Florence, 16/03/2020

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

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.

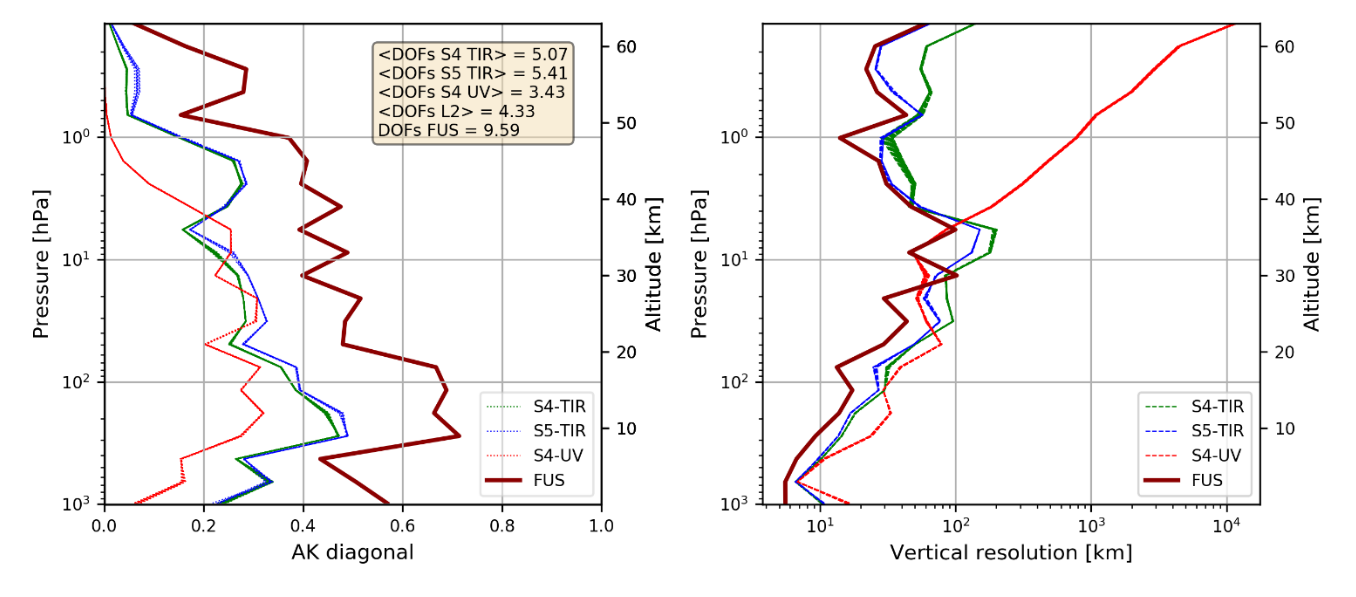

Fig. R1 (Fig3 new): ***Left panel:*** *AKs diagonal of: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). In the text box, the average number of DOFs for each type of L2 product, the average number of DOFs for all L2 products and the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

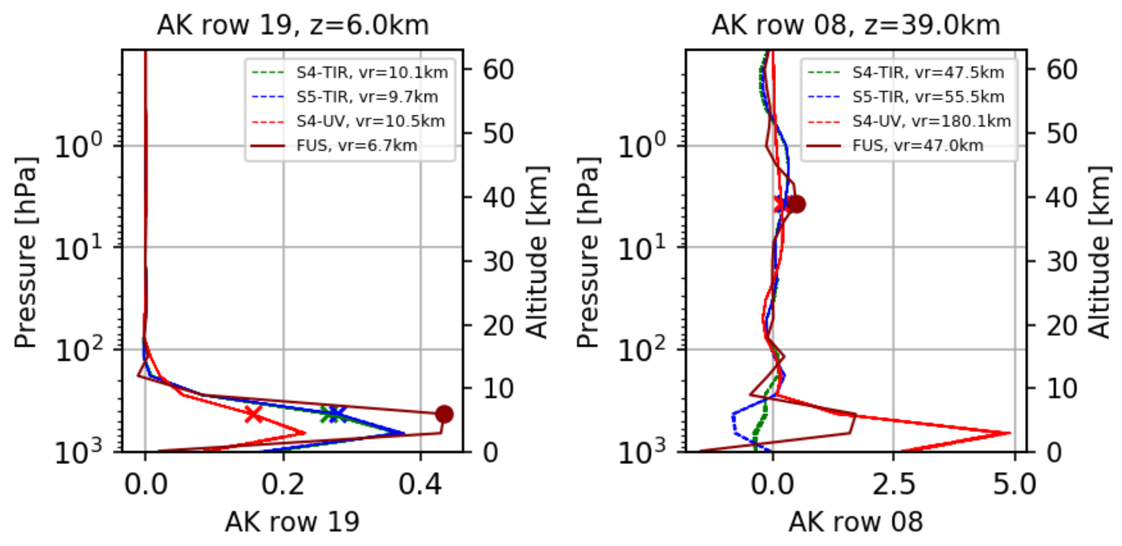

Fig. R2 (Fig.4 new): *Rows of AK matrices at 6 km (**Left panel)** and at 39 km (**Right panel)** for: S4-TIR products (red lines), S5-TIR products (blue lines) S4-UV products (red lines) and FUS product (dark red line). The value of the vertical resolution at the reference altitude is reported in the 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).

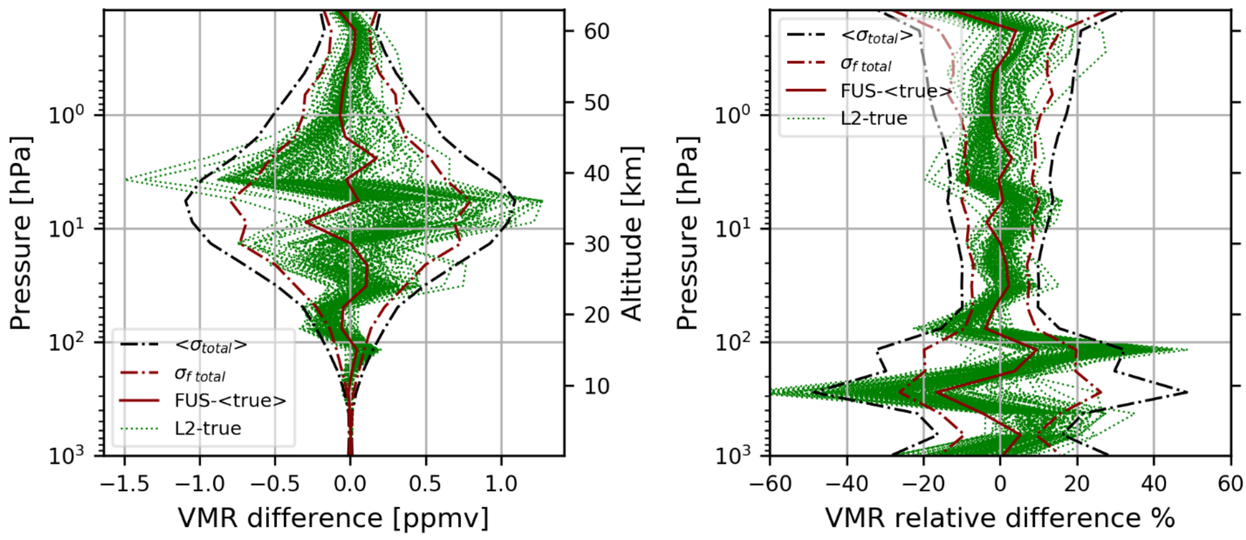

Fig.R3 (Fig2 new): *(Left panel): absolute differences between L2 profiles and their true profiles (green lines), absolute difference between the fused profile and the average of the true profiles (dark red continuous line), the average of σ_{total} of L2 simulations (black dash-dotted lines), $\sigma_{f\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

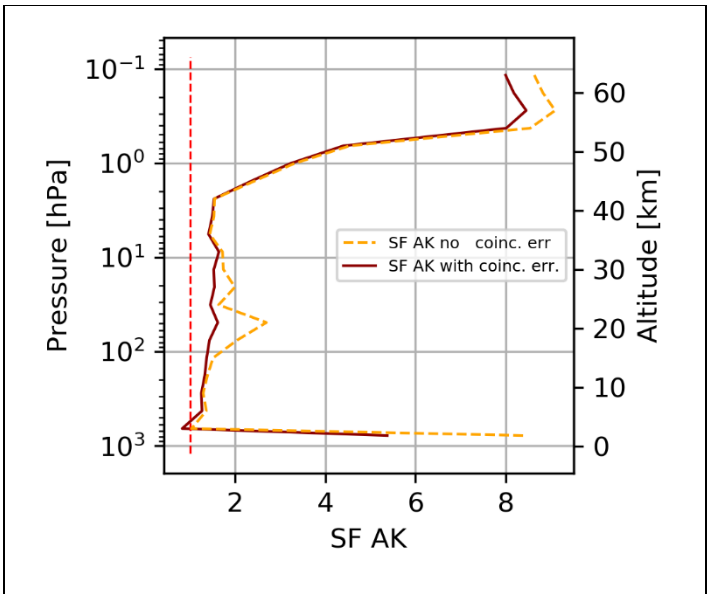

Fig. R4: *profiles of SF AK for a single cell in which 56 L2 products have been fused, considering the fusion both with the coincidence error added (dark red line) and without (orange dashed line).*

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.

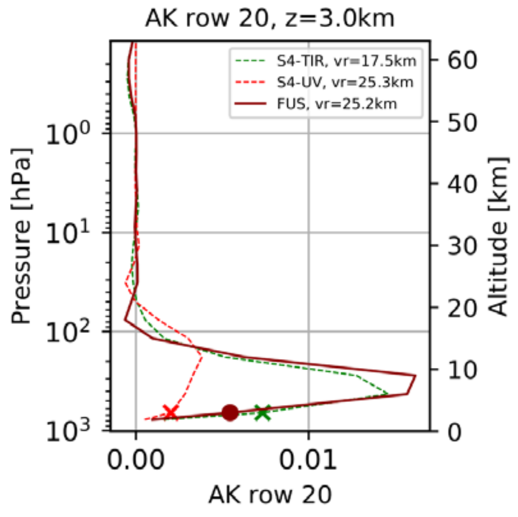

Fig. R5: *rows of the AK matrices referring to an altitude of 3 km for a cell in which two L2 products have been fused; FUS product (dark red line), S4-TIR product (red dashed line), S4-UV product (green dashed line).*

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.