Dear Referee,
dear Editor,

we would like to thank the referee for his interest in our work. In the following we reply to his comments (referee comments appear in blue font and our response in black font).

I make my congratulations to the authors for this very interesting paper that presents a way to combine L2 products of different co-located measurements of the same species.

Thanks!

This comment is just to suggest them to cite Warner et al. (2014) that, from what I can understand, used the same method proposed by the authors of this paper to combine CO products of AIRS and TES as well of AIRS and MLS.

Yes, thanks for this suggestion. In the revised manuscript we will discuss the Warner et al. (2014) approach and its similarities to our work. Warner et al. (2014) had the same principle idea as we, i.e. the application of a Kalman filter for combining different satellite sensor observations. However, there are also important differences.
(1) Warner et al. (2014) uses horizontal fields measured by AIRS with weak vertical details as the background and focuses on improving the vertical information for this large area by using the detailed vertical information provided very locally by the observations of TES (and MLS). In their method the vertical information comes mainly from TES (or MLS, see their Figs. 6, 7 and 11) and the horizontal information from AIRS. In our method we use two sensors, which both have good horizontal coverage (we do not need an analysis in the horizontal dimension), but different and rather synergetic vertical sensitivities. In our method we optimally combine the different vertical sensitivities and generate a combined observation that has more detailed vertical information than each of the two individual observations.
(2) In Warner et al. (2014) the combination is not made in a fully optimal sense (that would be quasi equivalent to the combined optimal retrieval products). They use a diagonal observational error covariance matrix (**R** in their Eq. 2) and a global statistics for the satellite sensors' noise errors and sensitivities (see their Figs 2 and 3 and the related text). In contrast, our method works with the individual noise errors and sensitivities of exactly the two observations that are combined, i.e. our method is rather similar to a combined optimal estimation retrieval product.

Then, as another method that uses the output of the individual retrievals to combine different co-located measurements, I suggest to cite the Complete Data Fusion method (Ceccherini et al., 2015). I have demonstrated that the method used by the authors of this paper is equivalent to the Complete Data Fusion method in the sense that starting from the formula of one method we can obtain, using algebraic operations, the formula of the other method. I have submitted the proof of this equivalence as a peer-reviewed comment to this paper to AMT in order that others can verify the proof as well. I hope that this peer-reviewed comment will be published in AMT Discussions as soon as possible, so that it is available also to the authors of this paper.

Ok, we will cite Ceccherini et al. (2015). We would like to thank the referee for the initiative of comparing the different approaches. We also think that this is useful and in the revised manuscript we will add some short discussions on this.

I have read the Ceccherini (2021) comment manuscript available since 26 April 2021 at https://amt.copernicus.org/preprints/amt-2021-98/ (submission date is 9 April 2021). This comment refers exclusively to the theoretical part of our work (the Appendix of our manuscript). I would like to suggest to the referee to consider the revised theoretical part of our manuscript for his comment. This revision has been elaborated in the context of the reply to Referee 1 and has been made available already on 29 March 2021 (https://doi.org/10.5194/amt-2021-31-AC1).

In the Appendix A of our revised theoretical part we distinguish two situations. We start with discussing the combination of two profile retrieval products (see Appendix A2.1). There the application of a Kalman filter is not needed and the combination can be easily achieved by using the a posteriori covariances of the two individual profile retrievals (see our Eq. A11). This completely avoids the problem of a potentially singular retrieval noise matrix. Our Eq. (A11) is the same as the second line of Eq. (10) of the Simone Ceccherini comment. So the Ceccherini (2021) comment paper could start already with its Eq. (10). The Ceccherini (2021) comment shows then the equivalence of Eq. (10) with Eq. (16), the Complete Data Fusion (CDF) method. However, I do not understand the advantage of Eq. (16) if compared to Eq. (10). Equation (10) needs less input than Eq. (16). The input for Eq. (10) is the retrieval products, the a priori data, the a priori covariances, and the a posteriori covariances. The input for Eq. (16) is the same input and in addition the two averaging kernels. Why should one use Eq. (16) (the CDF method) instead of Eq. (10), which is the same as our Eq. (A11)?

In Appendix A2.2 of our revised theoretical part we discuss the combination of a profile and a column data product. This is the problem, on which our work is focusing. It is only this situation, for which we suggest the application of the Kalman filter approach. We show the large similarity of the Kalman filter approach and a combined retrieval that uses the two individual measurements. To our understanding this combination of column data with profile data is not captured by the CDF method as written in Eq. (16) of Simone Ceccherini's comment. For a column observation, the profile averaging kernels and profile a posteriori covariances are not readily available (if they can be made available at all), but both are needed for the CDF (the respective Eq. 16).

In general and if data with different vertical representations (fine and coarse) are going to be combined, the CDF method (Eq. 16 of Simone Cecherini's comment) can only work on the coarse vertical representation. The averaging kernel and the a posteriori covariances of the fine gridded profile can be interpolated to the coarse grid (e.g., von Clarmann and Grabowski, 2007), but not vice versa. This means that when combining a profile product with a column product, the CDF method can (to my understanding) only generate a combined column product. In contrast, our Kalman filter-based approach can combine profile and column data and generate a profile observation that has an improved vertical sensitivity.

References:

Warner, J. X., Yang, R., Wei, Z., Carminati, F., Tangborn, A., Sun, Z., Lahoz, W., Attié, J.-L., El Amraoui, L., and Duncan, B.: Global carbon monoxide products from combined AIRS, TES and MLS measurements on A-train satellites, Atmos. Chem. Phys., 14, 103–114, https://doi.org/10.5194/acp-14-103-2014, 2014.

Ceccherini, S., Carli, B., and Raspollini, P.: Equivalence of data fusion and simultaneous retrieval, Opt. Express, 23, 8476-8488, 2015.

von Clarmann, T. and Grabowski, U.: Elimination of hidden a priori information from remotely sensed profile data, Atmos. Chem. Phys., 7, 397–408, https://doi.org/10.5194/acp-7-397-2007, 2007