

The response to the Referees shall be structured in a clear and easy-to-follow sequence: (1) comments from Referees, (2) author's response, (3) author's changes in manuscript.

Anonymous Referee #2

Received and published: 16 April 2018

General Comments:

This paper presents a comprehensive and lengthy assessment of the Ozone_cci CRDP of 13 nadir ozone profile data products from both UV-VIS and TIR instruments as well as 1 data assimilation product. The evaluation includes data content studies, information content, and validation against ground-based ozonesonde and lidar observations in terms of median relative biases and the IP68 spread as a function of various influence quantities and relative decadal drift. It is a very useful study and its scope is very suitable for publication in AMT. This paper is generally well organized, the methodology is generally very good and valid, and the results are well described. However, some of the sections are difficult to understand. For example, the section of L3 gridding could be made clearer and simpler and Figure 1 could be removed. The results of the vertical sensitivity are very difficult to interpret, and the derivation of vertical sensitivity could be improved. Also, the abstract does not include main conclusions. In addition, some texts need clarifications. Overall, I think that this paper can be published after addressing the comments mentioned here and specific comments below.

Specific Comments:

- (1) 1. In abstract, no conclusions are given. So what are the main conclusions of this study? Some of the sentences in the conclusions/discussion sections can be paraphrased here.
- (2) The authors agree that some of the major conclusions should be provided in the abstract as well, and have extended the abstract accordingly.
- (3) After “(WMO GAW).” the abstract has been extended as follows: “The nadir ozone profile CRDP quality assessment reveals that all nadir ozone profile products under study fulfil the GCOS user requirements in terms of observation frequency and horizontal and vertical resolution. Yet all L2 observations also show sensitivity outliers in the UTLS and are strongly correlated vertically due to substantial averaging kernel fluctuations that extend far beyond the kernel’s 15 km FWHM. The CRDP typically does not comply with the GCOS user requirements in terms of total uncertainty and decadal drift, except for the UV-VIS L4 dataset. The drift values of the L2 GOME and OMI, the L3 IASI, and the L4 assimilated products are found to be overall insignificant however, and applying appropriate altitude-dependent bias and drift corrections make the data fit for climate and atmospheric composition monitoring and modelling purposes. Dependence of the Ozone_cci data quality on major influence quantities – resulting in data screening suggestions to users – and perspectives for the Copernicus Sentinel missions are additionally discussed.”
2. In the introduction, full instrument names should be specified at their first occurrences.
- (1) 3. In Sections 2.2 and 2.3, it is useful to mention the unit of the retrieved ozone profile for each algorithm: partial ozone column in DU, average ozone mixing ratio in ppbv, etc.
- (2) In reply to a similar comment by reviewer 1, the authors have added two columns to Table 1 and extended its caption.
- (3) Two columns have been added to Table 1, and its caption has been extended as follows: “The products’ vertical range (with number of levels or layers between brackets) and original units are added in the last two columns.”
- (1) 4. Figure 1 and the text on P7 are difficult to follow and confusing. I guess that grid cells refer to those 1 x 1 boxes, but Figure 1 caption says TM5 assimilation grid. Is TM5 grid 1x1 (looks like it is 2 x 3 based on sect. 2.5)? Also grid cells boundaries typically are not parallel or perpendicular to ground pixel edges as shown in the figure. The naming of grid cells based on pixel corners also makes it more confusing as depending on the pixel size, the entire pixel can lie in one grid cell. Also, what is the size

of subpixels and how many subpixels for different instruments? I think that this can be described more clearly and also more concisely. The figure does not really help here and can be removed. Basically, each ground pixel is divided into subpixels (size, #), each subpixel contains the same value and uncertainty, then assign the subpixels to grid cells.

(2) In the context of the KNMI L3 product, a pixel refers to a satellite measurement, while a lat-lon grid cell refers to the regular 1x1 degree latitude-longitude grid for which the mean and standard deviation are calculated. Each pixel is divided into 25 subpixels, which are assigned to the grid cell containing the subpixel. The mean and standard deviation for the grid cell are calculated according to the equations given in the text. The authors agree with both reviewers that the text on which subpixels are assigned to which grid cell is unclear and the text of section 2.4 and the caption of figure 1, which is preferably maintained, have been updated accordingly.

(3) The paragraph before Eq. (1) has been replaced by the following: “Monthly-averaged L3 profile products are produced from the filtered RAL v2.14 GOME, GOME-2A, SCIAMACHY, and OMI data by the Royal Meteorological Institute of the Netherlands (KNMI). Version 0004 of the KNMI L3 products has been used in this work (see Table 1). The KNMI level-3 data consist of monthly ozone profile averages, also on a one-by-one degree latitude-longitude grid, containing 19 layers between 20 fixed pressure levels at each grid-point. The algorithm that calculates the monthly-averaged ozone fields assumes that the L2 satellite ground pixel vertices (labelled ABCD) are ordered as indicated in Figure 1. Each pixel’s across-track direction is defined by the lines AD and BC, while the along-track direction is defined by the lines AB and DC. The satellite pixel is divided into 25 subpixels, five in the along-track direction and five in the cross-track direction, and each subpixel is assigned to the L3 grid cell (the boundaries are indicated with the dashed lines in Figure 1) containing the subpixel. The subpixel values x_i are weighted by the square inverse of their uncertainties (σ_i^{-2}), so the weighted mean grid cell value x_c and the corresponding standard deviation σ_c are given by”
The caption of Figure 1 now reads as follows: “Figure 1: A L2 satellite pixel ABCD is divided into subpixels (diamonds 1 to 7). Each subpixel is assigned to a L3 grid cell (indicated with the dashed boundaries) and the average and standard deviation are calculated (see text). In this example, subpixels 1-3 would be assigned to the lower-right grid cell and subpixels 4-7 would be assigned to the lower-left grid cell. The satellite pixel ABCD may have any orientation with respect to the L3 grid.”

(1) 5. Please put table captions before the tables.

(2) This is done automatically in the final publication mark-up.

(3) Action is taken upon final submission of the manuscript.

(1) 6. P10, L17-21, temperature profiles are not required for conversion between number density and average layer VMR. Assuming a layer is well mixed, then average VMR = $1.25 * (\text{partial ozone column in DU}) / (\text{pressure difference of the layer in atm})$. Please see appendix B of Ziemke et al. (“Cloud slicing”: A new technique to derive upper tropospheric ozone from satellite measurements, JGR, 106, (D9), P 9853–9867, 2001) for more detail. The partial ozone column is related to number density and altitude difference of the layer.

(2) In these lines the level-related VMR value is intended, not the layer-average VMR. In that case the temperature profile is required. The authors are familiar with the literature referred to, but thank the reviewer for pointing this out again.

(3) No changes have been made.

(1) 7. P11, it is not clear about the three numbers in SPI column separated by “:”

(2) The authors admit that the meaning of the SPI values had erroneously not been mentioned in the text. Therefore the text has been extended with reference to Table 4, also in agreement with a comment by reviewer 1. The notation in Table 4 has moreover been changed.

(3) After the first reference to Table 4 in Section 3.3, the following sentence has been added: “The possible satellite pixel index (SPI) values within each cross-track scan and the resulting number of pixels per scan are provided for each instrument in Table 4 (taking into account pixel co-adding, see Section 2).” The notation of the possible SPI in Table 4 has been changed from X:X:X (start, step, end) to X,X,X,...,X (start, start+1, start+2,...,end).

(1) 8. P12, L18, even if it is difficult to know how much IASI data are screened as a function of latitude and time, the data providers should know on average how much data are screened out due to the use of cloud fraction greater than 13%.

(2) This section has been somewhat modified according to suggestions by the first reviewer. The authors agree that IASI's pre-screening does not allow a full assessment of the data (screening) distribution. Yet cloud screening has globally a similar effect on both the UV-VIS and TIR data. This information has been added to the text.

(3) The sentence on the IASI screening has been extended as follows: "The IASI screening on the other hand appears very low, yet this is due to the pre-screening by the product providers before data delivery, i.e., mainly the IASI cloud screening (if the fraction is higher than 13 %) cannot be observed from the plots, but is roughly of the same order as the UV-VIS data screening."

(1) 9. P15, L15-16, It is not clear why "quite stable in time" reflects the signal degradation correction? Please clarify it.

(2) The authors considered it quite obvious that a satellite signal degradation correction results in maintaining the signal-to-noise ratio close to its initial level and hence in a quite stable DFS upon nadir ozone profile retrieval. The DFS reduces in correlation with the signal if no signal degradation correction is applied. Yet for clarity, the authors have added this explanation in the text.

(3) "This correction maintains the instrument's signal-to-noise ratio close to its initial level and hence reduces the effect of the instrument degradation on the retrieval's DFS." has been added.

(1) 10. P15, L18-19, It is useful to explain the lower DFS under SAA: shorter wavelengths with weak signals cannot be used due to SAA, thus significantly reducing DFS in the stratosphere.

(2) The authors agree that some clarification would be helpful.

(3) "due to stratospheric intrusion of high-energetic particles (the tropospheric DFS is mostly maintained)" has been added.

(1) 11. P14, equation (6), based on the text, A_F is provided from the FORLI algorithm, so should the defractionalisation operation derive $A(m, n)$ from A_F rather than derive A_F from $A(m, n)$? I suggest changing this equation to $A(m, n) = A_F * \dots$

(2) The reviewer might have misunderstood: the absolute averaging kernel A is given in the FORLI data, yet this AK is calculated by the data providers from the fractional AK that results from the retrieval. This calculation is done using the prior profile, hence in this work we also use the prior profile to invert this operation and again obtain the fractional kernel matrix. The authors believe that this is clearly stated in the text already: "These fractional matrices are made unit-dependent by use of the prior profile before saving into the data files, allowing for more straightforward application (e.g. for vertical smoothing operations) by data users. For the information content studies presented here, this defractionalisation operation therefore has to be inverted"

(3) No further changes have been made.

(1) 12. P14, L22, and first paragraph of P17, Figure 5: It is not easy to understand the meaning of vertical sensitivity. Based on the definition on P14, it is an indication of the fraction of the information that is from the data. But on Figure 5 and P17, the vertical sensitivity values peak in the UTLS with a median value of 3, and are often greater than 1 even below 6 km, which does not seem to be consistent with the definition of the fraction of information from the measurement. Also the vertical sensitivity should not peak in the UTLS, as there is stronger vertical sensitivity in the stratosphere from UV-VIS measurements. Please check DFS at individual layers to make sure this is the case. It seems to me that this concept is not actually a good indicator of the vertical sensitivity or it might depend on how the vertical sensitivity is derived (e.g., from AKM or fractional AKM , what is the unit of state vector, e.g., DU or mixing ratio etc.). Based on the definition, when you sum the sensitivity of retrieved ozone at a layer to the perturbations of ozone at all layers, the units of state vector or the weighting of the perturbations at each layer are important. Using mass conserved units like DU or the weighting of perturbations at each layer by a priori error (rather than the a priori or the retrieved profile) might make more sense. Between IASI and UV-VIS retrievals, it is good to convert

the AK to the same units and then apply the same concept. Please clarify this on P4. You may also consider the use of DFS at each layer (diagonal elements of AK) normalized to the depth of layer (to account for non-uniform, variable vertical altitude grid) to show the vertical sensitivity, which is straightforward and independent of the retrieval scheme and might be more meaningful.

(2) The authors acknowledge that a fraction-like measure below zero or above one makes little sense. This sensitivity interpretation by Rodgers however is only a rough approximation (as stated in the text). In practice, the sensitivity at a specific retrieval level can nevertheless be negative or exceed unity (over-sensitivity) due to kernel fluctuations and correlations between adjacent retrieval levels, as reflected in the kernel width. This has now been made clear in the text. The over-sensitivity in the UTLS hence is no surprise, as especially the UV-VIS nadir ozone profile retrieval shows difficulties around the tropopause and below. The normalized layer-DFS (that is known to the authors) certainly is a useful quantity that has been considered in this work, but has been found to add little to the combined DFS and vertical sensitivity discussion. It has therefore not been additionally discussed. Both (overall) DFS and vertical sensitivity calculated from fractional averaging kernel matrices are fully unit-independent measures (the authors refer to Keppens et al., 2015, for a more extensive discussion), as requested by the reviewer.

(3) In section 4.2.1, right after the vertical sensitivity interpretation by Rodgers (2000), it has been added that “Note however that the sensitivity at a specific retrieval level can nevertheless be negative or exceed unity (over-sensitivity) due to kernel fluctuations and correlations between adjacent retrieval levels, as reflected in the kernel width (see below).”

(1) 13. P17, L12-14, it is not clear why the strong sensitivity variability affects vertical smoothing and Eq. 3 introduces a bias. Please make it clearer.

(2) The authors agree that this might not be clear immediately. It has now been pointed out that the sensitivity variability points at outliers in the averaging kernel matrices, thus introducing biases upon averaging kernel smoothing.

(3) It has been added that the sensitivity variability is “pointing at outliers in the averaging kernel matrices”

(1) 14. P17, L17-18 and also in Fig. 5 caption, it is not clear which is direct and centroid offset between 2nd and 3rd rows. Please make it clear in the figure caption. Also please mention the dotted lines in the figure caption.

(2) The authors agree that offset and spread indications can be improved. The caption of Figure 5 has been updated accordingly.

(3) The caption of Figure 5 has been updated as follows: “Global GOME-2A (left) and IASI-A (right) information content in terms of vertical sensitivity, retrieval offset (in km), and averaging kernel width (in km) and their dependence on DFS, SZA, or thermal contrast (TC). Black dashed lines represent median values, while out-of-range profiles are plotted in magenta. Different measures are used for the offset and kernel width in the second and third rows, which include the centroid offset and Backus-Gilbert spread, and the direct offset and FWHM, respectively. Plot titles provide the absolute and relative amounts of profiles after screening, and the number of ground-based overpass stations.”

(1) 15. P17, L31, in “decreases first to about 20 km”, it seems to me from the figure that the maximum median FWHM is 20 km, so should it be a smaller number here?

(2) At this point the BG spread was intended. The authors agree that this was not fully clear and have made this explicit.

(3) The beginning of the sentence has been changed into “At higher altitudes, the median BG kernel width”

(1) 16. P18, first sentence, “slant column density” should not be parallel to “the sensitivity” because the larger slant column density, the smaller the sensitivity from surface to the lower stratosphere. The real reason is because, the larger slant path length or slant column density, the fewer photons penetrating into the troposphere, the smaller the sensitivity in the troposphere, and the larger the resolving length values.

(2) The authors agree that the phrasing “and thus of the sensitivity” might be confusing to the non-expert reader. The reference to the vertical sensitivity has therefore been omitted.
(3) “and thus of the sensitivity” has been removed.

(1) 17. P20, L4, suggest changing “quarter” to “season”

(2) The authors agree that the ‘seasonal’ dependence is studied, yet for clarity and simplicity, full months are considered for grouping. This specification in terms of quarters is maintained throughout the text. In order to explicitly refer to the study of the “seasonal dependence” of the difference statistics however, this formulation has now been added to the text.

(3) The text has been updated as follows: “The influence quantities considered in this work are latitude (for meridian dependence), quarter (for seasonal dependence)...”

(1) 18. P20, L24-25, The sentence “This is with the exception of the Metop-B GOME-2 and IASI instruments however, that have not been used for drift studies” is difficult to understand. Suggest changing to “So Metop-B GOME-2 and IASI instruments are excluded for drift studies” and move it after “for this drift assessment”

(2) The authors agree that the reviewer’s phrasing is more clear, and have adopted it accordingly.

(3) The second part of the paragraph has been changed into “To avoid spurious effects due to a seasonal cycle in the differences, only time series of five years or longer are used for this drift assessment. Therefore Metop-B GOME-2 and IASI instruments are excluded from the drift studies (indicated with an asterisk in Table 4). Moreover, only fully available years of the satellite datasets have been considered for comparative analysis in order not to introduce seasonal effects at the beginning and the end of each time series.”

(1) 19. P20, L29, in addition to latitude and season, the influence quantity of time should be added.

(2) The authors agree; time has been added.

(3) The sentence has been changed to “These have therefore been limited to the latitude, and quarter, and time (drift).”

(1) 20. P21, the sentence above Eq. 9 is difficult to read. Suggest changing to “Yet both approaches introduce similar spatial and temporal representativeness errors into the difference statistics because taking (monthly) averages as a bias estimator XXX yields comparable outcomes:”

(2) The authors agree with the reviewer that the phrasing of this sentence can be simplified and have therefore adopted the reviewer’s suggestion.

(3) The sentence before Eq. (9) has been changed into “Yet both approaches introduce similar spatial and temporal representativeness errors into the difference statistics because taking (monthly) averages as a bias estimator X yields comparable outcomes:”

(1) 21. P22, L15, is the ex-ante uncertainty from the retrievals for random noise errors or for both random noise errors and smoothing errors? As the averaging kernels are applied to reference data to remove smoothing errors, ex-ante uncertainty of random noise errors should be shown here. Please clarify this.

(2) The authors acknowledge this point with respect to error budget assessments. The ex-ante errors referred to here only include random errors. This has now been made clear in the text, with additional reference to Miles et al. (2015).

(3) The beginning of the paragraph has been changed into “The individual L2 UV-VIS comparison graphs also contain information on the validity of ex-ante uncertainties provided for the satellite nadir ozone profile retrievals (thin grey lines). The relative random error reported in the RAL v2.14 data files amounts to about 5 % at the altitude of the ozone maximum, up to about 10 % at higher altitudes, and up to 40 % in the lower troposphere. In theory the IP68 spread should be close to the combined uncertainty of the satellite data, the ground-based data, and metrology errors due to remaining differences in vertical and horizontal smoothing of atmospheric variability (including co-location mismatch errors). The latter is difficult to assess, but one can expect that the bias and spread estimates resulting from the comparisons including AK smoothing are close to the combined uncertainty of

satellite and ground-based data, or at least the ex-ante satellite uncertainty in practice (Miles et al., 2015).”

- (1) 22. P22, L26, suggest adding “because the retrievals only include random noise errors and smoothing errors in the ex-ante uncertainty” after the “nadir ozone profile products”
- (2) The authors have only partially adopted this suggestion in order to clarify the statement, because the ex-ante errors referred to only contain random uncertainties.
- (3) “because the ex-ante uncertainty under consideration only includes random noise errors” has been added to the text.

- (1) 23. P22, L26, OMI is not an exception in that the total satellite measurement uncertainty is underestimated because the ex-ante uncertainty should be compared to comparison spread or the quadratic sum of comparison bias and spread rather than the comparison bias only.
- (2) The authors agree that this statement confuses the uncertainty contributions. It has therefore been changed.
- (3) “The only exception is given by the OMI tropospheric ozone data that have a bias below their ex-ante uncertainty.” has been replaced by “Only for the OMI tropospheric ozone data with a bias within 10 % does the combined uncertainty come close to the ex-ante uncertainty.”

- (1) 24. P23, L1, it is not generally true that there are smaller biases for larger total ozone columns based on the figures as the biases often increases when the total ozone increases from 300-400 to 400-500 or 500-600 DU (very clearly for GOME and OMI retrievals).
- (2) The authors agree that the dependence is not linear, and have modified the text to avoid giving this impression.
- (3) “Smaller biases are typically obtained in the northern hemisphere and for larger total ozone columns. The latter is indeed expected to result in an improved satellite measurement and retrieval sensitivity, and thus more stable averaging kernel behaviour with smaller vertical dependences.” has been changed into “Smaller biases are typically obtained in the northern hemisphere and for intermediate to larger total ozone columns. Larger ozone columns are indeed expected to result in an improved satellite measurement and retrieval sensitivity, and thus more stable averaging kernel behaviour with smaller vertical dependences.”

- (1) 25. P23, L3-4, the relationship between SZA/DFS and the biases are altitude-dependent. From the figures, the biases are typically smaller at larger SZAs/DFS in the troposphere, but are larger at larger SZAs/DFS. Larger SZAs typically lead to larger total DFS due to the increase of DFS in the stratosphere and often lead to smaller DFS in the troposphere due to reduced photon penetration. Smaller biases in the troposphere at larger SZAs/DFS could be due to the reduced retrieval sensitivity in the troposphere (i.e., retrievals are closer to the a priori). So the causal relationship is not as straightforward as larger DFS means better retrieval sensitivity and therefore smaller biases.
- (2) The authors thank the reviewer for clearly pointing this out. In order to introduce this vertical dependence of the SZA/DFS relationship, the text has been extended.
- (3) The original statement has been replaced by the following: “On the other hand, the DFS and SZA behaviour is somewhat smaller and, as one can again expect for UV-VIS observations, rather similar, with the higher solar zenith angles typically corresponding to the larger DFS values (mainly from the stratosphere), the largest stratospheric biases, and the smallest tropospheric biases. The latter could be due however to a somewhat reduced tropospheric sensitivity, bringing the retrieved profile closer to the prior profile.”

- (1) 26. P33, L11, based on figures, the spread is not always enlarged in the L3 comparison. Instead, the spread is typically significantly reduced below 6 km. This should be mentioned and explained.
- (2) The authors agree that this reduction of the L3 spread for the lowest level was not discussed in the original text. A statement has been added that points at this effect and provides a brief explanation.
- (3) The following statement has been added: “Note however that the lack of kernel smoothing instead reduces the L3 spread for the lowest level, which has a strongly reduced sensitivity in comparison with the levels above.”

(1) 27. P34, L1, says “GOME L3 data show a negative above-tropopause bias of 5-10%”. But based on the first panel of Fig. 11, I see mostly positive biases above 100 hPa, especially with large positive biases of 20% around 70 hPa and positive biases of 40% around 8 hPa. Please clarify this.

(2) The authors acknowledge this mistake and thank the reviewer for his/her thorough reading. This statement has been changed and extended for clarity.

(3) The first sentence of the paragraph now reads as follows: “GOME level-3 data show an above-tropopause bias of 5-10 % positive to negative, with strong outliers around 70 hPa and 8 hPa, especially in the tropical UTLS and Antarctic local spring (up to 50 %) due to ozone hole’s vortex conditions.”

(1) 28. P36, L11, again the spread values for the L3 comparison can be smaller below 6 km, which should be mentioned.

(2) The authors agree that this deviation should be mentioned for completeness. This has been done by reference to the previous section.

(3) “(except for the lowest-level spread, see previous section)” has been added.

(1) 29. P40, L7-8, it is useful to explain to the readers why there is stronger tropospheric reduction to 20% and why the drift is small (e.g., due to bias correction).

(2) The authors agree that some additional explanation is helpful to the reader. The text has been extended with the requested information.

(3) The last sentences of Section 5.5 have been changed into “The L4 spread remains close to the L2 and L3 values, though with an even stronger reduction (to 20 %) in the troposphere than the L3 comparisons as no monthly averages are considered. Moreover, due to the ozonesonde-based bias correction the remaining L4 drift is of the order of a few percent only and insignificant, i.e. within the 95 % CI, for all altitudes up to about 40 km globally.”

(1) 30. P45, last line, the 30% negative should be 30% positive in Antarctica as shown in last panel of Fig. 16. Also please change Table 5 correspondingly.

(2) The authors thank the reviewer for pointing out this error and have corrected the text and table accordingly.

(3) The text has been changed into “Despite this strong seasonality, and in agreement with the IASI-A L2 comparison statistics, median relative differences throughout the whole time series range between 25 % negative in the northern mid-latitudes and 30 % positive in Antarctica, with a nearly zero overall bias around the equator.” In Table 5, row 9 column 6, “-30” has been changed into “30”.

(1) 31. In table 5, suggest changing “Vertical resolution (6 km to troposphere)” to “Vertical grid/resolution”, changing to “115 km²”, “230 by 345 km²”, “12 km²”, “115 km²”

(2) The “Vertical resolution (6 km to troposphere)” QA quantity and user-requirement are given by GCOS and therefore preferably left unchanged. The horizontal resolution user requirements on the other hand are provided for a single dimension by GCOS (see first column) and therefore expressed in km.

(3) No changes have been made.

(1) 32. In Figs. 6-10, 14-15, change second bracket from “[“ to “[”]” in Fig. captions

(2) The open bracket “[“ is used commonly to indicate that the last value is not included, i.e. it indicates that values included in the set go up the last value (<), but don’t equal the last value (<=). The authors wish to preserve this indication as such, as it better represents the content of the sets.

(3) No action has been taken.

(1) Technical Comments:

1. P2 last line, and P8 L25, change “Keppens et al., 2015” to “Keppens et al. (2015)”

2. P4, L6, change to “time series”

3. P4, L13, P5, L15, P6, L13, change “priori” to “a priori”

4. P9, L14, change “beyond” to “above”

5. P11, change “prior” or “a-priori” to “a priori”
6. P11, suggest changing “averaging kernel smoothing of the FRM” to “smoothing the FRM with averaging kernel”
7. P12, L6, suggest changing “relative amount” to “percentage”
8. P12, L18, suggest changing “delivery. E.g.” to “delivery, i.e., “
9. P13, L10, change “prior” or “a-priori” to “a priori”
10. P14, L5-6, change to “help understand”
11. P14, L16 & L26, change “prior” or “a-priori” to “a priori”
12. P21, L13, suggest changing “Thanks to” to “Due to” to make it formal.
13. P21, L21, change “vertical averaging smoothing of ground-based reference data” to “vertical smoothing of ground-based reference data with averaging kernels”
14. P21, L29-30, change “smoothing difference error” to “retrieval smoothing error” and “Tropics” to “tropics”
15. P23, L6, change to “instruments except for GOME-2B”
16. P33, L12, change to “lack of”
17. P48, L10, change to “7 km by 7 km” or “7 by 7 km²”
18. P49, L11, change to “equivalent to”
19. P49, L12, suggest changing “as an alternative” to “along with”

(2) The authors thank the reviewer for the extensive technical check and suggestions. All suggestions have been incorporated, except for those suggesting to change “prior” into “a priori” (3, 5, 9, 11). The use of the synonym “prior” sometimes improves readability and has therefore been kept. Suggestion 14 has only been partially followed: “Tropics” has been changed to “tropics” at all instances, but “smoothing difference error” has not been changed. This phrasing indicates that the error is due to the difference in smoothing between the satellite and reference profiles, and not to the satellite profile smoothing only.

(3) All suggestions have been incorporated, except for those suggesting to change “prior” into “a priori” (3, 5, 9, 11), and to rephrase “smoothing difference error”.