

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2022-182-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2022-182

Anonymous Referee #1

Referee comment on "An improved formula for the complete data fusion" by Simone Ceccherini et al., Atmos. Meas. Tech. Discuss.,
<https://doi.org/10.5194/amt-2022-182-RC1>, 2022

General

Ceccherini et al. in this work provide an improved formalism for the complete data fusion of OE satellite retrievals, which they have developed earlier. The derivation of this new formalism is rigorous, but could be better situated with respect to the CDF development in previous work. Moreover, the demonstration of the 'improvement' in Section 3 looks quite poor. I therefore suggest major revisions, in line with the comments below.

Specific comments

Abstract and throughout the text: It sounds misleading to state that the CDF "combines independent measurements of an atmospheric vertical profile" – thus a single one – given that its extent (in space and time) is unknown and if later the formalism is extended "to include interpolation and coincidence errors, which must be considered when the profiles to be fused are measured on different vertical grids and at either different times or locations." The latter additionally implies that the indication of a single profile does not hold, in agreement with lines 26-28.

Eqs. (5), (10), and (12), are straightforwardly obtained from insertion of Eq. (8) of this work into Eqs. (5), (6), and (7) of Ceccherini et al., 2015, whereafter the need for non-singular matrices (or the generalized inverses of singular matrices) can be dropped from the CDF formalism. This more intuitive line of thought should be discussed, with appropriate reference to what has already been achieved in Ceccherini, 2022.

Given Sections 2.1 and 2.3, the conclusion of Section 2.2 is quite trivial and could be dropped.

A nice plus with respect to previous work is the brief expression for the full CM of the CDF as expressed in Eq. (14) and its extension in the presence of coincidence and interpolation errors by Eq. (32). Following up on the previous comment, these findings could be better situated with respect to the equations in Ceccherini et al., 2018.

In order to be exhaustive, several things are missing from Section 3. It would at least require (1) a paragraph on the IASI data retrieval (lines 150-151), (2) information on the definition of the IASI profiles (cf. Figure 1, partial columns for which layers?), (3) a view on the AKM and CM shapes for the IASI retrieval and a discussion of how representative these are for OE retrievals, (4) expressions for how eigenvalues and generalized inverses are calculated.

Given that (lines 159-161) "The distribution of the eigenvalues of S_{ni} is due to the fact that the AKM and the retrieval error CM provided to the users are compressed (Astoreca et al., 2017) and are reconstructed using the 6 largest eigenvalues of the Fisher information matrix." the result shown in Figure 3 does not come as a surprise. Figure 4 is then created for "the optimum number of eigenvalues" (please mention six explicitly). The most important conclusion of this work, however, is the following (last sentence of Section 4): "The use of the new CDF(2021) and operational CDF(2021) is recommended for data fusion processing, but the errors made with the old CDF(2015) do not appear to be important, even in the case of a significant data compression." Numbers should be given with this statement. In order to do so, Figure 4 should be reproduced for a range of generalized inverses (or numbers of eigenvalues), including the popular Moore-Penrose pseudo-inverse, to quantify what the effect of this choice is on the error that has been made by using CDF(2015). And finally, how does this error relate to the full CM S_f upon CDF of an increasing number of retrieved profiles? That should be the full merit of Section 3.

Technical corrections

Lines 23-25: "The method is equivalent to the simultaneous retrieval of all the measurements that are combined when the linear approximation of the forward models is appropriate in the variability range of the results of the individual retrievals." A proof or reference is needed for this statement.

References to (Ceccherini, 2021) should be updated to (Ceccherini, 2022).

Line 63: AKM and CM as a 'measure' for sensitivities and retrieval errors, respectively, does not sound correct. These are retrieval quantities that define / determine the sensitivities and errors.

Line 73: "With an iterative procedure that adds one by one the extra profiles to the fused product..." A proof or reference is needed for this statement.

The IASI data providers are not co-authors nor acknowledged, which seems inappropriate.