Atmospheric
Measurement
Techniques

Open Access

Discussions

EGU

Interactive comment on "The Complete Data Fusion for a Full Exploitation of Copernicus Atmospheric Sentinel Level 2 Products" *by* Nicola Zoppetti et al.

Anonymous Referee #1

Received and published: 14 January 2020

General comments:

In this work the authors demonstrate that the Complete Data Fusion method, which they have previously developed themselves, is very well suited for the combination of large amounts of retrieved satellite data without information loss. The demonstration is based on (here simulated) Sentinel-4 and Sentinel-5 data, for which the future use of the Complete Data Fusion method could be of high relevance. Although the importance of the Complete Data Fusion method and its insightful demonstration in this work is acknowledged, it is believed that the authors should address two important issues before the work can be accepted for publication. First, in the "Results and discussion"

section, 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 behavior and either modify their discussion and conclusions accordingly or demonstrate that only the vertical resolution is impacted by the method. The lack of a thorough averaging kernel discussion also points at the second issue: The presentation of this work appears somewhat sloppy, with important information being vague or even missing, while it certainly deserves comprehensiveness. Several suggestions for improvement are therefore being made in the specific and technical comments below.

Specific comments:

- 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.

- 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.

- Line 45: "whenever the user does not need the full spatial and temporal resolution"

Printer-friendly version

Discussion paper

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?

- 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.

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

- 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.

- 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.

- 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.

- 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?

- Line 168: The calculation of the barycenter should be specified. Why are only horizontal coordinates considered for calculating the location of the fused product? Figure 1 seems to suggest differently.

- 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?

- 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?

- 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.

- 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. - 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.

- 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).

Technical corrections:

- 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.

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

- Line 24: It looks as if "while" is missing between "dataset" and "limiting"?

- Lines 29-30: What errors (or uncertainties) are referred to here?

- 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. The program moreover is an initiative of the European Commission, not of the European Union.

- Line 64: Remove "with each other".

- 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.

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

- Line 192: What is the latitude-longitude range of the large domain? Repeating the single grid domain in the section title is misleading here.

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

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

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