11 July 2013

Dear Editor and reviewers,

Thank you very much for editing and reviewing for our paper. We corrected the manuscript according reviewers comments as follows. C is a comment and A is an answer.

Reviewer 1:

Minor comments:

C1. I am not sure if the title "Validation ..." is well chosen since the authors don't remove the systematic error . They say that they can reduce the uncertainty in the next retrieval version. I think "Intercomparison ..." would fit better. In addition nobody knows which satellite is right. It is a question if Aura, MIPAS, SMR, ACE are really independent from each other. So even if three of them agree and SMILES has a bias with respect to them, it can be due to the fact that the retrieval developers already "harmonized" their data sets.

A1. I completely agree that no satellite data shown true absolute value, and comparison with satellite measurements should be entitled "inter-comparison…". In this paper, we tried to evaluate the quality of SMILES ozone observation. That is why we performed not only the inter-comparison with satellite data, but also the full evaluation for L1b spectrum, L2 profiles, and comparison with ozone sonde data, which believed to give a true value in the stratosphere within a few percent. I think the title "Inter-comparison…" includes only a part of the work we did. Our paper might fit for title with "Evaluation of … ", but then people may not understand what we did from title. Actually, I could not find the best title for this study. I would like to keep the title with "Validation…." while this is not the best title, but better title than others.

C2. In the present paper, it is said in the conclusions that SMILES has a negative bias above 6hPa and I don't understand why a tangent height error should be induced by a gain calibration error? (page 2678). I like to repeat that the other satellite experiments don't know if their values are correct by 10% above 6hPa. The present paper of Kasai et al., provides the misleading impression that some experiments measure the true ozone values.

A2. Thank you very much for such right comments. I improved and add more detail explanations. Originals: 1) line 13-15 in page 2678 (shown in red) for the relation between tangent height error, a gain calibration error, and SMILES negative bias, and 2) line 25-28 in page 2678 for comparison results.

Improvements: line 791-799 in page 24. “This negative bias becomes larger with altitude, and can be explained by the error from retrieved tangent height. We retrieve the tangent heights from ozone spectrum, and the error is originally coming from the uncertainty of the gain calibration of L1b spectrum. c-yk: The next version of L1b data (version 008) will include an improvement in the gain calibration. The NICT SMILES data processing team will use that new calibrated measurements, L1b 008, for the new L2 data processing, L2 v300. We confirm that the negative bias of O3 at upper altitudes profile had been solved from preliminary analysis.”

C3. Abstract: The abstract should provide the important information that the data users should consider band B rather than band A.

A3. I add “SMILES O3 from observation frequency band B had better accuracy than that from band A.” in line 20 page 2.

C4. Abstract (and Introduction) p. 2645, line 10 ... components: error analysis; ... p.2645, line 21 maybe: " ...and the gain calibration." (calibration problem sounds not well) p.2647, line 5, Sect. 3 consists of ... p.2647, line 15: ... of O3 from SMILES is shown in Sect.5. p.2648 line 9, Did you explain band C?

A4. Thank you for the improvement. I corrected according the suggestion. You could find them with red words in page 2-4.

C5. Error analysis: The main result might be that the error analysis provides smaller error bars than the external comparison. Is it due to a too "optimistic" error estimation? It might be of general interest to comment on the advantages and drawbacks of error analysis.

A5. This is because we did not involved the error from retrieved tangent height in the O3 error analysis as shown in line 160 in page 6, which rise to the major error source of the systematic error in the upper atmosphere. The tangent height error is originally coming from the error of the spectrum calibration error (gain non-linearity behavior).

We had estimated the errors only the direct effect on O3 spectrum and profiles, not for the second order effects. For example, we also did not included the error from the uncertainty of temperature profiles, which also we had retrieved from ozone spectrum.

I add these descriptions in the error analysis and summary as “It should be noted that we estimated the errors only the direct effects on O3 spectrum and profiles, and did not estimated the second-order effects, such as an error of tangent height.” in line 177 in page 6, and as “This negative bias becomes

larger with altitude, and can be explained by the error from retrieved tangent height. We retrieve the tangent heights from ozone spectrum, and the error is originally coming from the uncertainty of the gain calibration of L1b spectrum" line 795 in page 24, respectively.

C6. p.2661 line 12 unclear: "This" refers to smoothing of ozonesondes (some sentences before)?
A6. "This" means the smoothing function for the comparison with ozonesondes, Odin/SMR, and MIPAS. I had modified to "Direct comparison are applied for MLS, SMR, and ACE-FTS since the vertical resolutions and sampling intervals are comparable with that of SMILES. We applied a vertically-smoothing triangle function as shown in Eq. (1), using the width of SMILES averaging kernel, for the ozonesonde, MIPAS, and Odin/OSIRIS datasets. The" in line 385-389 in page 12 .

C7. p.2663 ECC: I could imagine that "precision" is a problem for intercomparisons with ozonesondes in regions of high ozone variability since the ozonesondes perform point measurements.
A6. Yes, the observed air-mass is not exactly same, since the field-of-view is different.

C8. p. 2673, line 7, ther ? line 12: what do you mean with "lower sensitivity"? e.g., is it due to an unstable retrieval of SMR data? Or to a static retrieval of SMILES?
A8. We thought the lower sensitivity of SMR data. I have corrected it to "Odin/SMR instrument".

C9. p.2676 line 5, Is there somewhere a proof that Aura or Mipas measure the "true values"? Please give a reference to this proof.
A9. No, there is no proof. For example, we wrote "it was known that the MIPAS O3 profiles have positive biases (+0.9 ppmv at a maximum) around 37 km by (Stiller et al., 2012)," in line 730 page 22.

C10. abstract and conclusion: the diurnal variations of ozone are only treated on a half page.
So I would not mention the "diurnal variations" in the start sentence. I wonder that you did not consider ground-based remote sensing in the cross-validation since it is easy to find coincident profile pairs in case of a ground station. Some ground stations can provide the diurnal ozone variation too. Maybe you consider it in a later study.
A10. Thank you for the suggestion. What I wanted to say was we did the comparison under the condition of different local time for each satellite measurements. This was the first paper for ozone inter-comparison with different local time. I had changed the abstract and conclusion to make it clear.

C11. Figure captions Often it is not described if you took the difference "SMILES - X" or "X - SMILES" e.g. Figure 8 Comparison .... I would suggest something like this : Difference profiles of ozone: SMILES(NICT) - SMILES(JAXA) ...

A11: Thank you. I had changed it for all figures.

Reviewer 2:

C1. General comments: This article provides a comprehensive account of the validation of the SMILES ozone product from National Institute of Information and Communications Technology (NICT). The article also presents comparison of the NICT ozone with the operational ozone product from the Japan Aerospace Exploration Agency (JAXA). The article is well written. Regarding the structure of the article the section 5 (diur- nal variation) is quite isolated from the rest of the article. Perhaps this part could be extended and published separately? The main results of the paper show that ozone (NICT) deviates quite a much from other validating measurements above the ozone layer. The authors claim that the reason for this deviation is now well understood and the processing will be improved in the next version of data release. I wonder if it would be reasonable for authors to delay the publication of this article and use the new data version for these validations? Anyway, even now the article is worth of publishing in AMT.

A1. Thank you very much. We would like to keep the diurnal variation part, because we want to notice the inhomogeneous sampling issue in the paper for the diurnal variation for the data user.

My minor comments are listed below.

Detailed comments:

C2. Abstract and p 2646, lines 1: The abstract and article start by mentioning diurnal variation.
As mentioned above it does not reflect the main focus of the article.

A2. We have improved it. What I wanted to say was we did the comparison under the condition of different local time for each satellite measurements. This was the first paper for ozone inter-comparison with different local time. I had changed the abstract and conclusion to make it clear.

C3s. p. 2650, line 12: Is $\chi 2$ for each altitude or is it for the whole scan?

A3. We use $\chi 2$ for each altitude. I have improved all of these and shown in red in line 380 of page 12.

C4. p. 2650, line 18: The limit $\chi 2 < 0.8$ looks quite strict. Could you provide information about the distribution of $\chi 2$ -values. Do you have any outliers in data not detected by this limit?

A4. The most of the data (more than 85%) are smaller than $\chi 2 < 0.8$ in SMILES observation for its sensitivity. That is why we took $\chi 2 < 0.8$.

C5s. p. 2650, line 13: Define the measurement response m

p. 2653, line 25: inwhich -> in which

p. 2655, lines 18-20: Define parameters and n

A5s. I have improved all of these and shown in red.

C6. p. 2660, line 23-24: Provide an estimate how much of data is rejected by these two limits

A6. About 10 % of data were rejected. I have improved all of these and shown in red.

C7. p. 2660, line 23-24: Are the limits applied to all altitudes? Is the whole profile rejected if a limit is exceeded at one altitude?

A7. The limits applied to all altitude. We add the description in red in line 380 of page 12.

C8. p. 2662, Eq. (3): Are you sure that this kind of average is a good measure for the relative difference. As you notice later (page 2675), this formula weighs differences by the values themselves.

A8. This is not the best method, but standard method like Dupuy et al did. One reason we employed for the comparison between papers.

C9. p. 2662, Eq. (3): Dividing by the average of the measurements of these two instruments is reasonable when the values are near each other and well validated. Now the values often differ by a large amount and perhaps it would be wiser to use as a reference measurement values from the already validated instruments?

A9. Yes, we agreed. But the most of measurements has different local times. We would like to keep SMILES has an advantage for

C10s. p. 2663, line 4: Harris (2002) should be Harris (1998)

Fig. 3: Some misspellings: panel->panels, antscan->?

All figures: The font size of labels is quite small.

A10s. We have improved all of these. Figures are 1.5 times larger now.

Best Regards,

Yasko