The authors thank the reviewers for the constructive comments which have helped to improve the paper.

Answer to Reviewer #1.

The manuscript 'Merged ozone profiles from four MIPAS processors' by A. Laeng et al. presents the method of merging MIPAS ozone retrievals derived from four independent retrieval algorithms. The manuscript requires a major revision. I found it very difficult to read and follow the manuscript. I think it is partially due to the way the manuscript is organized. I would suggest authors to rearrange the order in the paper (see my comments below).

Specifically, I feel that authors should start Section 2 by describing the way they defined the error covariance matrices (currently Sections 4 and 3).
We rearranged the material; part of the introduction was moved into "Merging approach" section. We strongly feel that first the idea how the merging is done should be explained, then the details of how different ingredients of the merging formula (covariance matrices included) are calculated should be exposed.

I also feel that authors use too many equations in the manuscript, some of them are not really needed. Instead, I would suggest authors to describe some of the results in words.
We added the description of results in words, but kept all the original equations in order to avoid the ambiguity of language and provide precise description.

In some places, it is difficult to understand statements; some statements are vague and require more explanation and references, while other statements are irrelevant to the topic of the study and can be skipped. Finally, I feel that authors need to put more efforts to justify their approach of merging data and provide a stronger validation.
We enlarged the section justifying the approach and explaining while taking the weighting bases on the covariance matrices is mathematically clean. We deleted all the sentence you suggested (see below), and added description of equations in words.

Specific Comments:
Page 2, lines 50-54: It says that only KIT processor accounts for horizontal temperature gradients, however, it was noted above that the Bologna processor uses a "full 2D-approach". I thought that the 2D approach should take care about horizontal inhomogeneity (gradients) in T, O3 and other parameters. Please, comment on this.

The true atmosphere varies in all three dimensions. Variations in the altitude domain are accounted for by all four processors. Bologna and KIT also tackle horizontal variation, each by a specific approximation. Bologna allows a second dimension which is horizontal variation in the orbit plane (which is not always parallel to the line of sight). KIT allows a linear horizontal variation along the true line of sight direction. We have changed the text accordingly.

Page 3, line 61: Why did you merge data only for two years 2007 and 2008? Is there is a specific reason for that? Do you plan to merge data for entire time period of MIPAS operation? Please, clarify that in the text.
This merging exercise was performed on the datasets provided for MIPAS Round Robin study within European Ozone Climate Change Initiative project. Some of the four MIPAS teams provided only 2007-2008 data for this study because these are research processors and not operational processors. Future merging activities will depend on funding. We added the explanation in the text.

Page 3, line 63: It is not clear from the context what are the "corresponding" covariance matrices. I would suggest to call them "statistical covariance matrices" (as you do in Section 4) and add in parenthesis "(see Section X below for the details)".

To each profile corresponds a covariance matrix. We first introduce the idea of how the merging is performed, at this stage no need to go into details that the covariance matrices will be obtained in a statistical way. In this context, "corresponding covariance matrix" means "covariance matrix corresponding to this profile". For clarity, we now write '"error covariance matrix" throughout to avoid confusions with natural variability.

Page 3, line 66: I don't quite understand the statement: ": the source measurements can be considered as nearly independent with respect to the primary measurement error." Do you mean that the errors can be considered as independent because processors use different spectral micro-windows?

Yes. We mean that the source measurements provide almost independent information, these coming from different parts of the spectra. Thus, w.r.t. noise, the four retrievals can be treated as independent measurements.

Please, edit this sentence.

Done.

Page 3, line 67: I think that the statement "to be better" requires further explanation: do you mean better precision, accuracy or information content?

We meant "precision" and changed the text accordingly.

Page 3, lines 70-72: Which dataset do you refer as 'climatological datasets'? I think that this discussion is irrelevant to the topic of the study and can be removed.

We took the sentence about "climatological datasets" out, but left the discussion. We feel it should be explained why no *a priori* statement can be made about averaging of the biases of the parent datasets in the process of the merging.

Page 3, lines 73-74: Usually, term "biases" is used when one compares two datasets. In the context here do you mean biases relative to the ensemble mean or biases relative to the specific reference?

The term bias is also used for "deviation from the truth", and we use it here in this sense. We changed the text accordingly.

Page 3, lines 75-78: I don't quite agree with the statement that 'different degradation of instruments' is completely irrelevant. For instance, for the same instrument degradation could be different in different spectral ranges. Thus, if two algorithms use measurements from different spectral ranges, one could be more sensitive to instrumental degradation than another. If you make this statement specifically about MIPAS instrument, then please make it clear. I would strongly recommend adding a reference to the study that demonstrates that MIPAS degradation is similar in all spectral windows used by different groups for the O3 retrievals.

We indeed refer to MIPAS instrument only, we added a reference to A. Kleinert et al, "MIPAS Level 1B algorithms overview: operational processing and characterization", Atmos. Chem. Phys., 7, 1395–1406, 2007 to justify our assumption.

Section 2, page 3: The section starts by stating that the value of the merged profile at any given layer is a weighed sum of O3 values from all four parent processors and all levels. However, it remains unclear (until Sections 4 and 3) what is actually used as weights for the merged profile, which makes it very difficult to follow the paper. I think it would be better to provide a definition of errors first, and then move to the merging technique (put sections 4 and 3 first, and then move on with Section 2).

The weights are determined as the inverse error covariance matrices where the inter-processor correlations are taken into account. We rewrote the explanation.
We prefer to keep our order of presenting the information: first explain how the merging is done, then to go into details, and not the other way around.

Also, the motivation for considering contributions from all the layers is not clearly explained in the text. The standard approach would be to simply average O3 values from 4 processors at the given layer. You propose to weight contributions from each processor based on uncertainties, and additionally account for out-of layer contributions. You should clearly explain motivation for doing this. Why the contributions from other levels are so important for the merged product?
This is done because in retrieved atmospheric profiles the layers are not independent. We add a statement on this on the paper.

Page 4, equation 1: I don't quite understand this equation. The x_merged is a merged profile, and thus should be a vector with n elements (where I assume n is a number of altitude levels (n should be introduced in the text)).
Done, we introduced ***n*** in the text.
If matrix e has dimensions [n x n], then the matrix (e e e e) will have dimensions [n x 4n],
This is correct
and the final product of eq. 1 will be [n x n] matrix, and not a vector [n]. Please, correct this equation.
This is not correct.
The number of columns in a product of matrices (whatever is the number of entries) is the number of column of the last matrix, and the matrix $(\boldsymbol{x_1},\boldsymbol{x_2},\boldsymbol{x_3},\boldsymbol{x_4})^T$ has one column, because $\boldsymbol{x_i}$ vectors are n-dimensional column vectors.

Pages 4-5, lines 104-120: I feel that this discussion is irrelevant to the main topic of the paper.
We removed this paragraph.

Essentially what was calculated and evaluated in this study is variability relative to ensemble mean. It is not a random or systematic error. As authors noted in Sect.1, there are many common features in considered MIPAS retrievals, thus analysis of noise relative to ensemble mean will tell you very little about the instrument systematic errors.
This is true but it is not intended to derive information on systematic errors from this.

So I am not sure how you can isolate a random component here, and specifically it's not clear to me how you define a systematic component in eq. 6 (I guess it is assumed as zero). On page 6, line 140, authors state that "at fixed height systematic error component is constant". I don't quite agree with this statement, there are many examples of systematic errors that vary with season or over the lifetime of the sensor.
The distinction between systematic and random errors is not really relevant for our approach. In our revised version, we distinguish only between bias and errors with a random (=variation with time) component. Since we determine the bias from the data and subtract it to base the method on debiased profiles, it is not necessary to make any prior assumptions on the systematic part of the error.

I think that the statistical error considered in this study is a composite of random and systematic components. I would strongly advice to not use term "random".
We rewrote this part.

Page 5, line 116-117: Equation 5 assumes that the errors are additive. I am not quite sure what is the meaning of the following statement: "Our choice is to neglect both these facts : : :"

We rephrased this text.

Page 5, line 125-130: Authors discuss possible reasons for error correlations between a pair of processors. I would assume that a priori assumptions could play a role here. A priori assumptions together with the assumed instrumental noise define the amplitude of the retrieved noise. Additionally, a priori covariance matrix defines a range of layers where O3 variability should be correlated. I would think that if two processors use the same a priori information, you might expect stronger correlations.

Since we determine the error correlation empirically, assumptions on the causes of the error correlations are irrelevant. By the way, neither ESA nor KIT processors use a priori information in an "optimal estimation" sense

Page 6: Authors provide numerous lengthy equations, but it would be nice to summarize some results in words.

Done

Essentially you analyze the noise relative to ensemble mean by considering retrievals for all geolocations and entire time period from 2007-2008. I think you don't need to reproduce well-known equations for the expectations and correlations.

We deleted this part of text, as suggested.

You need to clearly define the way you compute anomalies. And I think it is not clearly explained or defined here. Do you derive correlation coefficients by considering all data over all locations or do you sort data by season and latitude bands?

This is defined by the Equation 11 (old Equation 12). Correlation coefficients are calculated on the whole sample of data.

It is interesting how robust these covariance matrices when you rearrange data. Did you check if these covariance matrices representative for all seasons and latitude bands?

Here it is important to note that all differences calculated in Eq. 11 (old Eq. 12) are differences of values related to the same geolocation k. Since no differences between profiles of different geolocations are calculated, no contribution by natural variability adds on $\boldsymbol{r}_{ij}^{pq}$.

Page 7, lines 171-173: The high inter-level correlations of retrieval noise between a pair of instruments do not necessarily suggest that the retrieved ozone from one processor at altitude z1 is connected to retrieved ozone from another processor at altitude z10, especially if these two layers separated in altitude. You should have some physical reasons to believe that errors at 20 km are connected with errors at 50 km.

The reason for this is that in a retrieved vertical profiles the levels are not independent. These correlations are attributed to the typical limb sounding error propagation patterns.

Section 4, page 8, lines 180-200: I think that these two paragraph are completely irrelevant. Here, authors discuss the error covariance matrices reported by different MIPAS algorithms and provide equations for these error covariance matrices, but these error covariance matrices have not been used in this study. Then what is the purpose of describing these covariance matrices? I think that this information is irrelevant and distractive for readers. I think you should limit this discussion by saying that due to a lack of reported retrieval metrics you decided to move forward with the statistical covariance matrices.

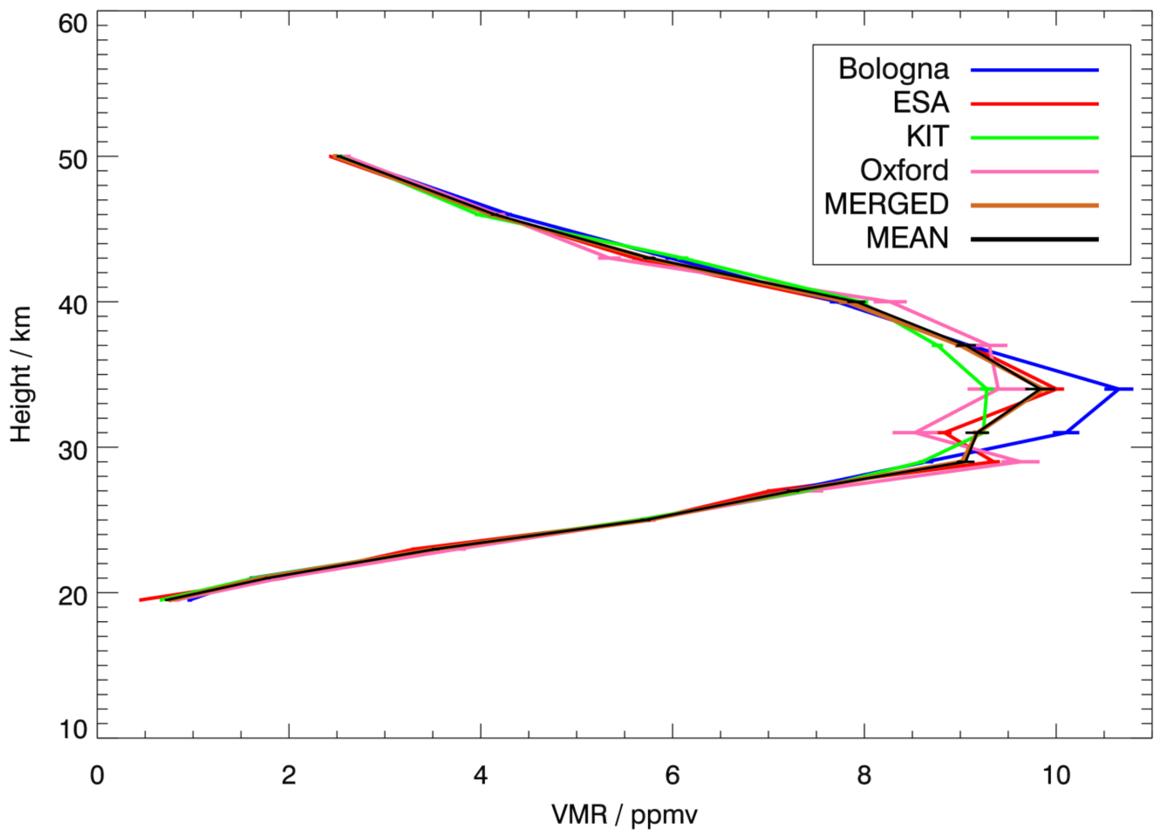
Done.

Section 4, page 8, equation 17: if I understood this equation correctly, the empirical matrices are constructed by calculating covariance between anomalies at different layers over entire range of

geolocations. Here, I have two questions. How sensitive these covariance matrices to the order of data (by time, geolocations etc.)? It's well known fact that the morphology of vertical ozone distribution is quite different in the tropics from that in mid-latitudes. In mid-latitudes the ozone peak is broader and is located lower compared to the tropics. I would expect the inter-level correlation to be different if you sort data by latitudes.
The statistical covariance matrices are calculated at the subsample of ozone profiles from the summer tropical stratosphere. This is done in order to minimize the impact of natural atmospheric variability, so yes, if you calculate the statistical covariance matrices via this formula at the other latitude bands, the result will be different, because there the influence of natural atmospheric variability will be larger. We consider our choice as the most adequate to infer inter-level error correlations, because here natural variability is minimal, and the contribution of the retrieval errors (as opposed to the natural correlations) to the total variability is largest

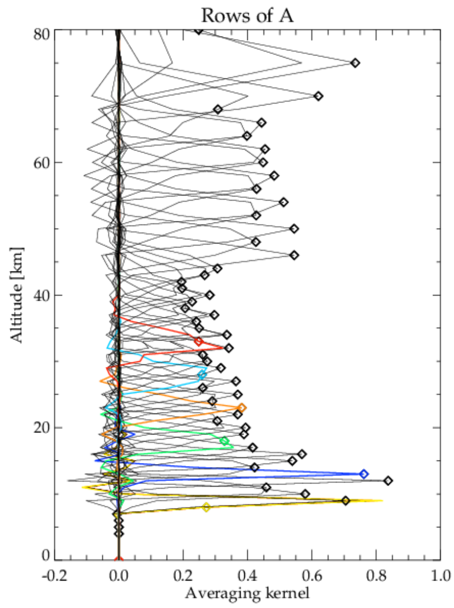
I am wondering how changes in constructing these empirical covariance matrices affect your final product. For instance, if you assume no inter-level correlation of errors, how will your merge product change? This is something that I would encourage authors to examine. This would demonstrate a benefit of this approach compare to the standard approach when merged value at the specific altitude is a simple weighted sum of parent values from the same layer.

By its nature, the limb-sounding retrieval (sensitivity varies most rapidly in the vertical dimension). Below is a figure showing the merged profile, parent profiles, the simple mean of parent profiles, and weighted mean of parent profiles.



How large is the contribution at any specific layer (in %) from other layers?

This is shown by averaging kernel matrix, below is an example of averaging kernels matrix for KIT processor.



Section 5. I feel that validation section can be greatly improved in the revised manuscript. The main difference of this proposed technique of merging several datasets is in accounting for out-of-layer contributions (by considering full covariance matrices including off-diagonal elements).

The main difference of the proposed technique is the careful weighting based on the uncertainties of parent profiles.

I would explore and validate this deeper in Section 5. Does your merged product perform better than the simple average? It would be nice if you can provide some stats on the contributions from other layers? Do you see a specific altitude range or latitude band or time period where the out-of-layer contribution is the largest? Do you see improvement in the merged product compare to the simple ensemble mean in the same places and time periods?

The Figure on previous page is the Figure 3 from the paper with the simple average curve for the comparison. We prefer not to include this version of the figure in the paper because merging by averaging is the a deficient approach.

Figure 3, page 9, lines 216-221: It is not clear to me what is the purpose of Figure 3 and the corresponding discussion. Why did you show a sonde profile if you believe it is biased? Could you find another sonde station that you trust?

We removed the ozonesonde profile from the Figure 3, and instead we added the error bars of the parent and merged profile. The purpose of this figure is to show an example of parent and merged profiles. The purpose of this study is improvement of the precision

Page 10, lines 233-235: I do not see that the merged product have smaller biases that any of four parent profiles. The orange line, which represents the merged profile, always goes close to the ensemble mean. Specifically, at altitudes above the ozone peak (_42 km) all four processors are biased high against ACE, and merged values agrees with KIT; around the ozone peak (39 km) the merged value is right in the middle; and below the peak (36 km) the merged value is close to zero, which is expected because ESA and KIT have negative biases and Oxford and Bologna positive. Please, explain what did you mean here?

Despite the fact that this study is not about the bias reduction, it was found that at the height range 33-35 km, the brown (merged) curve of the merged product is closest to zero. For other altitudes, the merged profile follows the parent profile.

Technical comments:
Page 1, line 2: should be 'infrared';
Done.
Page 1, line 4: I suggest to replace colon (:) with 'developed by';
Done.
Page 1, lines 9-10: Last two sentences of the abstract require some revision. I would suggest to replace them with 'Hence, information content of the merged product is greater and the precision is better than those of any parent dataset.'
Done.
Page 4, lines 95-96: I would recommend adding here "(see section 3)".
Done.
Page 4, lines 91-92: Replace Sx1 etc. with S11, because they are calculated in the same manner as any Sij, but for the case i=j.
This is true, but here we prefer to keep the standard notation for the covariance matrices on the diagonal. We will say that they are calculated as S_{11} is stated later in the text.
Page 5, line 109: Should be "example".
Page 5, line 110: Please add "a" to "a strong random component";
Following your suggestion, these parts of the text were removed.
Page 5, line 123-124: Please, re-phrase this sentence. Do you mean "algorithmic differences" when saying "or if the differences in the retrieval algorithm dominate";
Yes, we mean the differences in the algorithms. We rephrased this paragraph.
Page 5, line 126: Add "Bologna's correlation coefficients".
Done
Page 7, line 171: Replace "errors are non-negligibly correlated" with "errors are correlated".
Done.
Page 6, equation 12, the matrix "r" is in lowercase, while on page 4 it defined as "R" (uppercase). Please, use consistent terms throughout the paper.
Done

Answer to Reviewer #2.

This paper provides the theoretical foundation and first results for merging the four competing MIPAS ozone processors in what the authors call a "mathematically clean" way. The merging methodology is useful, but there are several shortcomings with the text, and in the validation and results. Overall the paper needs more explanation around the mathematical methodology and the meaningfulness of the calculations and results. The grammatical composition is also weak throughout the text, and really needs the engagement of the impressive consortium of co-authors to raise the level of quality of the writing. If this, and the issues raised below, can be satisfactorily addressed, this paper could be suitable for publication in AMT.
The text was improved and proofread by an English native speaker co-coauthor.

-Abstract: - What does it mean that the information content is more important? –
We rephrased it to "'Hence, information content of the merged product is greater and the precision is better than those of any parent dataset".
The phrase "parent profile" probably needs definition as it's not an overly common term.

Done.

- A statement is made about the change in relative bias with ACE-FTS and about the absolute bias with respect to MLS. Please make these statements consistent; the change in relative bias seems to be more relevant.

All the statements about the bias are in absolute terms. We added also the values in relative terms.

- Why is the study performed with only 2 years of data? It seems that it should be quite easy to apply the methodology to the entire data set and make it publically available. This seems especially important in light of the fact that one of the main motivations for this work, as claimed by the authors, is the confusion in the scientific community about which product is "better" and which should be used.

This merging exercise was performed on the datasets provided for MIPAS Round Robin exercise within European Ozone Climate Change Initiative project. Some of the participating processors are research processors rather than operational processors, and thus for the merging study data for the entire MIPAS mission were not available. There are no plans to merge larger datasets, the present paper aims to demonstrate the feasibility of the approach. We added corresponding explanation into the text.

- The different choice of microwindows for the four processors is a relatively important aspect of this approach, so it would be good to see an indication of just how 'independent' these choices are. What is the actual overlap in the source measurements? Surely it is not zero.

Below is the table of the microwindows used by the four processors, from (Laeng et al, 2015). This table has now been included into the manuscript.

Table A.2
Ozone microwindows of four MIPAS processors (some spectral gridpoints within the microwindows are masked).

	Microwindow (cm^{-1})	Altitude range (km)
ESA	729.25–732.25	15–42
	756.625–759.625	9–36
	1043.875–1046.875	27–68
	1117.000–1120.000	6–42
	1123.5625–1126.5625	9–68
Oxford	729.2500–732.25	15.0–46.0
	756.6250–759.625	7.5–37
	1043.8750–1046.875	27–70
	1117.000–1120.000	6–46
	1123.5625–1126.5625	7.5–70
Bologna	686.688–689.688	60–68
	689.750–692.750	30–52
	731.188–734.188	6–36
	790.625–793.625	27–47
	1036.313–1039.313	33–68
	1071.875–1074.875	6–39
	1651.000–1654.000	6–36
	1682.688–1685.688	27–68
KIT	760.6875–764.3125	6–68
	766.8750–767.1875	6–68
	776.1875–777.9375	6–68
	781.0000–782.8750	6–68
	787.0000–788.0000	6–68
	1029.0000–1031.0000	36–68
	1038.0000–1039.0000	36–68

It is noted later in the paper (line 122) that at least 2 of the processors use identical microwindows. How can this be taken into account? In this case it's not clear that this merging is a useful exercise, unless the errors from the retrieval algorithms are random somehow?

In order not to have to rely on assumptions about inter-processor error correlations, these have been analyzed statistically. This is described in the Section 2 of the paper. The considerations of these correlations accounts for the problem mentionned.

- The important conclusions of the comparison of these products in Laeng et al., 2015, should be summarized and potentially referenced in the discussion of the results. -
Done.

Do the authors really mean that the merging weights depend on the "quality" of the error estimates? Or simply the magnitude of the error? Is a small error a "better error estimate"? Please clarify.
Merging weights depend on the magnitude of the error: the smaller is the error, the bigger is the input. The text is corrected accordingly.

- The first paragraph of Section 3 seems mostly like a random (no pun intended) collection of facts and the point is not clear.
We removed this paragraph.

- Equation 5. Can any statements be made about the impact of this assumption?
We consider Eq. 5 a terminological convention rather than an assumption. We have corrected the text accordingly and present the material in a (hopefully) clearer way.

- Equation 6 should be typeset as an equation with an equality (i.e. set to nepsilon_{random})
Done.

-Line 140: Why can it be assumed that the systematic error component for each processor is constant?
Our original manuscript was misleading w.r.t. systematic errors. In our context the systematicity in the time domain is irrelevant. Only systematicity (correlations) between the processors is relevant. Biases between the processors (3^{rd} and 6^{th} terms in the numerator of the original Eq. 12) are subtracted. Remaining correlations are evaluated statistically. The text has been changed accordingly.

-Is Figure 1 calculated with the entire 2007-2008 data set? If so, the sharpness of some of the structures is difficult to understand. Comments on this would be insightful.
Yes, the Figure 1 is calculated with the entire 2007-2008 dataset

$$r_{ij}^{pq} = \frac{\sum_{k=1}^{N}(\hat{x}_{i,k}^{p} - \overline{x}_{k}^{p} - \frac{1}{N}\sum_{l=1}^{N}(\hat{x}_{i,l}^{p} - \overline{x}_{l}^{p}))(\hat{x}_{j,k}^{q} - \overline{x}_{k}^{q} - \frac{1}{N}\sum_{l=1}^{N}(\hat{x}_{j,l}^{q} - \overline{x}_{l}^{q}))}{\sqrt{\sum_{k=1}^{N}(\hat{x}_{i,k}^{p} - \overline{x}_{k}^{p} - \frac{1}{N}\sum_{l=1}^{N}(\hat{x}_{i,l}^{p} - \overline{x}_{l}^{p}))^{2}(\hat{x}_{j,k}^{q} - \overline{x}_{k}^{q} - \frac{1}{N}\sum_{l=1}^{N}(\hat{x}_{j,l}^{q} - \overline{x}_{l}^{q}))^{2}}} \tag{11}$$

In this formula, the third term in each bracket is the bias of corresponding processor, by taking it out of the first term we obtain a debiased profile, and the second term in the bracket is the mean around which the variation of debiased profiles is calculated.

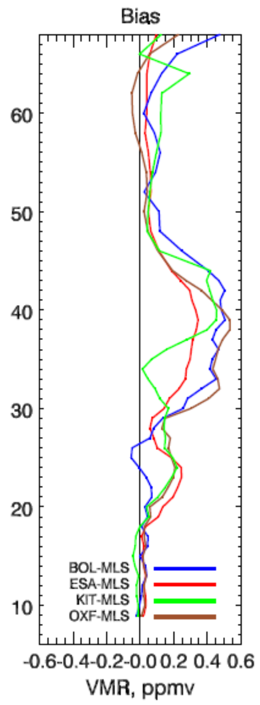


Fig: validation of the four processors wrt MLS from (Laeng et al, 2015)).

This gives a way to interpret the sharpness of the structures at the Figure 1 in the paper. The figure on the left is validation of individual processors against MLS.

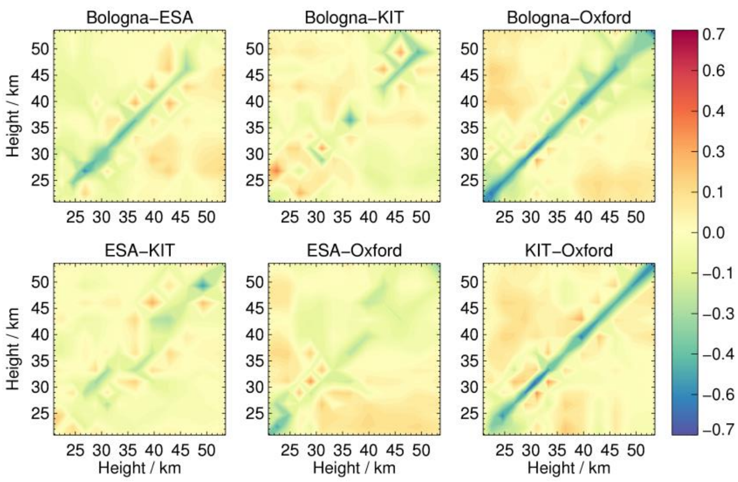
- Similar biases give a negative correlation: the more the bias are similar, the larger is the absolute value of this anti-correlation.
- Different biases give positive correlation; more the biases are different, the larger is the absolute value of the correlation.

<u>Example for the same heights, i.e. diagonal values of panels at Fig.1</u>
For example, at 20-35 km, for the same heights, the bias of KIT and Oxford wrt MLS is almost identitcal, which translates into blue (i.e. higher anti-corellated) peak at the diagonal of KIT-Oxford panel, same for 42-55 km heights.

<u>Example for the different heights, i.e. off-diagonal values at the panels at the Fig. 1</u>
The Bologna bias at 42 km is very different from KIT bias at 35 km (this is the highest deviation of the biases at the figure on the left, which translates into the most red spot of all six panels at the fig. 1, at the upper lefty part of the Bologna-KIT panel.

We should keep in mind that MLS itself tends to bias high at the upper stratosphere (see X. Yan et al.: Validation of Aura MLS …, Atmos. Meas. Tech., 9, 3547–3566, 2016), which could affect the absolute values of the correlation coefficients on Figure 1 of the paper wrt absolute values of the bias on the Figure at the left.

For comparison, below is the same Figure calculated on the summer tropical profiles only. As expected, the sharpness in KIT-Oxford and Bologna-Oxford are still present.



-Line 193: The term “statistical covariance matrix” is confusing. It seems that it means the authors calculate it directly from the data set and later in the text (eg. Line 201) they refer to the same as “empirical covariance matrix”. Also, are “genuine” and “analytic” covariance matrices the same?
We changed the terminology in a consistent way: ‘genuine” and “analytical” is now called just “covariance matrix of the profile”, while “statistical” and “empirical” is now called statistical throughout the text.

-It is not clear how using an analytic covariance can cause the intercorrelation matrix to be singular. Please explain.
Singular Value Decomposition of the term in the first brackets of Equation 1 was failing when original covariance matrices were taken into it, this for both ESA and KIT processors. Following the suggestion of the first referee, we removed these two paragraphs from the text.

-Figure 3 with the ozonesonde profile adds very little to the discussion and nothing of substance is discussed. To simply show an example of the merging, it would be better to provide a plot showing the error bars on the single profiles and the resulting error bars on the merged profile.
Done

-Line 230 and following: The authors state that the merging is “not supposed to remove the bias”; however, the merging will of course change the bias through the averaging and this could result in an improved situation depending on the sign and magnitude of the existing biases in the four products. This should be carefully explained given what is known from the independent validation of the individual products.
This is a very good point. But our merged product is a weighted mean with inputs from all the layers of all the processors coming into each value, and the correlation coefficient taking different signs. The result depend not only on the sign of the bias but also on the sign of the correlation coefficients. We feel that trying to interpret the comparison results from this point of view would be an over-interpretation.

-The discussion around the ACE-FTS and MLS comparisons details several altitude regions where the precision and/or bias is “better” in the merged product. In general however, it is not clear whether it is meaningfully better (in some cases the differences are quite small). The authors should be able to quantitatively state whether or not this is the case. Additionally, it is important to know if these comparisons and associated conclusions hold over all latitudes where the shape and magnitude of the ozone profile varies considerably. At least with the MLS data set, the sampling should be sufficient to test this.
With both comparison instruments, we have run the comparisons in 6 latitude bands, and found the the overall mean comparison is representative. We added the percentage values of the bias around ozone vmr peak in the text.

- The conclusions need to address the availability of the data product and give a recommendation to the community as to whether or not this data should be used as a general replacement for any specific processor in scientific studies, or if it is only of use in specific cases.
Done.
Minor/editorial points: Undefined acronyms in the abstract and throughout
Line 12: comma splice
Done
Line 31: comma required after “corresponding period”
Done
Is there no publication reference for the Oxford product?
There is no per-reviewed publication aside Laeng et al (2015)

Line 38: unspecific subject. Please rephrase to something like "The existence of these four products often leads to confusion: : :"
Done
Line 39: What is a "homogenized description"?
Description with unified notation performed by T. von Clarmann for Ozone_cci Project. We changed it into "unified".
Line 45: "but" not grammatically correct
Done
Line 63: comma splice
Done
Equation 3: It should be noted that the calculation of R is discussed below.
Done
Line 109: "A typical example" (note spelling)
Done
Line 131: Given values of i, j in brackets are typeset in a confusing way – almost looks like q is a function?
Brackets replaced with comma.
Line 151: Do you mean "We note that N is the number of profiles : : :"?
We are introducing notation; changed into "we note N …"
Equation 11: use an equality
Done
Line 217/219: Use conventional degrees symbols
Done

Bibliography:

Laeng, A., Hubert, D., Verhoelst, T., von Clarmann, T., Dinelli, B. M., Dudhia, A., Raspollini, P., Stiller, G., Grabowski, U., Keppens, A., Kiefer, M., Sofieva, V., Froidevaux, L., Walker, K. A., Lambert, J.-C., and Zehner, C.: *The Ozone Climate Change Initiative: comparison of four Level 2 Processors for the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS)*, Remote Sensing of Environment., 162 (2015) 316–343.

Raspollini, P., Arnone, E., Barbara, F., Carli, B., Castelli, E., Ceccherini, S., et al. (2014). *Comparison of the MIPAS products obtained by four different level 2 processors*, Annals of Geophysics, 56.