Atmos. Meas. Tech. Discuss., 7, C5336–C5349, 2015
www.atmos-meas-tech-discuss.net/7/C5336/2015/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2" *by* A. Keppens et al.

A. Keppens et al.

arno.keppens@aeronomie.be

Received and published: 8 April 2015

Structure: (1) comments from referees, (2) author's response, (3) author's changes in manuscript

Anonymous Referee #1 (Received and published: 24 December 2014): Specific comments:

1. (1) Abstract: all acronyms (like ESA, KNMI, WMO etc.) should be spelled out when they are used in the manuscript for the first time. (2) The authors agree that the acronyms should be spelled out and have modified the manuscript accordingly. (3)

The abstract and the third paragraph of the introduction section have been modified accordingly for the following acronyms: ESA, KNMI, RAL, MetOp-A, GOME-2, WMO, GCOS, and EUMETSAT.

2. (1) P. 11502, par. 1, l. 5-9: I don't quite understand the statement "...the relative DFS variation is much smaller...", because earlier on p. 11501, l.16 it is written that "The DFS difference between OPERA and RAL is remarkable". Please, clarify that. (2) The first statement on relative DFS variation is on its solar zenith angle dependence (which is quite small for both algorithms), while the second statement is on the (indeed remarkable) difference in DFS between both algorithms, irrespective of the SZA. The corresponding text might however not have been fully clear and has therefore been improved. (3) "As the relative DFS variation is much smaller than that of the MQQ. . ." has been replaced by "As the SZA dependence of the DFS is much smaller than that of the MQQ. . ."

3. (1) p.11508, par. 1, l. 5: I would suggest to replace the word "unrealistic" and list objective, scientific criteria used to filter data (2-3 sigma from climatology etc.). (2) The scientific criteria as exemplified by the reviewer are in fact not what was intended here. The authors rather meant "unphysical" and have modified the text to make this clearer. (3) "unrealistic" has been replaced by "unphysical (e.g. negative or NAN)"

4. (1) P11510, sec. 6.2, l. 19: There is a formula for computing space-time distance between two profiles. However, it is not numbered as all other equations. Please, do that. (2) The distance formula is an in-line equation that does not require numbering. Due to the AMTD text and page formatting however, the formula misleadingly appeared to be of the numbered type, with the number missing. This flaw should not be present in the AMT manuscript formatting. (3) No changes have been made.

5. (1) Sec 6.4. I found this section very confusing; I think it requires some work to clearly present the method (see specific comments below). (2) Responses are given for the specific comments below (6 to 12). (3) Manuscript changes have been made in

relation to the specific comments below (6 to 12).

6. (1) P. 11512, l.3-6: Another approach to smooth the ground-based profiles is suggested, that requires interpolation of the rows of the averaging kernel matrix onto the fine vertical scale. The motivation to use this approach is not clear for the reviewer. The averaging kernels are scale dependent retrieval characteristics, and interpolation on the fine vertical scale can introduce some additional uncertainties. Please, explain the reason/reasons for adapting this approach in the study. (2) The authors agree that the main reason for using the fine resolution smoothing method was not explicitly mentioned, but only referred to. An additional clarification has therefore been added to the text. (3) "which maximally exploits the high-resolution reference measurement without adding information to the retrieval data" has been introduced before the fine resolution smoothing reference of (Ridolfi et al., 2006).

7. (1) P. 11512, l. 22-26: From this sentence it is not clear that you mean unit conversion of correlative ground-based data and not satellite profiles. Please, change this sentence to make it clear. (2) The authors agree and have inserted a clarification in the text accordingly. (3) "unit conversion" has been replaced by "unit conversion of ground-based reference data"

8. (1) P. 11513, l. 1-2: It states that nine different approaches to covert correlative data had been used. Unfortunately, it is very difficult to see all approaches. I would suggest listing all approaches and numbering them from 1 to 9. So, it would be easy for readers to follow you, and, additionally, it might help you to organize results listed in section 7.2. (2) The authors agree that the distinction between the nine different evaluation methods might not have been clear straightforwardly from the text. It has therefore been decided to number the approaches using Roman numerals and correspondingly refer to the boxes in the sixth part of the validation scheme. (3) Roman numerals have been added in the listed text to denote the nine different comparison approaches. The preceding sentence has been extended to announce this numbering: "...resulting in nine evaluation approaches to be studied in total (corresponding to the nine boxes with

black text in the sixth part of Fig. A1), as denoted by Roman numerals:"

9. (1) P. 11513-11514, eq. 9-10, 12-13: what is x_p here? Is it a priori profile used in the satellite retrieval algorithm or ground-based mean? Please, specify. (2) x_p refers to the prior atmospheric state profile as in Eq. (1). This should be clear from the explicit reference to Eq. (1) in the sentence preceding Eqs. (9) and (10). The authors would therefore like to keep the text unchanged. (3) No changes have been made.

10. (1) P. 11513-11514, eq. 9-10, 12-13: In these equations some terms have bars and some do not. Does the bar on the top mean the coarse vertical resolution? Please, specify that in the text. I would also suggest to skip subscript 'g' (in x_g), since in this section only unit conversion and smoothing of the ground-based data are discussed. (2) The upper bar indeed means that the reference profile has been regridded to the coarse satellite grid. This information was already included in the text in lines 3 and 4 of page 11514 and is therefore not repeated. The authors moreover would prefer to maintain the subscript "g" to denote ground-based reference profiles, in order to keep notation consistency throughout the entire manuscript. Accordingly, only general profile manipulation statements (as in Appendix B) lack a subscript. (3) No changes have been made.

11. (1) P. 11514, eq. 11: Please, define Delta_L used in eq. 11. It had not been introduced. (2) The authors agree that the definition of Delta_L was missing here and have therefore introduced it in the text. (3) The sentence preceding Eq. (12) has been extended using "with Delta_L representing the layer thickness in km or log-pressure"

12. (1) P. 11514, eq. 12: Please, define smoothed profiles by adding subscription/superscription to the corresponding terms. Right now in eq. 12 the term on the left side is the same as the second factor of the first term on the right side of the equation. Please, keep consistent definitions throughout the manuscript. (2) As the Calisesi regridding technique also introduces some kind of profile smoothing due to mass conservation, the authors had decided to use an upper bar for both pseudo-inverse regridded

and averaging kernel smoothed ground-based reference profiles. The authors however agree that this approach may be misleading (especially in Eq. (12)) and have therefore correspondingly decided to denote AK smoothed profiles by an upper tilde ($\sim$), while maintaining the notation of merely regridded profiles with a bar. (3) Eqs. (9) to (15) and the guiding text have been updated to show an upper tilde for AK smoothed profiles.

13. (1) P. 11516, l. 9-14: Q has not been defined. (2) It was the authors' intention to define Q implicitly as a quantile on line 9 of page 11516, but this might have been somewhat too implicit to come out clearly. The symbol Q has therefore been repeated in the text at the first mentioning of the use of quantiles. (3) "median (50 % quantile) of the set of its absolute differences Q50. . ." has been replaced by "median (50 % quantile Q) of the set of its absolute differences, i.e. Q50. . ."

14. (1) P. 11516, l.21-22: Please, explain how re-gridded ground-based uncertainties were computed. Measurement uncertainties are closely tight to the vertical scale used in the retrieval algorithm, thus a simple vertical re-gridding from the fine scale to the coarse scale can lead to overestimation of the error. (2) The authors agree that it was not mentioned in the text how the ground-based uncertainty profiles are determined on the coarse satellite grid. As for the atmospheric state profiles, this is achieved by use of the Calisesi regridding approach. This has now been formulated explicitly in the text. (3) "Next to that, combined satellite and ground-based uncertainty profiles are assessed in relation to the spread." has been replaced by "Next to that, combined satellite and ground-based uncertainty profiles – which are also regridded using the method of Calisesi et al. (2005) – are assessed in relation to the spread."

15. (1) P. 11518, l. 5-8: It is not quite clear what is 'former' and 'latter' here. It would be better to replace them by 'fine' and 'coarse resolution' approaches. (2) The authors agree that the use of "former" and "latter" reduced the clarity of the text. As suggested by the reviewer, these terms have been replaced by "fine resolution smoothing" and "coarse smoothing", respectively. (3) "the partial column bias of the latter often seems to be slightly shifted ($\sim$3 %) towards more positive values with respect to the former."

has been replaced by "the partial column bias of the fine resolution smoothing often seems to be slightly shifted ($\sim$3 %) towards more positive values with respect to the coarse smoothing."

16. (1) P11518, l.12-23: I guess the difference between VMR/partial columns and ND is expected. ND represents absolute ozone concentration, while VMR shows the ratio relative to air concentration, and partial columns are integrated, smoothed characteristics of ozone distribution that also depend on air pressure. (2) The authors partially agree in the sense that the differences between the relative comparisons shown in Fig. 8 are also to a certain extent due to the unit definitions as such that are behind. This effect is however hard to quantify with respect to the differences that occur as a result of the several unit conversions that are required, which are subject of study here. Therefore the possible differences due to unit definitions are not explicated. (3) No changes have been made.

17. (1) P. 11521, l.24: I would suggest to replace "...some users" by "some user requirements". (2) Not applicable anymore; see next response. (3) Not applicable anymore; see next response.

18. (1) P. 11521, l.21-24: The color code that indicates whether the dataset requirements met or not is mentioned in the Conclusions. However, I don't see any color coding in Table 4. Did you mean something else? Please, explain. (2) The authors had intended to you use a colour coding for user requirement compliance in Table 4, but such coding seemed to be not straightforwardly permitted by the AMT(D) manuscript submission format. The colour coding has therefore been replaced by a more extended description of the user requirement compliance in the following paragraphs of the conclusions section. The reference to a colour coding in Table 4 has however been forgotten to be correspondingly removed. This has now been taken care of. (3) The sentence "The compliance of the OPERA and RAL GOME-2 2008 datasets with these user requirements is highlighted with a colour code: green indicates ascertained compliance with requirements from all users, orange indicates compliance with some users

but not all, and red designates that no user requirements are met." has been removed.

19. (1) Figure 1: Red crosses for lidar stations are not seen on the figure. I would suggest changing the color and type of symbols for lidar stations (e.g. yellow squares) to make them easy to see. (2) The authors agree that in the AMTD slide format the red crosses that indicate lidar stations are hard to see. Fig. 1 however is intended to be printed at the width of a single column in the AMT manuscript format and has been optimised for viewing as such. The authors have thereby tried several symbol forms and colours, and finally selected the red crosses for the following reasons: (1) Other colours are even more difficult to distinguish due to the mostly blue-green-yellow background of the colour plot, white areas with missing data, and the black land contours, and (2) symbols for sonde and lidar stations have to be chosen such that they can be easily discerned when plotted on top of each other. Circles and crosses thereby seemed to be among the best options. The authors therefore prefer to keep the Fig. 1 plot symbols and colours unaltered. (3) No changes have been made.

20. (1) Figure 3: Black lines on the plots show centroids, however the term "centroid" has not been introduced in the text yet. It would be better to move a definition of "centroid" in the text before discussing results presented on Fig. 3. (2) At the location in the text where it is first introduced (beginning of section 5), Fig. 3 is only referred to as an exemplary view on averaging kernel matrices. At that point the plots are not discussed, although the last sentence of that paragraph and the figure caption explicitly refer to the centroid definition that appears in the manuscript later on: In section 5.3 the figure is again pointed to for discussing the centroid representations therein. The authors believe that this text order fits the structure of the manuscript appropriately well and provides sufficient clarity, and therefore opt to keep the corresponding text unchanged. (3) No changes have been made.

21. (1) Figure 8: Plots on Fig. 8 are too small; it is very difficult to see anything. Please, enlarge all plots. (2) Fig. 8 is intended to fill a full single page in the AMT publication format (paper size "letter") and has been optimised for viewing as such. The AMTD

slide format unfortunately is much smaller, impeding proper figure representation. (3) The Fig. 8 format as optimised for full page printing has been maintained.

22. (1) Figure A1: The figure is small. Please, increase the font size for the text. (2) Fig. A1 is intended to fill a full single page in the AMT publication format (paper size "letter") and has been optimised for viewing as such. The AMTD slide format unfortunately is much smaller, impeding proper figure representation. (3) The Fig. A1 format as optimised for full page printing has been maintained.

Anonymous Referee #2 (Received and published: 16 March 2015): Specific comments:

1. (1) Cost function. In Line 4 of page 11485 it is written that the optimal estimation approach searches for the minimum of a cost function of the form |y-F'(x)|. Actually the cost function of the optimal estimation includes the constraint of the a priori! (2) The authors agree. The next sentence was intended to cover this information, but apparently this was not very clear. This next sentence has therefore been modified. (3) The next sentence has been changed from "The retrieval scheme must thereby include additional constraints, e.g. in the form of prior information on the profile, its shape, and its allowed covariance." to "The OE retrieval scheme includes additional constraints in the form of an extra term in the cost function that accounts for prior information on the profile, its shape, and its allowed covariance."

2. (1) Filtering of profiles. From Sections 4.2 and 4.3 I understand that RAL and OPERA applies their own filtering to GOME-2 data, resulting in different but comparable number of considered profiles. Then in the forth line of the Conclusions it is written that 'exactly the same level-1 radiance measurements' are used by the two algorithms. This is a bit confusing and a clarification is needed. (2) The authors agree that the term "exactly" might make to formulation too strong and somewhat confusing. It is meant that both algorithms start from MetOp-A GOME-2 L1 V4 data for 2008. The conclusions have been modified accordingly. (3) The first sentence of the conclusions

has been modified to contain “run on MetOp-A GOME-2 level-1 version 4.0 radiance measurements of 2008” instead of “run on exactly the same level-1 radiance measurements”.

3. (1) MQQ in Sect 5. MQQ is inserted in the list of the quality indicators of the retrieval that can be derived from the AKM. Actually also the Covariance Matrix is needed for its computation. (2) The authors are aware of this requirement, and therefore explicitly wrote that one can derive and study several diagnostic parameters and quality indicators of the retrieval “starting from” the provided AKMs. In order not having to remove the MQQ from this list of quality indicators, and because of the detailed discussion of the MQQ definition afterwards, the authors prefer to keep the text unaltered. (3) No changes have been made.

4. (1) Computation of MQQ and DFS in Sect. 5.1. The two main quantifiers chosen for studying the information content of the two datasets are the grid-normalised relative measurement quality quantifier MQQ (indicated as QR) and the number of degrees of freedom of the signal (DFS). In the cited paper Ceccherini et al., 2012, the MQQ is computed as the trace of the Fisher Matrix, while in this paper it is computed as the trace of the 'fractional' Fisher matrix FR=SR-1AR computed with the fractional AR and SR. It is not immediate to see that these two approaches are equivalent. If you don't want to report the demonstration of that, it could at least be written that it can be demonstrated that the grid-normalised relative measurement quality quantifier QR computed using either F or FR and the DOF computed using either A or AR are the same. (2) It is not the authors' intention to demonstrate that the MQQ as calculated from the absolute FIM on the one hand and the fractional FIM on the other hand are equivalent, as they are not. The authors recognise this difference and consequently suggest that the fractional FIM (and AKM) should be used for determining information content measures, so that the latter are independent of the retrieval units. This statement might however not have been fully clear from the initial text, and has therefore been more explicitly included. (3) The sentence containing the absolute and relative FIM definitions has been

modified, stressing the impact of their difference on MQQ calculations: “The fractional FIM on the other hand is calculated as FR=SR-1AR for optimal estimation retrievals, yielding unit-independent MQQ values that differ from those that are calculated from the absolute F=S-1A (Ceccherini et al., 2012).”

5. (1) Signal information load in Sect. 5.1. It has to be defined or reference to it has to be provided. (2) The two references preceding the list of information content metrics (Rodgers, 2000; Ceccherini et al., 2012) contain all background information that has been taken into account for the description of the five metrics. However, in order to provide more clarity on the signal information load and its relation with the MQQ, one of these references has now been repeated in the text. (3) A reference to (Ceccherini et al., 2012) has been added at the end of section 5.1 second paragraph.

6. (1) Vertical density distribution of information would be useful. In Fig. 4 it is reported the dependence on latitude and time of QR (MQQ) for OPERA and RAL, as well as their number of degrees of freedom. As written in the text, these two quantities provide complementary information, the former providing an assessment of the information contained in the measurement only, independently on the applied constraints, the latter providing information on the number of independent components of the unknown profile, depending also on the constraints used in the retrieval such as the retrieval grid and the constraints used in the cost function. Only the vertically integrated quantifiers (third row of Table 3) are provided in Fig. 4, while details on the information vertical distribution contained in the measurements, given by their density distribution (described in the second row of Table 3), would be very useful to understand the information provided by the two algorithms at the different altitudes. An example of this kind of analysis can be found in the paper [S. Ceccherini, B.Carli, P. Raspollini, 'Quality of MIPAS operational products', Journal of Quantitative Spectroscopy and Radiative Transfer, 45-55, 2013]. It has to be noted that while the information on which altitudes mostly contribute to the DFS is indirectly provided by the dependence on altitude of the vertical resolution (in Fig. 6), details on the information distribution contained in the

measurements, independently of the applied constraints, would help in evaluating the complementary behaviour of RAL and OPERA algorithms as emerging from the values of QR reported in the lowest plots of Fig. 4. It is true that part of the information on this can be deduced from the median sensitivity profiles reported in Fig. 5. However, as it is written in the text, contrary to the MQQ, the vertical sensitivity depends on both the information contained in the measurements and on the applied a priori constraint. In general, if the a priori constraint is larger than needed, you can have a small sensitivity even if enough information is contained in the measurements. This is not the case for MQQ. (2) The authors agree that information content quantifier (density) distributions, like the layer DFS and the relative information distribution, add to an understanding of the atmospheric state measurement and retrieval processes and their vertical dependence. An explicit mentioning thereof and an exemplary reference to (Ceccherini et al., 2013) have therefore been added to the text. Given the focus and outline of the manuscript however, it has not been the authors' intention to include a thorough discussion of these vertical distributions to the information content analysis. Within a round-robin type evaluation as presented, the respective total amounts of information that are retrieved and the vertical sensitivity of the retrievals – indicating the levels or layers with a valid (i.e. not mainly prior) comparison with reference data – are of major importance. (3) The third paragraph of Section 5.1 has been extended with the following text: "Vertically resolved information (density) quantifiers like the layer DFS density and the relative information distribution (see Table 3) are not discussed here, although their analysis and complementarity can add to the understanding of atmospheric state measurement and retrieval processes and their vertical dependences, e.g. see Ceccherini et al. (2013). Within the presented round-robin type retrieval quality assessment however, the respective total amounts of information and the vertical sensitivity distribution (see Section 5.2) have – for reasons that are further elucidated – been selected to be of major focus." and the following has been added to the references list: "Ceccherini, S., Carli, B., and Raspollini, P.: Quality of MIPAS operational products, J. Quant. Spectrosc. Radiat. Transfer, 121, 45-55, doi:10.1016/j.jqsrt.2013.01.021,

2013."

7. (1) Vertical sensitivity in Sect. 5.2. The definition of the vertical sensitivity can be guessed from the text, but it is not clearly stated. I guess that it is the area of the fractional AKM row that is written to be a unit-normalised measure for how sensitive the retrieval at a certain height is to all heights. What is it meant with unit-normalised? It seems in contradiction with the obtained values of 2. (2) The authors agree that an explicit definition of sensitivity was missing, and that indeed the term "unit-normalised" is wrong. Sensitivity instead is a regularly "normalised" quantity. (3) The first sentence of section 5.2 has been modified to start as follows: "The areas of the fractional AKM rows (i.e. AKM row sums in practice) are a unit-independent normalised measure for. . ."

8. (1) Sect. 6.4 In order to make the reading more focused, many details reported in this section could be skipped or moved in the appendix. (2) The authors agree that the text of section 6.4 may have been quite loaded and less straightforward. With the aid of several specific comments by AR#1 and matching manuscript adaptations, the authors however believe that by now the text has been made more accessible. Taking into account this improved presentation, the authors consider it most informative for the reader to keep all relevant regridding and smoothing information together, and therefore argue that (re)moving text parts (to the Appendix) seems not to be adding value any longer. (3) Several clarifying manuscript adaptations have been made in response to specific comments by AR#1.

9. (1) Bias compared with random errors. In various parts of Sect. 7.1 (line 17 of page 11516, line 10 of page 11517, and line 23 of page 11519) bias of the measurements is compared with the single measurement satellite random uncertainty. Furthermore it is also written that 'to be fully statistical significant the bias should exceed the combined satellite-ground random uncertainty'. This sentence is not true: the significance of the bias is determined by the fact that it is different of 0 within the error of the bias. It is true that when a single measurement is considered, if the bias is smaller than the

random error, the retrieval error is mainly given by the random error. However, we have to consider that, while the bias remains constant when different measurements are averaged, the random error reduces with the square root of the number of averaged profiles. The bias should be compared with the error of the bias (given by the spread divided by the square root of the number of averaged profiles) to be understood if it is statistically significant and text and figure should be changed accordingly. (2) The authors agree that the statement "to be fully statistical significant the bias should exceed the combined satellite-ground random uncertainty" is not true and confirm that indeed the bias should be compared with the error of the bias to be understood if it is statistically significant. In the round-robin analysis context of this paper, the authors had however intended to make a different statement here, while the other text parts referred to (line 17 of page 11516, line 10 of page 11517, and line 23 of page 11519) are nevertheless correct: In the paragraph "on the difference statistics in relation to the satellite and ground-based random uncertainties" it is indeed the authors' goal to compare the bias and spread of the (relative) differences with the (relative) satellite and combined satellite-ground random uncertainties, respectively. Such comparisons do not directly allow for an evaluation of the statistical significance of the bias, but for an appraisal of its meaning with respect to the satellite measurement noise or the random uncertainty of the satellite and reference measurements combined. For that reason those measures have been included in Fig. 8 explicitly, and the authors therefore prefer to maintain the plots as initially presented. Yet the statement on lines 23-25 of page 11519 has been changed. (3) The initial "To be fully statistically significant however, the bias should exceed the combined satellite-ground random uncertainty, which is hardly the case somewhere here." has been replaced by "To be significantly detectable above the random noise level in a single measurement however, the bias should exceed the combined satellite-ground random uncertainty, which is hardly the case somewhere here."

10. (1) Fig A1 and Fig. 8. They are not legible, if not expanded up to at least 220 %. Fig. A1 could be simplified, reporting only the elements that are discussed in the

paper. Fig. 8 must be enlarged. (2) Fig. 8 and Fig. A1 are both intended to fill a full single page in the AMT publication format (paper size "letter") and have been optimised for viewing as such. The AMTD slide format unfortunately is much smaller, impeding proper figure representation. (3) The Fig. 8 and Fig. A1 formats as optimised for full page printing have been maintained.

11. (1) General comment. In general RAL and OPERA measurements present many complementarities. It would be interesting to understand how this complementary could be exploited to improve the quality of Ozone profiles. (2) The authors strongly agree. Both retrieval teams have put substantial effort in the development of a joint retrieval scheme and the authors have therefore added a statement on the development of this joint scheme as a part of the future outlook at the end of the conclusions section. (3) The following text part has been added to the end of the conclusions section: "Moreover, the current RAL scheme does not use measurements between 307 and 322 nm (largely due to instrumental issues encountered during development), while this omitted range is included in OPERA and includes potentially valuable information on the ozone profile. Work is therefore ongoing within the Ozone CCI to assess whether the approaches used by the two schemes could be combined to allow this potential to be fulfilled in practice."

Anonymous Referee #2 (Received and published: 16 March 2015): Technical correction:

1. (1) Line 2 of page 11525: 'as shown in in the second column'. Please remove one 'in'. (2) The authors agree. (3) One "in" has been removed.

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 11481, 2014.