Response to the review of Anonymous Referee # 1 on paper:

Atmospheric ozone above Troll station, Antarctica observed by a ground based microwave radiometer

M. Daae, C. Straub, P. J. Espy, and D. A. Newnham

Earth Syst. Sci. Data Discuss., 6, 513-540, 2013

We thank the reviewer for the helpful comments and suggestions on our submitted manuscript. We have carefully considered all of the comments and used them to improve our revised manuscript. The methodology and analysis of ozone observations in this paper closely follows those described in our recent publication in Earth Syst. Sci. Data [Straub et al., 2013] reporting carbon monoxide observations. Please find a point to point response below; we print the referee's comment first followed by our response.

2.2 Dataset

Line 28, Page 517: *The authors described that the threshold of noise level is 0.71 K. Does this threshold mean peak-to-peak value or 1-sigma level? In Line 20 of Page 518 there is another noise level threshold of 0.4K. What is the difference between those two thresholds? In Figure 3b, there seems to be systematic wavy feature in the fitting residual indicated in red line. Does the noise include such wavy feature not only random noise*

We separated the description of these noise levels in the paper as they take action on the dataset during different operations and for different purposes. The 1-σ noise level on the integrated spectra, measured far from the line centre, is typically less than 0.2 K. However, occasional contaminated spectra will raise the noise level far beyond this and result in an unrealistic profile inversion. To prevent this, we only consider integrated spectra with a noise level lower than 0.71 K. Once the data have been inverted to ozone vmr profiles, the residuals, that is the integrated spectrum minus its fit, show small remnants of baseline components that were not included in the fit. While the 1-σ random noise component is less than 0.2 K, the variation due to the un-fit baseline components (the long period oscillation in Fig. 3b) in the paper) give an effective noise about twice this. Figure 3b) in the paper, also shows that while these components can vary from day to day, the total (random + unfit baseline) variation in the residuals remains on the order of 0.4 K. The fitting program will adjust the ozone profile so that the calculated spectrum matches the observed spectrum to within the specified noise limit. By setting a noise limit of 0.4 K, rather than 0.2 K, we ensure that the ozone profile is not modified in order to fit these small baseline components. We have added information to the paper hoping it will clarify the two noise levels we are operating with:
Line 103, Page 4, revised paper: The 1-σ noise level on the integrated spectra, measured far from the line centre, is typically less than 0.2 K. However, occasional contaminated spectra will raise the noise level far beyond this and result in an unrealistic profile inversion. To prevent this, we did not consider consider integrated spectra with a noise level greater than 0.71 K, that is 2-σ above the average noise level.
Line 126, page 5, revised paper: Removed explanation of the 0.4 K noise level.
Line 146-155, page 5, revised paper: Added explanation of the 0.4 K noise level.

3.2 Results of the retrievals

Line 26, Page 519: *It is reasonable way to evaluate the actual sensitivity of the observation, i.e. independence from the a priori by using the measurement response, but I think that the technical term of measurement response is not so common. So I think that the authors should give the definition of measurement response more definitely and justify the appropriateness of a threshold value of 0.8 by quoting appropriate references.*

The reference to the term "measurement response" as well as the "rule of thumb threshold of 0.8" is now given and we have replaced the text as follows:
Line 165-175, page 6, revised paper: "At a given altitude, the sum of all the AVK contributing to that altitude represents the degree to which the retrieved value there has been driven by the information from the measurement [Christensen and Eriksson, 2013, and references therein]. The value of this sum is generally referred to as the measurement response [Baron et al., 2002; Rodgers, 2004] and is shown by the black line in Fig. 3c) in the paper. The solid part of the curve shows where the measurement response is greater than 0.8, indicating that between 3 hPa and 0.02 hPa (about 38 and 72 km) the retrieved profile has a high degree of independence from the a priori [Christensen and Eriksson, 2013]. Outside of these altitudes (i.e. below 38 km and above 72 km) the measurement response weakens (shown by a dashed curve) and vmr values in these regions should be interpreted with caution as the information from the a priori becomes important."

Line 8, Page 520: *The authors described a weak secondary maximum of 1ppmv at 91km. However, both the AVK and the measurement response are almost zero at this altitude, so I think this statement is meaningless. If the authors think this statement is significant, please clarify the reason. I think that similar description 3.2ppmv at 93 km in line 12, Page 520 is also meaningless.*

We agree and it has been removed. We instead added this to Line 181, page 6 in the revised paper: The secondary maximum is also seen in the summer and winter profiles, but the measurement response is < 0.8 indicating that the data at these altitudes are dominated by the a priori.

3.3 Error characterization

Line 7-9, Page 521: *The treatment of calibration error is not clear. The hot/cold-load intensity scaling, standing waves, and the tropospheric correction factor are independent and completely different type of quantities. There is no quantitative description about the detailed breakdown of the calibration errors. Why those different types of errors are imposed only in the tropospheric corrections? Why an upper limit of ± 10% is appropriate value? Please explain the treatment of the calibration error more in detail and more quantitatively.*

For the estimation of the errors on the spectra induced by the instrument we take into account uncertainties in the hot/cold load calibration temperatures, standing waves on the spectra, the line of sight and the tropospheric correction, and refer to these collectively as the calibration error. Although ideally one would like to test the effect of each on the spectrum individually, this is not possible as their spectral responses are virtually identical, and therefore their effects cannot be individually identified or separated. However, the parameters in the calibration equation can be randomly varied by their typical measurement uncertainties to heuristically determine the cumulative uncertainty. It is found that the collected calibration uncertainty has the same contribution and spectral signature to the uncertainty in the spectrum as varying the tropospheric correction factor alone by less than ±10% (which includes the contribution of the uncertainty in the tropospheric correction factor itself) [Jarchow and Hartogh, 1994; De Wachter et al., 2011; Straub et al., 2013]. Thus, the total calibration uncertainty is accounted for by imposing this upper limit on the uncertainty of the tropospheric correction factor. Since this is a minor contribution to the overall error budget, we have re-worded the text slightly to try to make this clear, and added references where the reader may find the detailed derivations:

Line 205-210, page 7, revised paper: For the estimation of the calibration error we take into account the uncertainties in the hot/cold load calibration, standing waves, line of sight and tropospheric correction factor, and find that it is equivalent to, at most, a 10% variation of the tropospheric correction factor [Jarchow and Hartogh, 1994; De Wachter et al., 2011; Straub et al., 2013]. Thus the total calibration error, including the uncertainty in the tropospheric correction, is taken to be 10% of the tropospheric correction factor.

4.4 Temporal variations

Figure 6, Left panel: *In the time series plot of VMR at 0.04 hPa (upper left panel), the black data points (BAS-MRT) seem to be distributed bimodal in spring and autumn seasons. One group of data points show good agreement with SD-WACCM rather than MLS, and the other group of data points are concentrated almost zero. Are these nearly zero values of VMR in spring and autumn are real or artifact? If these data points are artifact, the authors should eliminate these data points from the statistics.The black line (BAS-MRT) and red line (MLS) at 0.04hPa show apparently good agreement in spring and autumn, but in fact, there are very few black (BAS-MRT) data points near the black line. This is very strange to me. The black line is in between the bimodal distributions. If the nearly zero values are artifacts, then the present black line should be meaningless. Such a bimodality of VMR values can be seen at 0.39 hPa (middle left panel) between Jan-09 and may-09 as well. In my impression, such bimodalities are due to artificial effect occurring in the retrieval calculation. The authors did not mention about these bimodal distributions in the manuscript. I think that the authors must discuss more carefully and more in detail whether these bimodalities are real or artifact. In case the bimodality is due to artificial effect, the authors should eliminate the wrong data points from the statistics. In the right panel of Figure 6, the correlation coefficients at 0.04hPa are high and almost the same both for MLS-BAS and WACCM-BAS. However, in the left panel, the apparent correlation between the blue and black line (WACCM-BAS) is clearly worse than that of the red and black lines (MLS-BAS). In my impression, the correlation coefficient in the right panel seems to be wrong.*

During winter or summer, the middle mesospheric ozone remain in near complete darkness or sunlight, respectively, the entire day. Thus, the odd oxygen is present either as ozone (winter darkness) or atomic oxygen (summer sunlight). This accounts for the high and low ozone vmr's measured by both MLS and BAS-MRT during those seasons. During spring and fall the mesosphere experiences diurnally varying solar illumination conditions, and thus the ozone undergoes a strong diurnal cycle. This is described in the introduction of the paper. Since the ascending and descending orbits of MLS occur in sunlight and darkness, respectively, both the MLS and BAS instruments measure alternately low and high ozone concentrations each day. The WACCM-SD results are only for midnight, and thus represent only the dark, high-ozone values.

The lines represent the low pass filtered data, which removes the diurnal cycle and yields an average day-night value for MLS and the BAS radiometers, much the same way a mean meteorological temperature is reported. Thus, the low-pass filtered data contains only variations with periodicities greater than 20 days, and the correlation analysis depends only on the synchronization of the variations present in the data and not on their magnitude. Thus the correlation analysis would be valid for all seasons.

We now mention the diurnal cycle of ozone clearly seen in spring and fall when describing Fig. 1 (Figure 6 in the paper). We have also separated the intercomparison with MLS and WACCM into two different figures, Fig. 1 and 2, (Figure 6 and 7 in the updated paper) for easier comparison of the data points to its belonging fitted line. We also repeat the data sampling time in the Figure captions.

4.5 Comparison of profiles

Line 25-26, page 525 and Figure 8: *The authors discuss the data quality by using only one averaging profile. There is, however, strong seasonal variation between winter and summer, especially in the middle mesosphere, as the authors mentioned in the manuscript. I think that the authors should compare the profiles at least summer and winter seasons separately in order to evaluate the quality of retrieval results.*

Thank you for suggesting to differentiate between winter and summer in this section too. The result can be seen in Fig. 3 below (Fig. 9 in revised paper). I have labeled the respective co-incident data sets with (m) for MLS and (w) for SD-WACCM to highlight that the co-incident dataset represent different hours. Your suggestion made us aware that we should use the 2-σ systematic errors in addition to 1-σ, and this has now been added to the figure. Thanks to the reviewers suggestion, we also discovered and

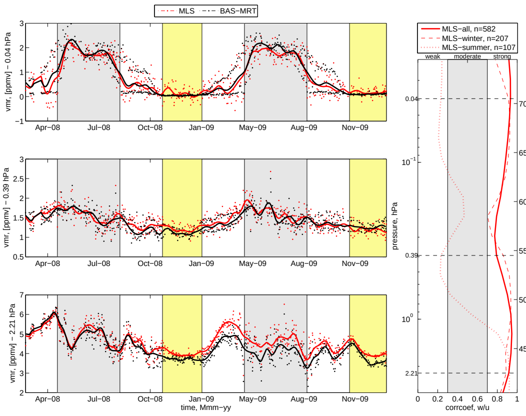

Figure 1: Time series of co-incident data between MLS and BAS-MRT at three independent pressure levels. Data are taken near 00 UTC and 15 UTC. *Left panels*: Dots mark the individual data points for the datasets, and solid lines show the low-pass filtered data with a 20-day cut-off. Grey shading marks the winter season defined by solar elevation angles below -30°, while yellow shading marks the summer defined by solar elevation angles above 0°. The areas without shading represent spring and fall, and in the mesosphere the diurnal signature can be seen in the data points. *Right panel*: Calculated correlation coefficients for the low-pass filtered data between BAS-MRT and MLS. The thick solid line is for the entire measurement period, while dashed and dotted lines represent the correlation coefficients for the winter and summer respectively. The number of correlated data points (n) is listed in the legend for the respective calculations.

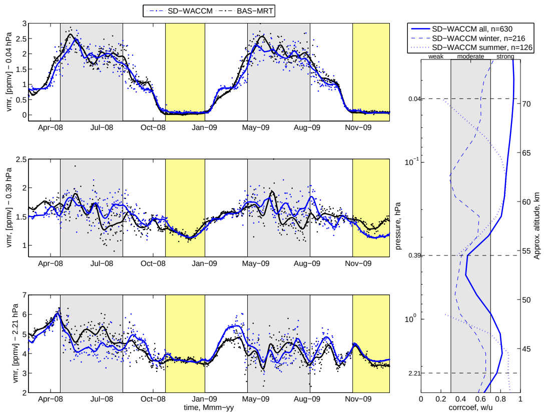

Figure 2: Same as Figure 1, but for BAS-MRT and SD-WACCM. Data are from midnight only.

error in the way we convolved the errors, which is now corrected in the revised manuscript. The text in section 4.5 in the revised paper has been edited to describe the new version of Fig. 3 below (Fig. 9 in revised paper), and reads as follows:

The comparison between O_3 profiles from BAS-MRT and Aura MLS, and those from BAS-MRT and SD-WACCM, are carried out following the procedures by Stiller et al. [2012]. Below, the BAS-MRT profiles that are co-incident with MLS (00 UTC + 15 UTC) and SD-WACCM (midnight only) at different times are highlighted with (m) and (w) respectively. Figure 3a) shows the overall mean of the O_3 profiles from BAS-MRT(m), MLS, BAS-MRT(w) and SD-WACCM. In general there is a very good agreement between the profiles. MLS shows slightly higher values at about 3 hPa than BAS-MRT(m) in both winter and summer (Fig. 3d) and 3g)), whereas SD-WACCM shows slightly lower values in winter at about 3 hPa than BAS-MRT(w).

Figure 3b) displays the average difference (bias) and standard error of that difference between the profiles together with the estimated systematic error of BAS-MRT. The total biases between the instruments and models are within the 2-σ systematic error estimate of BAS-MRT at all altitudes, and within 1-σ at most altitudes. While the altitude behavior of the bias varies from winter to summer, particularly between BAS-MRT(w) and SD-WACCM, these variations are sill within the 2-σ systematic error of the instrument and at most altitudes within the 1-σ. Thus, there is no significant bias between BAS-MRT(m) and MLS, or between BAS-MRT(w) and SD-WACCM.

Figure 3c), 3f) and 3i) display the standard deviation of the difference between the profiles, together with the individual and combined random errors of BAS-MRT and MLS, for overall, winter and summer respectively. If the instrument observations are truly co-incident in space and time, we expect the standard deviation between the observations to fall within the combined random error estimates of the instruments. In the mesosphere the combined random error estimates of the BAS-MRT and MLS instruments are large and accounts for the standard deviation of the difference between their observations. In the stratosphere, where the MLS measurement precision becomes better, there is more sensitivity to the spatial and temporal differences between the measurements. This is reflected by the standard deviation between the two observations becoming greater than the estimated random error, particularly in winter where small-scale wave activity is higher [Alexander et al., 2011]. However, it may be that the estimated random error of BAS-MRT is underestimated and should reflect the observed summertime differences.

Apart from the summer mesosphere, when O_3 levels are very low, the random deviations between BAS-MRT and SD-WACCM are larger than the estimated instrumental random error at all altitudes. This is likely because SD-WACCM does not resolve the short period or short wavelength variations that will affect the observations. Similar to the MLS comparisons in the stratosphere, the larger deviations in the winter than in the summer point to gravity waves as the source of these variations [Alexander et al., 2011]. However, it may also be that the estimated random error of BAS-MRT is underestimated and should reflect the observed summertime differences.

Technical corrections

Line 13, page 523: *One of the three pressure levels 0.21 hPa must be 2.21 hPa.*
Thanks for spotting this.
Lines 18, 24, and 25: *Replace green by blue for the color legend in Figure 6.*
Done. Thanks for seeing this.

References

Alexander, S. P., Klekociuk, A. R., and Murphy, D. J.: Raleigh lidar observations of gravity wave activity in the winter upper stratosphere and lower mesosphere above Davis, Antarctica (69°S, 78°E), J. Geophys. Res., 116, D13 109, doi:10.1029/2010JD015164, 2011.

Baron, P., Ricaud, P., de la Noë, J., Erikssona, J., Merino, F., Ridal, M., and Murtagh, D.: Studies for the Odin sub-milimetre radiometer. II. Retrieval methodology, Can. J. Phys., 80, 341–356, doi:10.1139/P01-150, 2002.

Christensen, O. M. and Eriksson, P.: Time series inversion of spectra from ground-based radiometers, Atmos. Meas. Tech., 6, 1597–1609, doi:10.5194/amt-6-1597-2013, 2013.

De Wachter, E., Haefele, A., and Kämpfer, N.: The Seoul Water Vapor Radiometer for the Middle Atmosphere: Calibration, Retrieval, and Validation, Geoscience and Remote Sensing, IEEE Transactions on, 49, 1052–1062, doi:10.1109/TGRS.2010.2072932, 2011.

Jarchow, C. and Hartogh, P.: Retrieval of Mesospheric Ozone from Groundbased millimeter wave Observations, Geoscience and Remote Sensing Symposium, 3, 1700–1703, doi:10.1109/IGARSS.1994.399542, 1994.

Paganini, L. and Hartogh, P.: Analysis of nonlinear effects in microwave spectrometers, J. Geophys. Res., 114, D13 305, doi:10.1029/2008JD011141, 2009.

Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and Practice, vol. 2 of *Series on Atmospheric, Ocean and Planetary Physics*, World Scientific, 2004.

Stiller, G. P., Kiefer, M., Eckert, E., von Clarmann, T., Kellmann, S., García-Comas, M., Funke, B., Leblanc, T., Fetzer, E., Froidevaux, L., Gomez, M., Hall, E., Hurst, D., Jordan, A., Kämpfer, N., Lambert, A., McDermid, I. S., McGee, T., Miloshevich, L., Nedoluha, G., Read, W., Schneider, M., Schwartz, M., Straub, C., Toon, G., Twigg, L. W., Walker, K., and Whiteman, D. N.: Validation of MIPAS IMK/IAA temperature, water vapor, and ozone profiles with MOHAVE-2009 campaign measurements, Atmos. Meas. Tech., 5, 289–320, doi:10.5194/amt-5-289-2012, 2012.

Straub, C., Espy, P. J., Hibbins, R. E., and Newnham, D. A.: Mesospheric CO above Troll station, Antarctica observed by a ground based microwave radiometer, Earth Syst. Sci. Data, 5, 199–208, doi:10.5194/essd-5-199-2013, 2013.

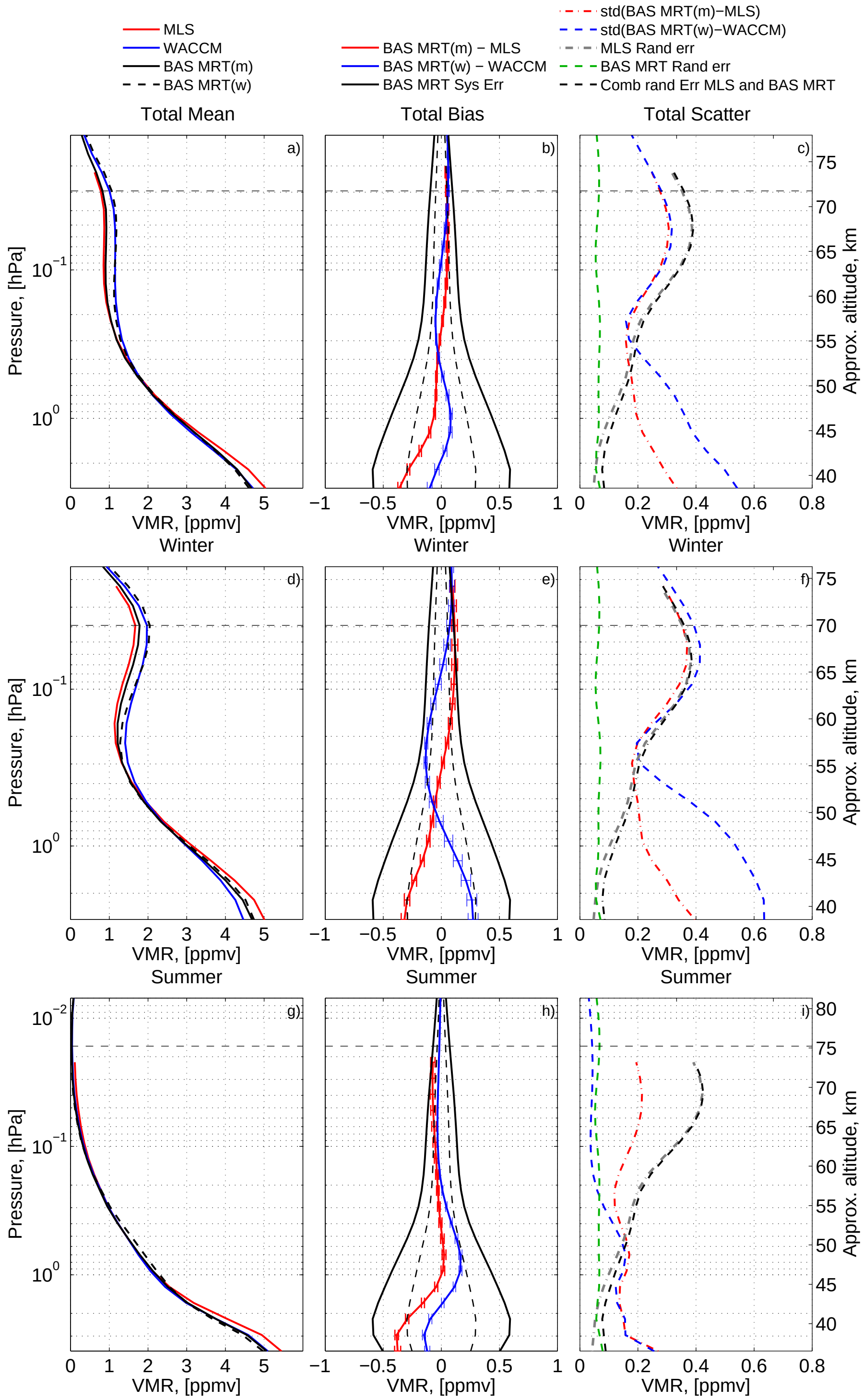

Figure 3: Intercomparison between BAS-MRT, MLS and SD-WACCM using co-incident profiles; (m) labels MLS and (w) labels SD-WACCM: *Panels a), b) and c)*: Overall comparison from available co-incident data. *Panels d), e) and f)*: Comparison from summer time (yellow area in Fig. 1 and 2). *Panels g), h) and i)*: Comparison from winter time (grey area in Fig. 1 and 2). *Left panels*: mean of midnight profiles. *Middle panels*: The bias with the 1-σ of the standard error of the bias. The black solid (dashed) line is the ± 2-σ (± 1-σ) of the total systematic error of BAS-MRT. *Right panels*: Standard deviation of the bias between the co-incident profiles of BAS-MRT to MLS (red) and SD-WACCM (blue). The combined (magenta) and individual (dashed) random errors of MLS and BAS-MRT for midnight values are also shown.