

More information is needed on SMR in Section 2, for instance Are the 2 side bands folded over on each other, or are they separate? (Could be part of more detailed explanation of Fig. 3). Section 2.1 has been updated with the following: “The output signal of the mixer is a function of the sum and difference of the frequencies of the input signals, therefore consisting of two sidebands. SMR is however run in single sideband mode so, of the two sidebands, only the one containing the signal of interest should be transmitted and the other one suppressed. This is achieved, before mixing, by accordingly setting the length of the arms of the Martin-Puplett interferometer.”

Is the scan continuous, or a step scan? It is a continuous scan with a constant speed of 0.75 km/s. This information has been added to Section 2.1.

Section 3.1: The method looks conceptually straightforward. Is there a reason the this has not been done before? The reason was a lack of personnel that would take care of this.

It looks as though the method described will work from 40 to 70 km, but what is done above that? Once the line has been identified via the spectra in the 40 – 60 km range, all the spectra in the scan are moved in bulk to the theoretical position. This is done by calculating the mean spectrum of all spectra in the scan, then the center frequency of the line is identified and the difference to the theoretical center is calculated. This same value is applied as a frequency correction for all spectra in the scan. After that a second correction, different for each spectrum, is applied, as described in Section 3.1 of the paper. We have clarified that point in Section 3.1.

Is the monthly a priori also divided into latitude bins? Yes, it is divided into 10° latitude bins. This information has been added to the text.

What is the “measurement response”? Is this the same as the Degrees of Freedom of the Signal (DFS)? The following was added to the text: “, where the measurement response is a measure of the degree to which the result may be contaminated by the a priori. It is defined as the sum over the row of the averaging kernel matrix (Rodgers, 2000)”. Also, the measurement response profile was added to Fig.5.

Section 3.2: Does the retrieval depend on temperature? If so, what temperature is used? What are its Biases? Yes, the retrievals depend on temperature. ERA-interim reanalysis data (Dee et al., 2001) is used up to 60 km and the Mass Spectrometer Incoherent Scatter model (version NRLMSISE-00; Picone et al., 2002) is used from 70 km upwards. Between 60 – 70 km, a spline interpolation of the two is applied. Information added to the text.

P. 7, Line 12 (regarding Fig. 5)- isn't it more that the concentrations above 90 km are influenced by the signals there? Yes, concentrations between 90 – 100 km are influenced by the signals there. In the text it is written that they are *also* influenced by higher altitudes measurements, as can be seen by the shape of the averaging kernels.

P. 9, l. 25: Please relate the vertical motions associated with the SAO with these signals. We have added the following explanation to the manuscript: “Although the SAO-induced variations in zonal mean vertical wind are too small to be measured directly, the accumulated effects can be seen in the distribution of the long-lived constituents like CO (Hamilton,

2015). Sinking of air, or at least anomalously weak rising motion, leads to the observed increase in the CO mixing ratios around the equinoxes.”

l. 28: clarify which is “this latitude band”. This has been changed to “high latitudes”.

P. 10, l. 1: Isn't it more accurate to say SSW's occur almost entirely in the N.H. We have changed “occur mainly in the N.H.” to “occur almost entirely in the N.H.”

P. 12: l. 2: What was the criterion for sufficient coincidences? Coincidences are considered sufficient if they reach at least the number of 10. This information was added to the text.

l. 11: If this is absolute difference, please denote with abs. value brackets. Why use absolute values- aren't you interested in whether the difference is positive or negative? If not, why? Absolute difference just means the subtraction $x_{\text{SMR}} - x_{\text{comp}}$ (without any normalization, as it is done for the relative difference), which of course can be negative, as it can be seen in the figures showing the comparisons. Absolute difference doesn't mean “the absolute value of the difference” as the reviewer suggests.

ll. 17-18: Wording is unclear- is the average difference being calculated, or the median, or some combination? Is $\bar{D}(z)$ the mean of $\bar{d}_{\text{abs},i}(z)$? Where does the median come in? We changed the text to “the median difference $\Delta(z)$... is calculated.”

The SEM expression is the standard for calculating the standard deviation of the mean. True, it is the same as for the standard deviation of the mean, only that here $\Delta(z)$ is the median, not the mean.

P. 13, l. 12: Why were the 2 different periods with different MIPAS spectral resolution lumped together? Does the MIPAS spectral resolution not make a difference? The reviewer is right. The comparison has been re-done to consider the two periods separately and the paper has been updated accordingly. Figures 9 and 14 have been updated to show comparison with FR-NOM and OR-NOM separately. The following has been added to the text: “Regarding FR-NOM mode, with the above mentioned coincidence criteria we found 6088 coincident measurements with SMR over the period July 2002 - March 2004. The profiles and differences between 50 and 70 km altitude, averaged over the whole time period and over the whole globe, are shown in Figure 9a. The median relative difference between SMR and MIPAS has a value of +8% at 50 km which decreases to reach -15% at 60 km, and eventually remains constant between 60 - 70 km altitude. No bias characterising a particular latitude or season over time has been identified for this specific comparison, since FR-NOM mode has been operational for only a short period.

Comparing with OR-NOM mode, 88902 coincident measurements have been found over the period January 2005 - April 2012. Figure 9b shows that the median relative difference between SMR and MIPAS remains positive and almost constant with altitude, staying below +15%. The main contribution to this difference comes from latitudes between -25° and +50°. In particular during northern summer, around +25°, where it reaches values above +40% between 65 and 70 km. Everywhere else relative difference values remain within $\pm 20\%$ (see Figure A1).”

The difference plots in Fig. 9a stop at ~ 70 km. What is plotted in the right panel above that- is it all a priori? Nothing is plotted above 70 km in the right panel, as well as in the center

panel. In fact, differences between the two instruments are plotted in these panels and, since MIPAS nominal data extend only up to 70 km, there can be no difference shown above that altitude. In case the reviewer meant the left panel, note that what is plotted above 70 km is SMR CO concentration which is retrieved up to 100 km. So no, it is not all a priori.

P. 14, Fig. 9: What is the explanation, or theory for, the negative peak in the comparisons ~ 85 km? Since this occurs for all global comparisons, it would seem that it is a feature of the SMR measurements and/or analysis. On the same subject, there are very large and fluctuating differences as functions of season and latitude. Do the authors have any explanation for these variations? As explained more extensively in the dedicated points, SMR-MLS differences are consistent with a known bias of MLS (positive systematic error of 20–50%). Regarding concentration differences with other instruments we have checked that they are not due to differences in vertical resolution. This has been done by smoothing SMR profiles with a Gaussian filter to account for differences in vertical resolutions; however, this smoothing didn't lead to significant changes in the profile. Also, we isolated MIPAS daytime profiles (which have a vertical resolution comparable to SMR) but the same relative differences with SMR were observed as in the globally averaged profiles. Moreover, it is unlikely that SMR-MIPAS relative differences are due to a misestimation of non-LTE effects from MIPAS, since systematic errors due to non-LTE were estimated to be less than 5% (in Funke et al., 2007, doi:10.1029/2006JD007933). This being said, the causes for the observed concentration differences, both globally and for different latitudes and seasons, are unknown. This information has been added to the Section "Conclusion" of the paper.

Note that the large number of comparisons make the SEM's very small and perhaps not useful. It is true that the SEM is sometimes not visible, but the fact that it is very small is a useful information. Moreover, there are some cases where it is not negligible and, for the sake of consistency, we should indicate it in all comparisons.

P. 16, l 3 MLS: Since MLS is another microwave instrument, it is surprising that its data are more different than that from MIPAS or ACE-FTS. The authors should comment on why this might be the case. Is it instrumental, or arising from the data analysis? It is a known bias of MLS, already referred to in the text in Section 5.3. MLS CO, as reported e.g. in Errera et al.(2019) and Livesey et al.(2018), is known to be affected by a positive systematic error of 20–50% throughout the mesosphere. This explains the observed SMR-MLS differences. Moreover, although SMR and MLS are both microwave instruments, they are not using the same spectral line, so there may be a variety of spectroscopic errors which would affect one instrument and not the other. However, CO is a species that radio astronomers have been studying for many decades; we rather suspect that its spectroscopy is better known than that of most other molecules. When MLS validation has been performed in Pumphrey et al. (2007), CO spectroscopy revealed itself to cause a 10% effect at worst; less serious than a variety of other systematic errors.

P. 17, ll. 4-5: Again, what is measurement response? Why a criterion of 0.75? For definition see point above. The value of 0.75 has been used for a long time. The reason is that, using a higher value as a threshold would result in the filtering out of a great part of the data.

Were OSO data only available for 2002-2007? The description seems to indicate an instrument improvement, especially after 2014. Most of the OSO data after 2014 has not been processed yet and the corresponding L2 data is therefore not available. The instrument improvement has been described just to shortly point out that OSO does not operate with a single sideband mixer anymore.

P. 19, Fig. 14: The difference of the comparisons between MLS and the other instruments is striking, as noted above. Also, the consistent decrease from higher to lower than the comparisons with altitude. As mentioned above, MLS CO is known to be characterized by a positive systematic error of 20–50% throughout the mesosphere.

Appendix: Latitudinal and seasonal differences in the biases at high altitudes- who is right? The causes for the observed concentration differences are unknown.

Looking at Figs. A2 and A3, it strikes this reviewer that the strong differences with MIPAS occur at the summer mesopause, the coldest region of the atmosphere. However, this difference does not show up in the ACE-FTS comparisons. This suggests that the MIPAS CO mixing effects might be too high, which could result from their temperatures being too low. At this level MIPAS temperatures must take non-LTE effects into account. These are notoriously difficult, and a negative error could lead to CO amounts too high.

1) non-LTE modelling of MIPAS CO has been extensively validated in Funke et al., 2007 (doi:10.1029/2006JD007933) and systematic errors due to non-LTE were estimated to be less than 5%. This information has been added to Section 5.1.

2) MIPAS CO (at least during daytime and hence polar summer) is rather independent on kinetic temperature, since CO daytime vibrational temperatures are dictated by solar excitation. A mapping of T errors is thus very unlikely, except for indirect effects via hydrostatics.

3) Further, the MIPAS kinetic temperature product used in the CO retrievals has been extensively validated in Garcia-Comas et al, 2014 (<https://www.atmos-meas-tech.net/7/3633/2014/>) and differences to correlative measurements were found to be less than 4K in the polar summer mesopause region.

4) the polar summer mesosphere, where the most striking differences occur, is characterised by very large vertical gradients in the CO distribution. In consequence, differences in vertical resolution between both instruments are likely important, there. While the vertical resolution of mesospheric MIPAS daytime CO is around 5 km (similar to that of SMR), nighttime vertical resolution is worse (>10 km) because of the lower sensitivity caused by smaller nighttime non-LTE emissions (no solar excitation). Therefore, it could be possible, that the SMR-MIPAS comparison is affected by the lower MIPAS nighttime resolution. This has been checked by restricting the comparison to MIPAS daytime profiles, but the resulting difference profiles are very similar to the ones considering data altogether. This suggests that differences in vertical resolution are not the cause of the observed biases. This information has been added to Section 5.1.

Technical Comments:

Page 1, Line 7: Much of the level 1. . . (The not needed) OK

l. 8: front end OK

l. 17: comparison instruments OK

P2, l. 19, 20- Suggest wording- “. . . obtained overall agreement over a two order of magnitude agreement” Not applied. We think the current wording is appropriate.

l. 22: front end OK

P3, l. 19 downward) scans Not applied. We would like to avoid using the word “scan” in the definition of scan.

P4, l. 7 front end OK

P. 10, Fig. 7- tick marks on abscissa can't be seen, also Fig. 8. OK

P. 15, l. 13: Does SCISAT-1 need to be explained and capitalized? The name of the satellite is written in lowercase on the ESA website so we don't think it needs to be written in capital letters.

l. 24: monotonically, not monotonously OK

P. 16, l. 17: better English would be . . . the MLS bias described is. . . OK