Author reply to Referee #2

Ground-based assessment of the bias and long-term stability of fourteen limb and occultation ozone profile data records

Hubert *et al.*, Atmos. Meas. Tech. Disc., 2015, 8, 6661-6757.

Thank you very much for your detailed review of our paper. We are glad you liked it and appreciate your kind words. Our response to your questions, comments or suggestions can be found below, with different text formatting for referee comments and author replies.

GENERAL COMMENTS

This is an excellent and well written paper. It is a long paper but I think that it needs to be this long. It is a paper that people will reference and I don't see how the information provided in the paper could be cut back without loss of value. The paper is as long as it needs to be. I have made some minor suggested corrections below. The paper is very close to being acceptable for publication in AMT and I commend the authors for writing a paper that will be of wide interest to the ozone measurements and research community.

SPECIFIC COMMENTS

Page 6664, line 8: Was there a specific reason why you decided not to include SAGE I in the mix?

> Not really. Our main objections were the short record (34 months, Feb 1979 – Nov 1981) and a lower availability of correlative data in the early 1980s. This limits the estimation of data quality, especially stability. E.g. for POAM II (another short, sparse record, 37 months) the drift uncertainty is above 10% per decade, which is not really helping most data merging/trend applications. Estimates of bias patterns will also be less precise, although we acknowledge the value of having at least some estimate, since including SAGE I data in a trend assessment can be very valuable if done correctly.

Page 6664, line 10: It is not clear to me what you mean by 'harmonized' in this context. You may need to state more clearly what you mean by this.

> Referee #3 suggested "consistent", which is what is now in the revised version.

Page 6664, line 17: Do you mean that the *differences* in the short-term variabilities are better than 5–12%?

> No, the short-term variability (i.e. the spread in the comparison results) is better than 5-12% for all records. This observable relates to the precision of satellite ozone profile measurements, but also includes other random terms (explained in Section 3).

Page 6667, line 2: Saying "to another profile representation" is a little too vague and I suspect that many readers will not know what you mean by this. I would suggest that you be more specific and say "from pressure to geopotential height (or vice versa) as the vertical height coordinate".

> Done.

Page 6667, line 5: But how many of the limb and/or occultation data sets that you are assessing extend down to the ground?

None. We rephrased to "… from the troposphere up to the stratopause…".

Page 6667, lines 15-18: Will readers know what you mean by "native" and "non-native in this context?

This distinction is made clearer, by replacing "in non-native representations" by "when the original profiles are converted to another vertical or ozone coordinate".

Page 6670, lines 4-5: When converting between vertical coordinate systems, do you assume that geometric height=geopotential height or do you also account for that difference?

Indeed, we did not account for differences between geometric and geopotential height since the differences (at most ~300m at the stratopause) are negligible compared to the vertical resolution of most satellite records. This assumption will be clarified in the text.

Page 6670, line 25-26: Are these the total number of profiles measured over the analysis period, or the total number of profiles analyzed over the analysis period? i.e. were some measured profiles excluded though QA/QC before being used in this analysis?

These are the total number of profiles measured over the analysis period after QA/QC. This was mentioned in the caption of Table 1, but not in the text. This is now clarified.

Page 6671, line 12: Is this a geopotential height grid or a geometric height grid?

The assumption in our analysis of geometric=geopotential height will be clarified in the text, as mentioned before. We will also clarify which of both is provided in the satellite data. Both in text and in Table 3.

Page 6671, lines 22-24: Is this a true bias or is it simply that the diurnal cycle in ozone results in the sunrise profiles being 8-10% below the sunset profiles above 35 km? i.e. would a state-of-the-art chemistry-transport model also show such a bias? The same question applies to the sunrise-sunset biases seen in HALOE as mentioned on page 6673, line 8.

Diurnal variation certainly plays a role in the difference between sunrise and sunset profiles. However, it is difficult to quantify how much of the bias is explained this way. Different authors conclude to different strengths of the diurnal variation (see e.g. Sakazaki, Parrish, Schanz). In the uppermost stratosphere 10% differences are possible, but the diurnal effect at 35km could be much smaller, just a few %.

In the text, we will add that diurnal variations may (at least partially) explain observed sunrise/sunset differences.

Page 6672, line 14: How is it that the SAGE III data are useable to 85 km altitude while the SAGE II data are usable to only 60 km altitude? What does SAGE III do differently that extends the uppermost altitude coverage?

The SAGE III instrument has a larger coverage of the UV band than the SAGE II instrument. As a result, ozone retrievals can be done up to the mesosphere.

Page 6673, line 10: 10% per decade is a very large drift that would significantly compromise trend analyses. So if the drift between SAGE and HALOE was 9% per decade I certainly wouldn't refer to this as insignificant.

In the entire paper we reserved the term "(in)significant" only for statements on statistical (in)significance. We will clarify where confusion is possible.

Page 6673, line 19: Does this latitude range apply to both hemispheres?

Yes, the revised text makes this clear (was already the case in Table 3).

Page 6673, line 25: It is not clear to me how you can retrieve a profile of geopotential height.

Clarified in text. We meant that the hydrostatic equation is used to reconstruct MLS GPH from a reference point (the 100 hPa level) and retrieved temperature.

Page 6675, line 14: Is this a fixed geopotential height grid or a fixed geometric height grid? Likewise for Page 6676, line 5, and elsewhere.

This will be clarified.

Page 6675, line 16: The UKMO acronym needs to be expanded here and, if possible, an additional phrase needs to be added regarding exactly which UKMO reanalyses were used for this purpose. Likewise for page 6676, line 7.

The UKMO acronym is now explained. Unfortunately we could not trace which UKMO analysis was used (e.g. Lumpe et al. (1997, 2002) and Hassler et al. (2014)).

Page 6677, line 18: I don't know what is meant by "the Chalmers v2.1 ozone profiles". Shouldn't this be "SMR v2.1 ozone profiles"?

The reference was to the group that developed the processer. Text is modified.

Page 6677, line 27: What is the procedure for determining when noise is an issue?

Such a procedure is subjective and up to each SMR data user to decide. In this paper, we compare co-located pairs of single profiles, so no averaging is done before comparison statistics are computed. We replaced "When noise is an issue, …" to "When a user needs more precise SMR data, …".

Page 6680, line 5: It's not clear to me what you mean by "scientific MIPAS data". Is there another kind of MIPAS data i.e. non-scientific data? The same applies to the term "scientific ozone data" on line 24. How do "scientific ozone data" and "ozone data" differ? If they're the same thing, I would suggest that you just refer to "ozone data". The same applies to the reference to "scientific SCIAMACHY data" on line 24 of page 6714.

This is indeed jargon, and confusing for the unacquainted. We now avoid the term "scientific" and make it explicit that the statements relate to another Level-2 processor.

Page 6683, line 20: I am not sure exactly what you mean by "atmospheric cycles". Do you mean the diurnal cycle in ozone? I think that you need to be more specific here.

We rephrased to "If satellite biases have a multiplicative nature then any time-dependence in ozone levels (e.g. seasonal, interannual, solar cycle) is divided out in the relative differences. Another advantage is …".

Page 6683, line 21: I don't know what an "ozone coordinate" is.

Replaced by "ozone unit" everywhere in the text.

Page 6684, line 1: And, for higher altitudes, rarely coincide perfectly with regards to the diurnal cycle in ozone.

The original phrase does not exclude the uncertainty from temporal mismatch due to diurnal cycle. We therefore kept the original phrasing.

Page 6684, line 13: I don't understand what is meant by "reduction of smoothing". Do you just mean where the application of smoothing plays a central role?

We meant the reduction of the uncertainties due to differences in smoothing between satellite and correlative instrument. We replaced "in which the reduction of smoothing, mismatch and auxiliary uncertainties play a central role" by "in which the reduction of the uncertainties due to differences in smoothing, geolocation and auxiliary data plays a central role".

Page 6684: Regarding dealing with co-location. I am not suggesting that you implement what I have detailed below in your paper, but you may want to give this idea some thought when next dealing with the issue of co-location. One option is to systematically account for lack of perfect coincidence when two measurements (V1 and V2) are being difference as: V1-V2= ad_lat + bd_long + cd_alt + dd_time + ed_SZA + f where the separation in latitude, longitude, altitude, time and SZA between the pair of measurements is d_lat, d_long, d_alt, d_time, and d_SZA respectively. By training this equation on a large ensemble of pairs of differences over some prescribed space and time domain, the systematic bias between the two data sets, and its uncertainty, is obtained via the f term. The remaining fit coefficients (a to e) capture the sensitivity of the differences to gradients in space and time. Coincidence criteria still need to be selected and used, but these can now be far more inclusive since absolute coincidence is no longer a requirement for determining the systematic bias. The key assumption that the application of this equation makes is that gradients in the field of interest (e.g. ozone concentrations) are linear in the space and time domain defined by the choice of coincidence criteria. An additional advantage of the use of such a method is that a robust estimate of the uncertainty on the systematic bias is also obtained. Such an approach would minimize mismatch uncertainties.

This is an interesting idea, certainly worth exploring in future work.

Page 6684, line 28: But 6h is far too large for higher altitudes where the diurnal cycle in ozone is strong. Isn't there a need to impose a rather strict SZA coincidence criterion at altitudes where ozone shows a strong diurnal cycle?

Schanz et al. (ACP, 2014) showed that both SZA and sunshine duration play a role in the diurnal effect, leading to larger peak-to-valley differences at the polar circle (up to 15%) than at lower latitudes (3-5%). The use of both parameters could improve the coincidence and reduce the effect of diurnal cycle on the results above 35 km. However, the application of more refined co-location criteria is outside the scope of this work. We will make this more clear in the text.

Page 6691, lines 2-3: But 6%/decade is the maximum uncertainty in the drift not the minimum uncertainty in the drift.

This phrase discusses the ensemble of estimated drift uncertainty over all stations in the ground network. Here, 6% per decade is the station with most precise determination of drift, while at most other the drift estimates are at least twice less precise.

Page 6693, line 2: Mitigation of what? Likewise for line 19.

"More details and mitigation follow in" replaced by "More details and a recommendation to avoid such representation-dependences follow in"

Page 6699, line 14: I think that it is incorrect to refer to these as metrological uncertainties. Metrological uncertainties can relate only to the measurement process itself e.g. to calibration standards and the like. Differences in airmasses therefore cannot be included in a metrological uncertainty. The same is true for the use of the word on line 16.

We replaced all occurrences of "metrology" by "metrology of the comparison".

Page 6700, line 6: But, given the structure of the diurnal cycle in ozone, is it expected that the sunrise and sunset ozone profiles will be identical?

We will add a comment on the effect of the ozone diurnal cycle on comparisons, so that it is clear that sunrise/sunset biases may be (at least partially) be due to non-instrumental effects.

Page 6700, line 7: I think that you need to be careful here, and throughout the manuscript, with your sign nomenclature i.e., for me, an underestimate of -10% is the same as an overestimate of +10%. So when you say an underestimate of -(10-15)%, I am no longer sure if the ozone is underestimated or overestimated.

This is done.

Page 6706, line 4: ERA-Interim definitely cannot be considered as an "actual measurement". It is no more an actual measurement than any of the reanalyses data sets included as auxiliary data with some of the satellite-based instruments.

This was confusing, so we rephrased to "… taken either from actual measurements […], or from reanalysis fields […]".

Page 6706, line 21: I presume that this climatology is an annually repeating climatology and therefore does no incorporate in any way inter-annual differences in temperature and pressure profiles nor long-term trends in temperature and pressure profiles?

Yes, this is an annually repeating climatology.

Page 6708, line 27: Why would there be increasing biases in the pressure readings?

Very interesting work was done by Stauffer et al. (AMT, 2014). They showed a more or less constant absolute pressure bias with altitude for some often-used radiosonde types. What we meant was that the impact on vertical registration becomes increasingly important towards the top of the profile. We will rephrase so this point is made clearer.

Page 6709, line 10: I would have thought that a lower signal to noise ratio would have caused the differences between ozone records to become less obvious.

The vertical dependence of the S/N ratio differs per instrument. The ground-based comparisons of some instruments (e.g. GOMOS, UARS MLS) are clearly more noisy around 15-20km than for other satellite instruments. We argue that this increase in observed comparison noise is due to the much lower S/N ratio of GOMOS and UARS MLS in this altitude region.

We made this point more explicit, replacing "differences between records" by "differences in comparison spread between records".

Page 6712, line 3: Again, I think that the term "metrological uncertainties" is not being used correctly here. Natural variability cannot be a component of metrological uncertainty.

We replaced all occurrences of "metrological uncertainties" by "uncertainties in the metrology of the comparison".

Page 6717, line 18: I don't know what is meant by "due to quality features in the pressure and/or temperature information". What, specifically, is a "quality feature".

We rephrased the sentence.

GRAMMAR AND TYPOGRAPHICAL ERRORS

While the paper is very well written, there are some places where the grammar could be improved. I have made some suggestions to that effect below. The suggested changes may, however, also simply reflect my own writing style and the authors should takes these changes as suggestions. I commend the authors on taking the time to develop a polished manuscript. Both reviewers and readers will greatly appreciate this.

Thank you for having such a detailed look at our writing. We have implemented nearly all of your typographical and grammar suggestions, except for the ones below.

Page 6703, line 12: Replace "more important negative" with "more negative" - I didn’t understand how the bias was more "important".

We rephrased as “more pronounced negative”.

Page 6704, line 6: Replace "representation, if one uses" with "representation i.e. if one uses".

The phrase now reads “We also noted a dependence of the MIPAS bias on the profile representation when the pressure and temperature data included in the MIPAS product are used to perform conversions.”

Page 6717, line 21: Replace "important biases" with "significant biases". It is not clear in what way a bias can be important.

Now rephrased as “more pronounced biases”, since we reserved “significant” for statistics statements.