Interactive comment on "On the improved stability of the version 7 MIPAS ozone record" by Alexandra Laeng et al.

Anonymous Referee #1

Received and published: 6 December 2017

This is a very brief paper that assesses the stability of the new version 7 MIPAS ozone data product, which incorporates new time dependent detector calibrations. This is an important improvement in the MIPAS record and it should be documented. This paper could eventually be suitable for publication in AMT; however the current state of the paper is seriously lacking in important details, as listed in the comments below, some of which I consider major and a barrier to understanding what has been done and interpreting the results. A second major point that needs to be addressed, is that the work is rightly motivated to assess/use MIPAS for long term trends; however, the effect of the correction on derived trends is not addressed. The authors should revise or update previously reported trends using MIPAS data or perform at least a basic trend analysis on the old and new data records.

Page 2-3: The description and brief discussion on the new ozone retrieval algorithm (section 2) needs more detail/rigor. These changes and potential improvements are an important part of understanding the changes to the ozone product and several questions arise. For example, the text implies the ECMWF 2D temperature fields are used in the forward calculation for the temperature retrieval. This doesn't make sense. Are they used as the a priori for the temperature retrieval, or they used instead of the retrieved temperature in the ozone retrieval? What is the impact of using WACCM CO_2? How/why is it appropriate to fit the background up to 60 km? Why was it fit to only 33 km previously? What is the quantitative effect of the joint retrieval with H_2O? How is it that the max vertical resolution (6 km) is less than the max sampling (8 km)? Is the FR processor unchanged? Have the averaging kernels shown in Fig 1 changed from the previous version in a substantial way? If so, this should be pointed out, otherwise maybe consider not including the figure.

How exactly is the collocated MIPAS profile shifted to "prolong" the sonde profile? (I think "extend" is a better word to use rather than prolong). Is there some averaging of the two in an overlap region? At which altitude does this become important in the comparisons in Fig 3? The smoothed sonde profile should be explained, and not just presented with "edge effects". Presumably it trails off near 30 km because the authors have assumed a vmr of zero above the sonde measurements when applying the averaging kernel?

Is this level of "improvement" in bias between versions really significant in the Boulder comparisons? What does the statement mean exactly that "all biases are significant at 2 sigma level"? If so, then what can be said about the increased bias between 15-18 km? Is this also significant?

The discussion around the change in bias with respect to ACE and MLS around the 35 km vmr peak should be edited for clarity. It seems in both cases, the bias has increased in value, but for ACE it started out negative and is now closer to zero and for MLS it started out positive and is now more positive. This is not really what is implied with the

current discussion and as it stands it is difficult to reconcile.

Figure 5 and surrounding discussion is hard to interpret without a quantitative explanation of the non-linear coefficients and how they are applied, i.e. is an offset of -2 nW/sr/cm2/cm-1 a substantial change or not? In all cases shown, the year to year variability in the NLC oscillates about a trend. Is there an explanation for this? Are the NLC changes linearly interpolated when applied to the data as implied by the plots, or fit to a line, or something else? Are these NLC changes at 16 km and 51 km typical or extreme? What about altitudes between 16 and 51 km, such as 20-40 km where the trend results are quite important? Are the changes always monotonic with time? Why is spectral band D always unchanging?

The multilinear regression model used to assess the drift should be explained briefly so that a reference to the previous papers isn't necessary for a basic understanding of what is done here. A plot of the fit and the residuals would be informative. How linear is the correction and/or the remaining drift? Is it a good assumption to model it as linear in the regression?

Technical corrections:

Page 2 line 6: phrase in the past tense, something like, "the aging of the instrument was observed to change the detector response"

Page 2 line 12: what is a "parent" dataset?

Page 2, line 29: What is the "nominal mode"?

Page 5, line 2: "contrast" is not the right word to use to describe the difference in averaging kernels

Page 8, line 7-9: not a well formed sentence

Page 9: line 2: revise the wording of "a lot fewer"

C3