

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

## **Anonymous Referee #1**

### **Review**

#### **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.**

In the revised version of the paper, we changed “combines independent measurements of an atmospheric vertical profile” in “combines independent measurements of atmospheric vertical profiles” in the abstract. On the other hand, at line 15, where reference to “an atmospheric vertical profile” is still present, it is specified that the given definition refers to how the CDF was introduced in Ceccherini et al. (2015).

**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 nonsingular 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.**

We followed an approach that was driven by the new information published in the literature and developed a rigorous demonstration; however, the line of thought suggested by the reviewer is probably more intuitive. Therefore, in the revised version of the paper we added a paragraph at the end of Section 2.1 where the alternative line of thought is presented.

**Given Sections 2.1 and 2.3, the conclusion of Section 2.2 is quite trivial and could be dropped.** The last sentence of Section 2.2 wants to provide a comment to the previous Eq. (18) rather than a conclusion. In the revised version of the paper we modified this sentence to make this more evident.

**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.**

A sentence was added in the revised paper at the end of Section 2.3 as a comment to Eq. (32).

**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.**

In the revised version of the paper, we added all the information requested by the reviewer. We modified also Fig. 1 in order to show the a priori profile, the AKM and the errors.

Reference to McPeters et al., 2007 was added in order to describe the a priori information used in the retrieval of the IASI measurement.

References to the Python functions used to calculate the eigenvalues and the generalized inverses were added.

**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.**

In the revised version of the paper we explicitly mention six as the optimum number of eigenvalues, furthermore we specify that the calculation of the generalized inverse is made with the Moore-Penrose pseudo-inverse of a matrix using its singular value decomposition and a threshold for the eigenvalues (the Moore-Penrose pseudo inverse without a threshold does not have a physical meaning).

The effect of choosing a wrong threshold for the eigenvalues is shown in Fig. 3 and the comparison of this figure with Fig. 4 is sufficient to highlight that to be wrong in the number of eigenvalues of one unity produces an error comparable with the retrieval error. We added this consideration in the revised version of the paper at the end of the paragraph after Fig. 4. We prefer to make this comment and maintain Fig. 4 as simple as possible.

The question “how does this error relate to the full CM  $S_f$  upon CDF of an increasing number of retrieved profiles?” is quite important, however, it is difficult to provide examples that are valid for all the possible cases encountered with real data. The error of the fused profile decreases more slowly than the square root of the number of fusing profiles because: a) part of the information is used to increase the number of degrees of freedom and b) the contribution of the a priori errors to the components that are not measured does not decrease. On the other hand, the error due to the generalized inverse does in general decrease with the square root of the number of fusing profiles, but also in this case a lower decrease may be encountered if systematic effects are present. We considered making this comment to the paper, but we think that it would increase the complexity more than the information.

## 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.**

The proof is given in the Appendix of Ceccherini et al. (2015). We added the reference in the revised version of the paper. Furthermore, we rephrased the sentence.

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

In the revised version of the paper, we updated the references and also substituted CDF(2021) with CDF(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.**

In the revised version of the paper, we replaced “measure” with “determine”.

**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.**

In the revised version of the paper, a proof of this statement is given in Appendix A.

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

In the acknowledgments of the revised version of the paper, we thank EUMETSAT that is the originator and the distributor of the IASI data used in the paper.