profiles that is equal to the random $\boldsymbol{\delta}$ plus the so called smoothing error caused by the limited vertical resolution, see also Eq. (7))~~

$$\mathbf{S} = \langle\boldsymbol{\delta}\boldsymbol{\delta}^T\rangle = \left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}\left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1} \quad (4)$$

The CM $\mathbf{S}_{total}$ associated to the total error $\boldsymbol{\delta}_{total}$ (that is the difference between the simulated and the true profiles, equal to the random $\boldsymbol{\delta}$ plus the so-called smoothing error caused by the limited vertical resolution of the measurement; see Eq. (7)), is given by Eq. (5) (Rodgers, 2000):

$$\mathbf{S}_{total} = \langle\boldsymbol{\delta}_{total}\boldsymbol{\delta}_{total}^T\rangle = \left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1} \quad (5)$$

It should be noted that through the term $\boldsymbol{\delta}$ it is possible to simulate additional error components with respect to the random one considered in this study and this fact adds flexibility to the simulation method.

~~The~~ In this study, the above formulation was used to simulate ozone profiles in the two spectral bands (UV and TIR) for both S4 and S5, after considering the instrument specifications and accounting for the differences in the two spectral bands. In particular, if a fixed geo-location is considered, starting from the same true profile and the same a priori information, the L2 products of the different instruments are obtained by the choice of the suitable Jacobian matrix **K** and of the CM $\mathbf{S}_y$ that have been synthetized using the technical requirements of the considered platforms and their foreseen performances.

It is important to remind here that ozone profiles measured in the UV region will be retrieved from spectral radiances acquired by the UVNS/Sentinel-5 spectrometer on-board Meteorological Operational satellite - Second Generation (MetOp-SG) and by the UVN/Sentinel-4 spectrometer on-board Meteosat Third Generation Sounder (MTG-S). For ozone and other targets observed in the TIR, the atmospheric Sentinel missions will be using the operational products of IASI-NG on MetOp-SG and of IRS on MTG.

In the framework of the AURORA project, simulated ozone products from the above-mentioned instruments operating in the UV and TIR regions were generated by using the most up-to-date information available. A detailed description of the instrumental and observational features, as well as of the characteristics of different L2 products goes beyond the scope of this article. All the relevant information was reported in the Technical Note on L2 Data Simulations (AURORA 2017) and can also be found in (Cortesi et al., 2018) and in table 1 of (Tirelli et al 2020), where the spectral bands and the available products specifications used in for the simulation of ozone products in the context of the AURORA project are summarized.

For the sake of shortness, in the next sections of the paper, we will refer to UVNS/MetOp-SG as S5-UV1, to UVN/MTG as S4-UV1, to IASI-NG/MetOp-SG as S5-TIR and to IR/MTG as S4-TIR.

~~A detailed description of the specific characteristics of different L2 products goes beyond the scope of this article and can be found in (Cortesi et al. 2018).~~

~~This simulation method offers indisputable advantages in terms of speed of calculation compared to a complete retrieval.~~

2.3. The CDF method

In this Section, we briefly recall the formulas of the CDF method (Ceccherini et al., 2015). We assume to have *N* independent simultaneous measurements of the vertical profile of an atmospheric species that can be referred to the same geo-location. Performing the retrieval of the *N* measurements, we obtain *N* vectors $\hat{\boldsymbol{x}}_i$ (*i*=1, 2, …, *N*) ~~that provide~~ providing independent estimates of the profile, here assumed to be represented on a common vertical grid. Using as inputs these *N* measurements, the CDF produces as output a single product characterized by a profile $\boldsymbol{x}_f$, an AK matrix $\mathbf{A}_f$ and a CM matrix $\mathbf{S}_f$ with the procedure summarized by Eq.s (6). These three quantities are function of the input products, $\mathbf{A}_i$, $\mathbf{S}_i$, hereafter referred to as fusing products, and depend on the a priori information ($\boldsymbol{x}_a$,$\mathbf{S}_a$) used as a constraint for the fused product.

$\boldsymbol{\alpha}_i = \hat{\boldsymbol{x}}_i - (\mathbf{I} - \mathbf{A}_i)\boldsymbol{x}_{ai} = \mathbf{A}_i\boldsymbol{x}_t + \boldsymbol{\delta}_i + \mathbf{A}_i\boldsymbol{\delta}_{coinc,i}$	(6a)
$\tilde{\boldsymbol{S}}_i = \mathbf{S}_i + \mathbf{A}_i\mathbf{S}_{coinc,i}\mathbf{A}_i^T$	(6b)
$\boldsymbol{x}_f = \left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i + \mathbf{S}_a^{-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\boldsymbol{\alpha}_i + \mathbf{S}_a^{-1}\boldsymbol{x}_a\right)$	(6c)
$\mathbf{A}_f = \left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i + \mathbf{S}_a^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i$	(6d)
$\mathbf{S}_f = \left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i + \boldsymbol{S}_a^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i\left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i + \mathbf{S}_a^{-1}\right)^{-1}$	(6e)
$\mathbf{S}_{f\,total} = \left(\sum_{i=1}^{N}\mathbf{A}_i^T\tilde{\mathbf{S}}_i^{-1}\mathbf{A}_i + \boldsymbol{S}_a^{-1}\right)^{-1}$	(6f)

For what concern the profile and the error the CDF can be thought as a "smart average" in which the a priori information is removed from the L2 profiles and ~~V~~CMs before they are put together in the average. The total error of the single product whose a priori information has been removed is higher than the original one and this error increase is only partially compensated by the effect of the average. This is the reason why, even if the total error of the fused product is generally lower than the one of the single L2 fusing product, it is in general higher than the error of the average. The behaviour of the AK matrix is less intuitive and will be thoroughly analysed in the presentation of the results.

A coincidence error characterized by a CM $\mathbf{S}_{coinc}$ is added if the input products are not coincident in time and space. When the CDF is applied to not perfectly coincident products, the diagonal elements of $\mathbf{S}_{coinc}$ are calculated as the square of the 5% of the a priori profile ~~estimates in the ozone climatology~~$\boldsymbol{x}_a$~~;~~. this value has been chosen considering the size of the coincidence grid cells used in this study. ~~-~~The off-diagonal elements of $\mathbf{S}_{coinc}$ are obtained applying an exponential decay with a correlation

length of 6 km (Ceccherini et al. 2018). In (Ceccherini et al. 2019) the dynamical choice of $\mathbf{S}_{coinc}$ is presented and in particular the a priori error (coincident with the climatological variability) is used as reference for the diagonal elements and a fixed exponential decay is applied too but the multiplicative factor ~~here fixed to 0.05~~ is calculated by imposing that the cost function of the retrieval is equal to its expectation value. ~~minimizing an opportune cost function.~~ That study, which is based on simulated products similar to the ones of this work, shows ~~that 0.05 (5%) is a good compromise in the considered cases and~~ that even if the coincidence error is strictly needed for the correct behaviour of the CDF product, this is not strongly dependant by its exact amount until it is smaller with respect to the errors of the individual L2 products.

The formulas of Eqs.(6) refer to the case of measurements made on the same vertical grid. In general, also an interpolation error may be needed considering that the retrievals of the products to be fused can be furnished on different vertical grids. In (Ceccherini et al. 2018) the general expressions of CDF in the case of the fusion of products characterized by different vertical grids are presented and discussed together with the expression of the interpolation error that depends on the involved grids and on the AK matrices of the fusing products. ~~, but s~~However, since the interpolation error ~~it~~ does not apply to the present study (the L2 products have been simulated on the same vertical grid) it has ~~is~~ not been considered in Eqs. (6) and in the following discussion.

2.4. Arithmetical average and biases

Before proceeding, it is necessary to clarify why the arithmetic average of the profiles cannot be considered as a good option to represent a set of products retrieved with optimal estimation techniques.

To do this, we consider N coincident L2 measurements (i=1,..,N) ~~that~~ referring to the same true profile, the same AK matrix and the same CM but ~~have~~ having different (noise) errors $\boldsymbol{\delta}_i$ randomly generated according to Eq. (3). The total error expression for the *i-th* measurement is given in Eq. (7) that can be easily derived from Eq. (1).

$$\boldsymbol{\delta}_{i,\text{total}} = \hat{\boldsymbol{x}}_i - \boldsymbol{x}_t = (\mathbf{I} - \mathbf{A}_i)(\boldsymbol{x}_a - \boldsymbol{x}_t) + \boldsymbol{\delta}_i \quad (7)$$

Considering that the individual measurements are co-located in space and time, thus they refer to the same truth, the same a priori profile and the same AK matrix **A**, the mean total error is equal to:

$$\langle \boldsymbol{\delta}_{i,\text{total}} \rangle = \langle \hat{\boldsymbol{x}}_i \rangle - \boldsymbol{x}_t = \ldots = (\mathbf{I} - \mathbf{A})(\boldsymbol{x}_a - \boldsymbol{x}_t) + \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_i \quad (8)$$

It follows that the averaging process reduces the random component of the total error, but does not reduce the bias, due to the a priori information and equal to the term $(\mathbf{I} - \mathbf{A})(\boldsymbol{x}_a - \boldsymbol{x}_t)$ of Eq. (8), which therefore becomes a dominant component. The existence of this bias is one of the reasons why the arithmetic mean cannot be considered as a reference algorithm to collect the information of several products into one. Further reasons concern the choice of a suitable AK matrix to be assigned to the average (see also von Clarmann. and Glatthor 2019) and the management of possible coincidence and interpolation errors. An explicative comparison of the application of CDF and standard averages in the case of 1000 coincident L2 products is reported in the supplementary material ~~section~~.

3. Results and discussion

3.1. Fusion in realistic spatial and temporal resolution conditions: the L2 Datasets

To analyse the behaviour of CDF in realistic spatial and temporal resolution conditions four sets of measurements were considered. These measurements correspond to the cloud free observations that were possible between 9:00am and 10:00am

on the 1st April 2012. ~~Table 1~~Table 1 lists the L2 product types, namely S4-TIR, S4-UV1, S5-TIR and S5-UV1, used in the remaining of the article. The L2 datasets have been generated according to the equations (1)-(5) described in the paragraph 2.2. The details of the simulation process can be explored in the technical note (AURORA 2017) considering that here we simulated all the pixels corresponding to a clear sky line of sight in the atmospheric scenario without applying any additional selection criteria. In fact, in the AURORA project 4 months of data have been considered, but a subset of the clear sky pixels has been simulated to reduce the computational cost of the simulations; for this study all the clear sky pixels in the considered hour of data have been simulated, without additional filter, choosing the orbits so that S4-S5 coincidences occur; the spatial distribution of the simulated products is indirectly represented in the left panel of Figure 5.

3.2. Single grid-box analysis (0.5ºx0.625º)

We consider first the case of a single grid-box (~~Fig.1~~Figure 1). In the selected grid-box, 118 measurements were available (55 of S4-TIR, 55 of S4-UV1, 8 of S5 TIR, no S5 UV1). The cell has the size of 0.5 degrees in latitude and 0.625 degrees in longitude, centred on the Egina Island in the Aegean Sea. The cell size has been chosen to be comparable with the assimilation grid used in the AURORA project. We assign the geo-location of the fused product to be the barycentre of the horizontal coordinates of the L2 measurements in the grid-box. In this particular case, since the horizontal distribution of the 118 L2 profiles is quite homogeneous, the barycentre is practically placed at the centre of the grid-cell.

~~**Errore. L'origine riferimento non è stata trovata.**~~ Figure 2 shows with green lines the absolute (left panel) and relative (right panel) differences between each L2 profile and the corresponding true profile, with a red line the difference between the fused profile and the mean truth (computed as average of the 118 true profiles), with a black ~~solid~~dash-dotted line the average of the estimated standard deviation of total error of the individual L2 measurements $\boldsymbol{\sigma}_{total}$, and with a red ~~solid~~dash-dotted line the estimated standard deviation of the total error of the fused profile $\boldsymbol{\sigma}_{ftotal}$. The last two quantities have been calculated as the square root of the diagonals of the $\mathbf{S}_{total}$ and $\mathbf{S}_{ftotal}$ CMs given by Eqs. (5) and (6) respectively. ~~Fig.~~Figure 2 shows that the fused product is in better agreement with its truth than the individual profiles with their own, and presents a smaller estimated total error than the individual L2 products. In particular the right panel allows to see in the detail the performances of CDF in the tropospheric region.

The representation of a retrieved profile is always a compromise between the amplitude of the errors and the vertical resolution. The latter can be quantified by the AKs, which ideally would be equal to the identity matrix in the case of a profile that has a vertical resolution equal to that defined by the sampling grid. Diagonal elements with values smaller than 1 correspond to a loss of vertical resolution. In the left panel of ~~**Errore. L'origine riferimento non è stata trovata.**~~ ~~compares~~ Figure 3, the diagonal elements of the AKs of the L2 products, ~~and those~~are compared with the one of the fused product . ~~**Errore. L'origine riferimento non è stata trovata.** shows that the fused product has better vertical resolution than all the other L2 products at all pressure levels.~~ ~~W~~where we have also computed the number of Degrees Of Freedom (DOFs), given by the sum of the diagonal elements of the AK matrix (Rodgers, 2000), for both L2 and fused products, and reported the values in the text box of the left panel ~~**Errore. L'origine riferimento non è stata trovata.**~~. ~~T~~; as it can be noted the number of DOFs of the fused product is about twice the number of DOFs of the best L2 one. In the right panel of Figure 3, the comparison of the vertical resolution profiles of L2 and FUS products is shown, where the vertical resolution is calculated starting from AK matrices according the Full Width Half Maximum (FWHM) approach (Rodgers, 2000) and in particular with the algorithm defined in (Ridolfi and Sgheri, 2009).

From the comparison of the left and the right panel of Figure 3 it can be noted that the increase of the AK matrix diagonal values of FUS product, and consequently the increase of the number of DOFs, implies an improved vertical resolution only in

a subset of the vertical levels. To better understand the effect of the fusion on the AK matrices, it is useful to analyse the behaviour of their individual rows. In Figure 4, two rows are represented, one that refers to the troposphere (left panel, 6 km) one to the middle stratosphere (right panel, 39 km), where the reference altitude is the one corresponding to the diagonal value of the row. The value of the vertical resolution at the considered altitude is reported in the legend (the minimum of vertical resolution at the considered vertical level for each type of L2 product) while the diagonal value of each row is evidenced in the graphs with cross (L2) and dot markers (FUS). At lower altitudes (left panel), as suggested by one of the reviewers, the DOFs increase can be attributed to three distinct phenomena. The first is the constriction of the main FUS AK lobe and the consequent improvement (of more than 30%) of the vertical resolution with respect to L2 products. The second phenomenon is linked to the fact that, while for the FUS product the maximum value of the AK row corresponds to its diagonal element, for the L2 products these maxima are shifted with respect to the reference altitude of the rows. The last phenomenon is a stronger contribute of the measures with respect to the a priori in the FUS product, where the latter effect can be evidenced considering the sum of all the elements of the rows that assume 0.913 as maximum value for the L2 products and 0.956 for the FUS product. In the particular case all these three effects go in such a direction that can be considered as benefits of CDF application. The results at higher altitudes (39 km, right panel) are primarily influenced by the shape of the AK rows that exhibit large secondary lobes that degrade the vertical resolution.

3.3. Statistical analysis for a large domain ~~(0.5°x0.625°)~~

While the analysis of the previous paragraph focuses on a particular grid-box~~cell~~, here an ~~overview~~ analysis of the CDF behaviour is presented, referring to all the 1939 fusion grid-boxes ~~with size of 0.5 degrees in latitude and 0.625 degrees in longitude~~ in which more than one of the 79781 L2 simulated products, considered in ~~Tab.~~Table 1 is placed. The fused products can be classified depending on the types of L2 measurements falling inside the coincidence grid cell. Since S4-TIR and S4-UV1 products are in perfect coincidence and S5-UV1 products have a horizontal spacing larger than the cell size, only six fused product types (FUS type), listed in Table 2, effectively occur. In this table, the FUS type and its description are reported together with the following complementary data:

- Ncells: the number of grid-boxes characterized by the considered FUS type.
- <NL2>: the mean number of individual L2 fusing profiles per grid-box.
- Max NL2: the maximum number of individual L2 fusing products per grid-box.

The left-hand side panel of Figure 5 ~~Fig.4~~ shows the geographical distribution of the FUS products. Different colours have been used to classify the fused data according to their provenance type. The irregular geographical coverage is due to the realistic distribution of the cloud free measurements. The histogram in the right-hand side panel of Figure 5 shows the number of cells that contain a given number of measurements, divided in different colours depending on the FUS type. The FUS cells, in which only S5 platform L2 products fall, are characterized by a small number of L2 measurements, while when S4 products are present, many L2 measurements can be present.

With the selected grid-box size and the multitude of different products that are present in each cell, the question is which product can be used in alternative to the fusion process in those operations in which a single product is requested in each grid-box. Since the averaging process is affected by a large bias error, a viable alternative is the use of the best fusing product present in the cell and we want to compare the CDF result with this product. This comparison is the ~~so-called~~so-called Synergy~~ic~~ Factor (SF), introduced by Aires et al. (2012). Although Aires introduces SF only for errors (Eq. (11)), we extend~~ed here their~~ his definition also for other quantities because they constitute a useful tool to synthetically represent the performances of fusion algorithms.

The *SF DOF*, defined by Eq. (9), is a pure number that can be calculated for every FUS pixel by the ratio of the number of DOFs of the FUS product and the maximum number of DOFs of the L2 measurements that have been fused. In this equation the index *l* enumerates the vertical levels and the index *i* enumerates the L2 products fused in each grid-box.

$$SFDOF = \frac{\sum_l \mathbf{A}_{f,ll}}{\max\limits_{i \in L2} \sum_l \mathbf{A}_{ll}} \quad (9)$$

When *SF DOF* is larger than ~~one~~1.0, the FUS product carries more information than the individual L2 measurements. Figure ~~5~~ 6 shows that the *SF DOF* computed for all the fused products (and plotted as a function of the number of L2 profiles in each grid-box) is always larger than ~~one~~1.0. This means that the information content of the fused product is always larger than that of the standard L2 retrievals. It is also worth noticing that *SF DOF* increases approximately linearly with the logarithm of the number of fusing products, although the proportionality depends on the FUS type. The two different clusters of red symbols (S4:TIR+UV1) are caused by the different latitude bands in which these products are distributed (see also left panel of Figure ~~4~~5). It is important to underline that the improvement in vertical resolution, which cannot be obtained with the arithmetic averaging, is the most demanding requirement (in terms of observation time and instrument sensitivity) in remote sensing observations and, considering the significant gain obtained relative to the single product selection, is the most important feature of fused products.

While *SF DOF* is a scalar quantity, both ***SF AK*** and ***SF ERR***, defined by Eqs. (10) and (11), are vertical profiles of pure numbers. ***SF AK*** represents an expansion on the vertical dimension of *SF DOF* and, in particular, is calculated, level by level, as the ratio between the diagonal elements of the AK matrix of the FUS product and the maximum of the corresponding elements of the AK matrices of the fusing L2 measurements.

A value of ***SF AK*** larger than 1.0 at a specific vertical level (indicated by the index *l*) means that, at that level, ~~the vertical resolution~~the diagonal value of the AK matrix of the FUS product ~~is better~~ has a ~~higher~~larger value than that of all the individual products. As we have seen commenting Figure 3 and Figure 4, the increase of the AK diagonal values at a specific level can happen for different reasons but all of them can be considered as an improvement in the product quality.

$$SFAK_l = \frac{\mathbf{A}_{f,ll}}{\max\limits_{i \in L2} \mathbf{A}_{i,ll}} \quad (10)$$

The *SF ERR* (Eq. (11)) at a given level is calculated as the ratio between the minimum total error of the L2 measurements that have been fused and the total error of the FUS product. A value of ***SF ERR*** larger than 1.0 means that at a specific level the error of the FUS product is smaller than that of all the individual products.

$$SFERR_l = \frac{\min\limits_{i \in L2} \sigma_{total,i,l}}{\sigma_{ftotal,l}} \quad (11)$$

The SFs defined by Eqs. (10) and (11) provide a conservative comparison because the fused product is compared with the L2 product that at that level has the largest diagonal value in its AK matrix and with the one that has the smallest total error at the same level (generally, these are two distinct L2 products).

~~**Errore. L'origine riferimento non è stata trovata.**~~Figure 7 shows the ***SF AK*** (left panel) and ***SF ERR*** (right panel) profiles for the 1939 FUS products considered in ~~Tab.~~Table 2. We have used different colours to denote the provenance of the L2 data

contributing to the fused products and different symbol size to infer the number of L2 fusing measurements (the larger the symbol size, the larger the number of L2 fusing profiles). As mentioned above, the merit of the fused product in terms of SF is higher than that of the L2 retrievals if the ***SF AK*** and ***SF ERR*** are greater than 1. The significant improvement obtained with the fused products is confirmed by ~~Fig.~~Figure ~~6~~. It can also be noticed that considering symbols of the same colour the symbol size (*N*) tends to increase moving horizontally in the graph (same vertical level) from left to right (SF increasing) denoting that, for each FUS type, SF increases with *N*. This is not in contradiction with the fact that symbols with different colours (FUS types) and different sizes (*N*) can share the same position (SF, vertical level) on the graph. Some ***SF AK*** values, both in the troposphere and in middle upper atmosphere are smaller than one; ~~where~~ in the troposphere this happens in 20 cells out of 1939 while in the middle upper atmosphere this happens in almost 500 cells for two possible and sometimes simultaneous circumstances. The first one happens when the introduction of the coincidence error provokes a sensible degradation of the quality of the FUS AKM. The second circumstance happens for example when one of the L2 products is characterized by a vertical resolution that is much better than all the other fusing products and in particular the peaks of their AKM rows tends to not coincide with the nominal vertical level of the row itself.

~~A few *SF AK* values are slightly smaller than one in the middle to upper stratosphere. This is likely caused by the coincidence errors that have to be added in the fusion process.~~

3.4. Statistical analysis on a coarse horizontal resolution ~~(1°x1°)~~

We have seen that, starting from 79781 L2 measurements (~~**Errore. L'origine riferimento non è stata trovata.**~~Table 1), when a coincidence grid-box with size 0.5°x0.625° is used, the number of fused profiles is 1939 (~~**Errore. L'origine riferimento non è stata trovata.**~~Table 2), with a reduction of the data volume of more than a factor 40.

~~**Errore. L'origine riferimento non è stata trovata.**~~Table 3 provides a summary of the number of fused profiles and the provenance of the L2 profiles that contribute to them for a fusion grid resolution of 1°x1°. In this case, the total number of FUS products is 775 with a reduction of data volume of more than a factor 100.

The Synergy~~ie~~ factors *SF DOF*, ***SF AK*** and ***SF ERR*** have been considered also in this case and the figures (similar to ~~Fig.~~Figure ~~5~~ 6 and ~~Fig.~~Figure ~~6~~7) are reported in the supplementary material. In summary, the greater number of fusing observations in each fusion cell produces a further improvement for both the vertical resolution and the total error proving that the CDF method can be used with a wide range of grid-box size and data compression and the quality of the products generally improves with larger cells. An upper limit to the grid-box size is caused by the coincidence error amplitude, which increases with the geographical variability degrading the quality of the fused product; the study of this aspect will be of crucial importance if the CDF will be applied to species with greater spatial-temporal variability than ~~the~~ ozone or in any case to very large spatial-temporal domains.~~.~~

4. Conclusions

This paper presents a sensitivity study of the Complete Data Fusion technique, applied to L2 measurements simulated with the characteristics of the measurements, which will be acquired in the context of the atmospheric Sentinel missions~~with the characteristics expected for the atmospheric Sentinels 4 and 5~~. This analysis allows to evaluate the performances of the CDF algorithm in ideal conditions (i.e., with no systematic errors added) and to quantify the possible benefits of the application of CDF to real Sentinel data.

In particular, we show the application of CDF to a single cell with size of 0.5 degrees in latitude and 0.625 degrees in longitude in which more than 100 L2 products are fused. Results show that the fused product is characterized by higher information content, smaller errors and smaller residuals (i.e., anomaly from the true profile) compared to individual L2 products. The information content being, with its improvement of the vertical resolution, the most important achievement.

This analysis is then extended to a larger domain consisting in 79781 L2 products subdivided in 1939 grid boxes with 0.5ºx0.625º size. In this case, the comparison of L2 products and CDF output are carried on in terms of synergyic factors. This analysis shows that the CDF can be applied to a wide range of situations and that the benefits of the fusion strongly depend on the number of the measurements that are fused together and from their characteristics. It is also shown that CDF can be run customizing grid resolutions, e.g. to match the resolution requirements of the process that will ingest the products, with full exploitation of all the available measurements.

As the fused products are traced back to a regular, fixed horizontal grid and, as shown here, are not affected by the bias introduced by the a priori information, they can be considered as a new type of level Level 3 products with improved quality (reduced bias) and the same characteristics (AK included) with respect to L2 products, even if further analysis are needed especially for what concern the coincidence error to be applied fusing data on large spatial-temporal domains.

Data availability

The data of the simulations presented in the paper are available from the authors upon request.

MERRA-2 data (atmospheric scenario) are available at MDISC (https://disc.gsfc.nasa.gov) , managed by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

The ML climatology (McPeters and Labow, 2012) is available online from the Goddard anonymous ftp account: ftp://toms.gsfc.nasa.gov.

Author contributions (according to CRediT https://casrai.org/credit/)

N. Zoppetti: Conceptualization, Methodology, Software, Writing – Original Draft, Writing – Review & Editing, Investigation, Data curation, Visualization S. Ceccherini: Conceptualization, Methodology, Investigation, Writing – Review & Editing B. Carli: Conceptualization, Methodology Writing – Review & Editing, Supervision S. Del Bianco: Investigation, Data curation, Project Administration M. Gai: Investigation, Data curation C. Tirelli: Investigation, Data curation, Project Administration F. Barbara: Resources R. Dragani: Investigation, Data curation, Writing – Review & Editing A. Arola: Investigation, Data curation J. Kujanpää: Investigation, Data curation R. Van Der A: Investigation, Data curation U. Cortesi: Funding Acquisition, Project Administration, Supervision, Writing – review & editing.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgments

The results presented in this paper arise from research activities conducted in the framework of the AURORA project (http://www.aurora-copernicus.eu/) supported by the Horizon 2020 research and innovation programme of the European Union (Call: H2020-EO-2015; Topic: EO-2-2015) under Grant Agreement N. 687428.

Financial support.

This research has been supported by the European Commission, H2020 (AURORA, grant no. 687428).

References

Aires,F., Aznay,O., Prigent,C., Paul,M. and Bernardo,F.: Synergistic multi-wavelength remote sensing versus a posterior combination of retrieved products: Application for the retrieval of atmospheric profiles using MetOp-A. J GEOPHYS RES, Vol. 117, D18304, https://doi.org/10.1029/2011JD017188, 2012.

AURORA consortium, (Advanced Ultraviolet Radiation and Ozone Retrieval For Applications, grant no. 687428): Technical Note On L2 Data Simulations [D3.4], 35 pp., available for download at https://cordis.europa.eu/project/id/687428/results, 2017.

~~Ceccherini, S., Carli, B., Pascale, E., Prosperi, M., Raspollini, P. and Dinelli, B.M.: Comparison of measurements made with two different instruments of the same atmospheric vertical profile, Appl. Opt.,42,6465-6473, https://doi.org/10.1364/AO.42.006465, 2003.~~

~~Ceccherini, S., and Ridolfi, M.: Technical Note. Variance-covariance matrix and averaging kernels for the Levenberg-Marquardt solution of the retrieval of atmospheric vertical profiles, Atmos. Chem. Phys., 10, 3131-3139, https://doi.org/10.5194/acp-10-3131-2010, 2010.~~

~~Ceccherini, S., Carli, B., and Raspollini, P.: The average of atmospheric vertical profiles, Opt. Express, 22,24808-24816, https://doi.org/10.1364/OE.22.024808, 2014.~~

Ceccherini, S., Carli, B., and Raspollini, P.: Equivalence of data fusion and simultaneous retrieval, 2015. Opt. Express, 23, 8476-8488, https://doi.org/10.1364/OE.23.008476, 2015.

Ceccherini, S., Carli, B., Tirelli, C., Zoppetti, N., Del Bianco, S., Cortesi, U., , Kujanpää, J., and Dragani, R.: Importance of interpolation and coincidence errors in data fusion. Atmos. Meas. Tech., 11, 1009–1017, https://doi.org/10.5194/amt-11-1009-2018, 2018.

Ceccherini S., Zoppetti N., Carli B., Cortesi U., Del Bianco S., and Tirelli C.: The cost function of the data fusion process and its application. Atmos. Meas. Tech., 12, 2967–2977, https://doi.org/10.5194/amt-12-2967-2019, 2019.

Cortesi, U., S. Del Bianco, S. Ceccherini, M. Gai, B.M. Dinelli, E. Castelli, H. Oelhaf, W. Woiwode, M. Höpfner, D. Gerber, Synergy between middle infrared and millimeter-wave limb sounding of atmospheric temperature and minor constituents, Atmos. Meas. Tech., 9, 2267-2289, https://doi.org/10.5194/amt-9-2267-2016, 2016.

Cortesi, U., Ceccherini, S., Del Bianco, S., Gai, M., Tirelli, C., Zoppetti, N., Barbara, F., Bonazountas, M., Argyridis, A., Bós, A.,Loenen, E., Arola, A., Kujanpää, J., Lipponen, A., Nyamsi, W.W., van der A, R., van Peet, J., Tuinder, O., Farruggia, V., Masini, A., Simeone, E., Dragani, R., Keppens, A., Lambert, J.-C., van Roozendael, M., Lerot, C., Yu, H., and Verberne, K.:

Advanced Ultraviolet Radiation and Ozone Retrieval for Applications (AURORA): A Project Overview, Atmosphere, 9, 454, https://doi.org/10.3390/atmos9110454, 2018.

Cuesta, J., M. Eremenko, X. Liu, G. Dufour, Z. Cai, M. Höpfner, T. von Clarmann, P. Sellitto, G. Foret, B. Gaubert, M. Beekmann, J. Orphal, K. Chance, R. Spurr and J. M. Flaud: Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements over Europe, Atmos. Chem. Phys., 13 (19), pp.9675-9693, 2013.

Gelaro, R., McCarty, W., Max J. Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419-5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

~~Kroon, M., de Haan, J. F., Veefkind, J. P., Froidevaux, L., Wang, R., Kivi, R. and Hakkarainen, J. J.: Validation of operational ozone profiles from the Ozone Monitoring Instrument, J. Geophys. Res., 116, D18305, https://doi.org/10.1029/2010JD015100, 2011.~~

Lahoz, W.A. and Schneider, P., Data assimilation: making sense of Earth Observation, Frontiers in Environmental Science, 2, https://doi.org 10.3389/fenvs.2014.00016, 2014.

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

McPeters, R.D., and Labow, G.J.: Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, J. Geophys. Res., 117, D10303, https://doi.org/10.1029/2011JD017006, 2012.

Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G., and Richards, N. A. D.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385-398, https://doi.org/10.5194/amt-8-385-2, 2015.

Natraj, V., Liu, X., Kulawik, S., Chance, K., Chatfield, R., Ed-wards, D. P., Eldering, A., Francis, G., Kurosu, T., Pickering, K.,Spurr, R., and Worden, H.: Multi-spectral sensitivity studies for the retrieval of tropospheric and lowermost tropospheric ozone from simulated clear-sky GEO-CAPE measurements, Atmos.Environ., 45, 7151–7165, 2011

Ridolfi,M. and Sgheri,L. 2009: A self-adapting and altitude-dependent regularization method for atmospheric profile retrievals, Atmos. Chem. Phys., 9, 1883–1897, 2009 https://doi.org/10.5194/acp-9-1883-2009

Rodgers, C.D.: Inverse Methods for Atmospheric Sounding: Theory and Practice. Vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics. World Scientific: Singapore, 2000.

Sato, T. O., Sato, T. M., Sagawa, H., Noguchi, K., Saitoh, N., Irie, H., Kita, K., Mahani, M. E., Zettsu, K., Imasu, R., Hayashida, S., and Kasai, Y.: Vertical profile of tropospheric ozone derived from synergetic retrieval using three different wavelength

ranges, UV, IR, and microwave: sensitivity study for satellite observation, Atmos. Meas. Tech., 11, 1653–1668, https://doi.org/10.5194/amt-11-1653-2018, 2018.

Tirelli, C., Ceccherini, S.; Zoppetti, N., Del Bianco, S., Gai, M., Barbara, F., Cortesi, U., Kujanpää, J., Huan, Y., Dragani, R. Data fusion analysis of Sentinel-4 and Sentinel-5 simulated ozone data. J. Atmos. Ocean. Technol., 37 (4), 573–587, https://doi.org/10.1175/JTECH-D-19-0063.1, 2020.

von Clarmann, T. and Glatthor, N.: The application of mean averaging kernels to mean trace gas distributions, Atmos. Meas. Tech., 12, 5155–5160, https://doi.org/10.5194/amt-12-5155-2019, 2019.

L2 Type	Platform	Band	Number of simulated measurements	Minimal distance between measurements across x along track [km]
S4-TIR	S4	TIR	35594	5.7 x 7.4
S4-UV1	S4	UV1	35594	
S5-TIR	S5	TIR	8023	12.2 x 12.3
S5-UV1	S5	UV1	570	46.2 x 46.7
TOTAL			79781	

Table 1: Characteristics of the simulated measurements. For S4 platform across-track is South-North direction and along-track is East-West direction.

FUS Type	Description	*Ncells*	*<NL2>*	*max NL2*
S4:TIR+UV1	Two or more S4 pixels, no S5 pixels.	908	29.3	160
S4:TIR+UV1_S5:TIR+UV1	Two or more S4 pixels, one or more S5_TIR pixel, one or more S5_UV1 pixel.	245	114.7	163
S4:TIR+UV1_S5:TIR	Two or more S4 pixels, one or more S5_TIR pixel, no S5_UV1 pixels.	299	69.4	165
S4:TIR+UV1_S5:UV1	Two or more S4 pixels, one or more S5_UV1 pixel, no S5_TIR pixels.	2	20	37
S5:TIR+UV1	No S4 pixels, one or more S5_TIR pixels, one or more S5_UV1 pixels.	247	11.1	24
S5:TIR	No S4 pixels, two or more S5_TIR pixels, no S5_UV1 pixels.	238	6.2	14
TOTAL		1939	41.1	165

Table 2: types and characteristics of fused product when a coincidence grid cell size of 0.5°x0.625° is used. Ncells is the number of grid-boxes characterized by the considered FUS type; <NL2> is the ~~the~~ mean number of individual L2 fusing profiles per grid-box and Max NL2is the maximum number of individual L2 fusing products per grid-box.

FUS Type	Description	N_{cells}	$<NL_2>$	*max NL2*
S4:TIR+UV1	Two or more S4 pixels, no S5 pixels.	354	73.1	420
S4:TIR+UV1_S5:TIR+UV1	Two or more S4 pixels, one or more S5_TIR pixel, one or more S5_UV1 pixel.	140	289.4	504
S4:TIR+UV1_S5:TIR	Two or more S4 pixels, one or more S5_TIR pixel, no S5_UV1 pixels.	79	115.4	442
S4:TIR+UV1_S5:UV1	Two or more S4 pixels, one or more S5_UV1pixel, no S5_TIRpixels.	0	0	0

S5:TIR+UV1	No S4 pixels, one or more S5_TIR pixels, one or more S5_UV1 pixels.	142	26.2	71
S5:TIR	No S4 pixels, two or more S5_TIR pixels, no S5_UV1 pixels.	60	8.9	26
TOTAL		775	102.9	504

Table 3: Like in Table2 but with a grid-box size of 1ºx1º. Ncells is the number of grid-boxes characterized by the considered FUS type; <NL2> is the the mean number of individual L2 fusing profiles per grid-box and Max NL2is the maximum number of individual L2 fusing products per grid-box.

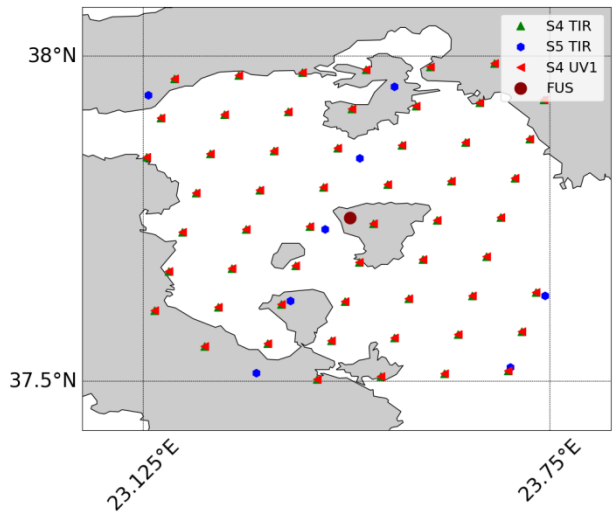

Figure 1: geographical distribution of the simulated L2 measurements and geo-location of the fused product. The black dash-dotted lines represent the borders of the 0.5ºx0.625º grid cells.

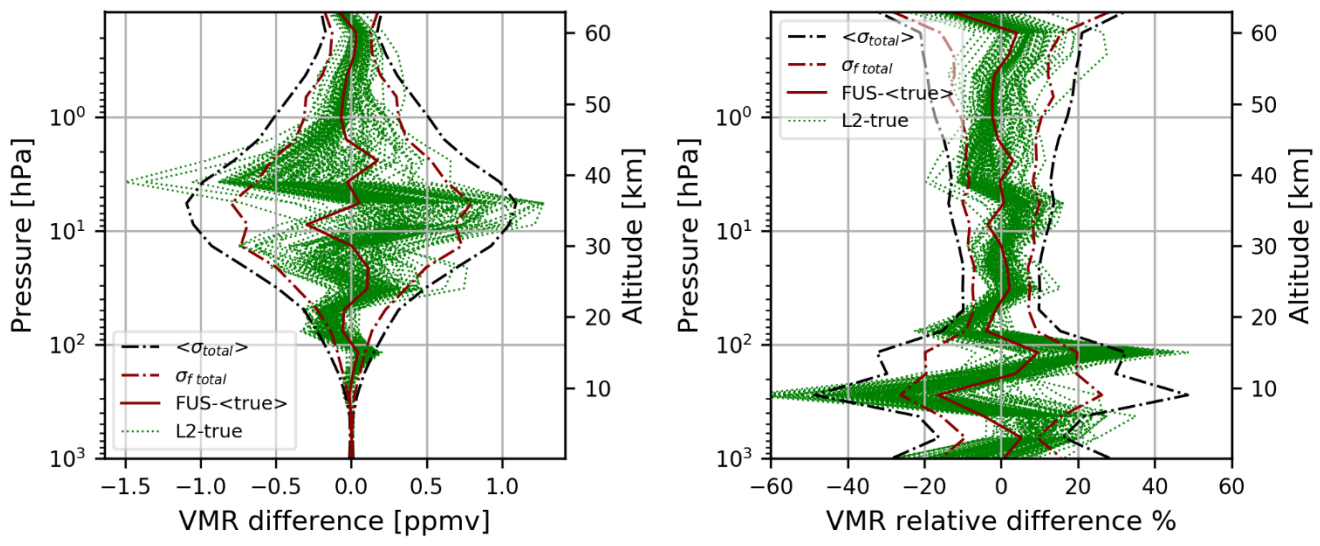

Figure 2 (Left panel): absolute differences between L2 profiles and their true profiles (green lines), absolute difference between the fused profile and the average of the true profiles (dark red continuous line), the average of σ_{total} of L2 simulations (black dash-dotted lines), $\sigma_{f\ total}$ (dark red dash-dotted lines). (Right panel): relative percentage differences between L2 profiles and their true profiles (green lines), relative percentage difference between the fused profile and the average of the true profiles (dark red continuous line), the average of σ_{total} of L2 simulations normalized wrt the true profile and expressed in percentage (black dash-dotted lines), $\sigma_{f\ total}$ normalized wrt the true profile and expressed in percentage (dark red dash-dotted lines).

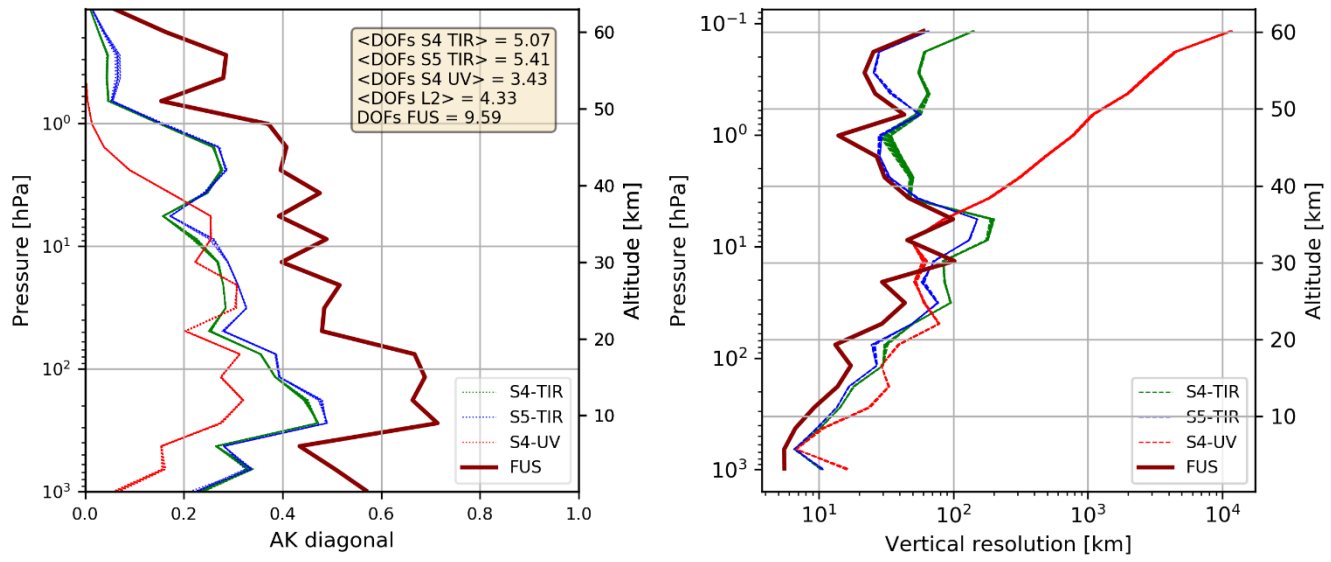

Figure 3 (Left panel): AKs diagonal of: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). In the text box the average number of DOFs for each type of L2 product, the average number of DOFs for all L2 products and the number of DOFs of the FUS product are reported. (Right panel): Vertical resolution (FWHM) profiles of: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). In each panel, while solid dark red line is a single one, red and green lines are both 55 overlapped lines and blue lines are 8 overlapped lines (one for each L2 product).

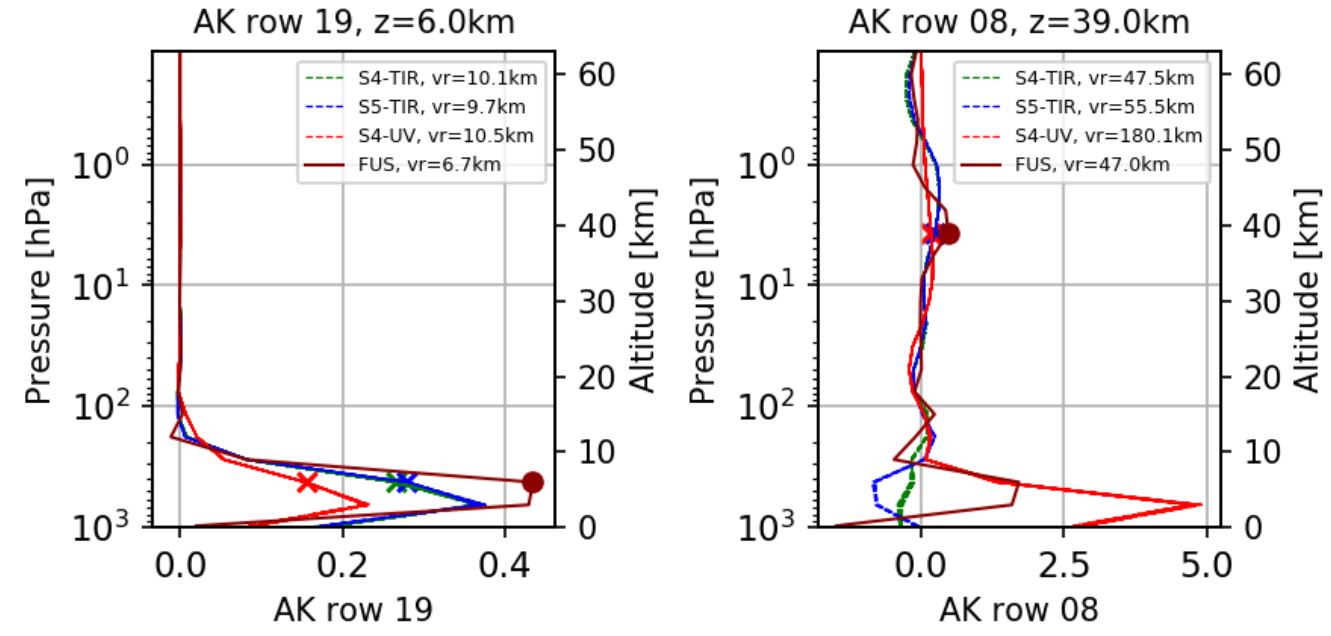

Figure 4 (Left panel): Rows of AK matrices at 6 km. (Right panel): Rows of AK matrices at 39 km S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line).

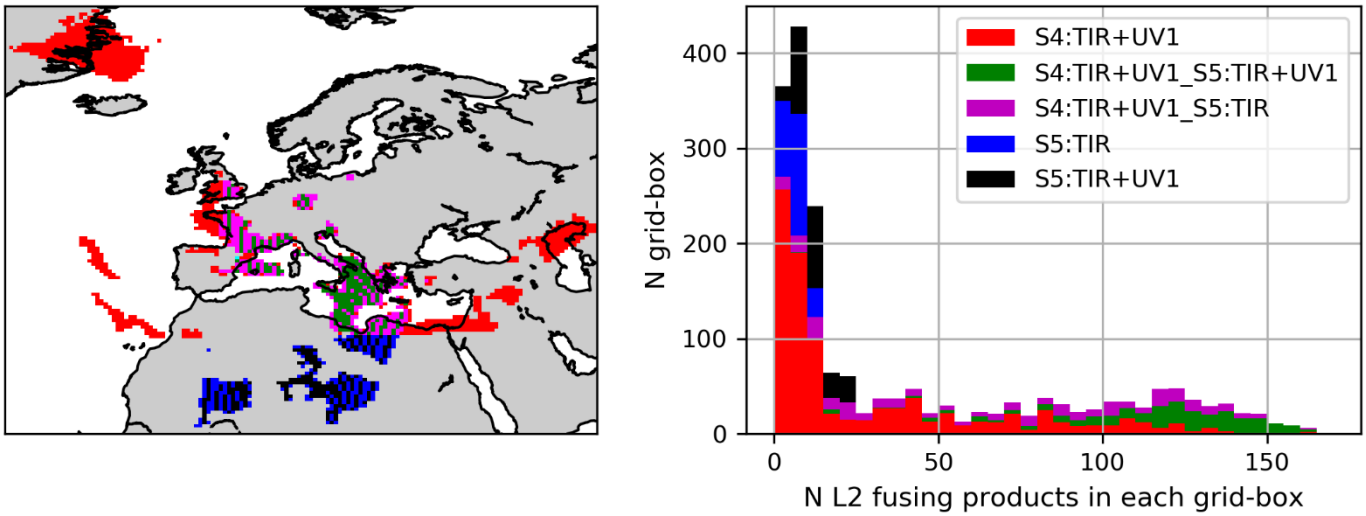

Figure 5. Left panel: geographical distribution of FUS products differentiated by FUS type where the effect of the lower resolution of S5-UV1 respect to the other L2 products is the cause of the periodic FUS type transitions in the Mediterranean area. Right panel: histogram of the number of cells with a given number of L2 measurements differentiated by FUS type.

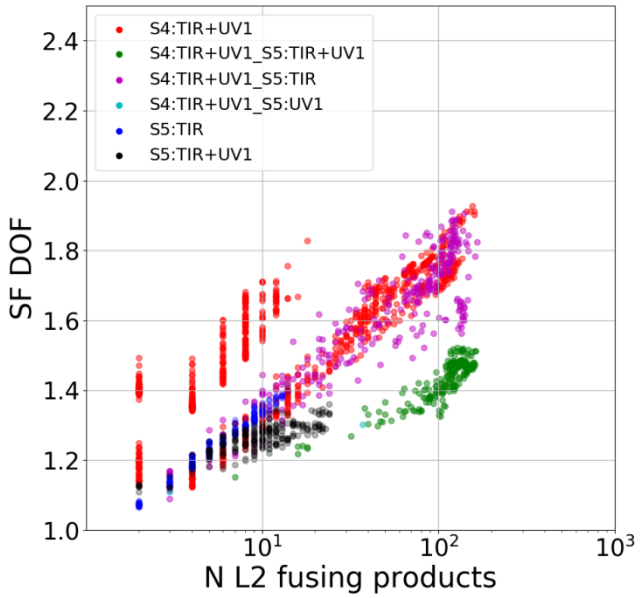

Figure 6: scatter plot of SF DOF as a function of the number of L2 measurements fused in each coincidence grid cell; different colours represent different FUS types.

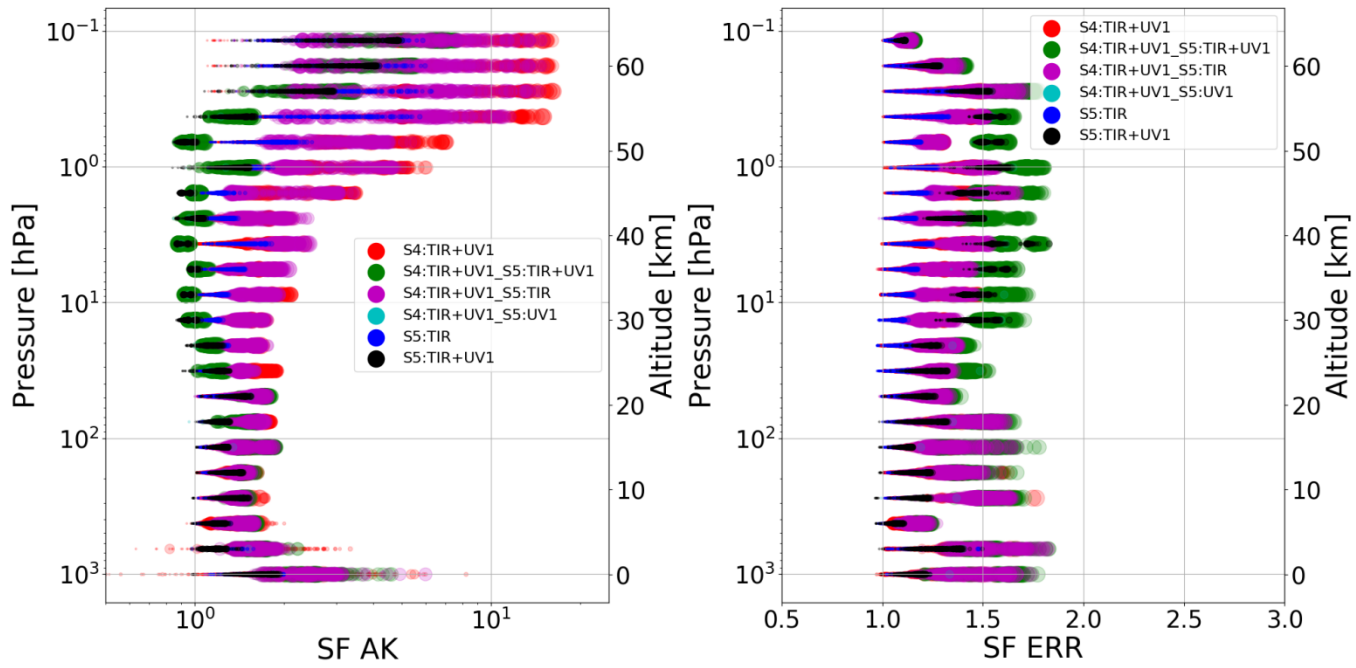

Figure 7 (Left panel): SF AK versus vertical level. (Right panel): SF ERR versus vertical level. In both panels, different colours of the symbols represent the FUS type, different sizes of the symbols represent the number of measurements that have been fused. The maximum symbol size shown in the legend corresponds to N=160.