Major

There seems to be two major obstacles to overcome during periods when the PLL was malfunctioning: frequency shifting and line broadening. The approaches taken to overcome these obstacles seem reasonable, but more information must be provided in the paper in order to assess whether this is the case.

1. The paper is not entirely clear about this, but my impression is that from 2001-2003, the PLL was malfunctioning but only a frequency correction was need. In 2003-2004, the PLL was working and no corrections were applied. Then from 2004 onward, the PLL was malfunctioning but both the frequency and broadening corrections were needed. My naive assumption was that the broadening was a result of the frequency shift occurring during the time of a single spectrum, but this seems inconsistent with how the corrections were applied. In any event, the paper needs to have more information to better clarify the malfunctions of the PLL during the pre 2003 and post 2004 time periods. Section 3.3 has been updated to better clarify that we think two different kinds of PLL malfunctioning affected the instrument during these two periods: one before 8 October 2003 which had effect only at longer time scales and caused only frequency shifts; and another one after 8 October 2004 which have effects both within the integration time and at longer time scales, causing both line broadenings and frequency shifts.

2. For the frequency correction, a "basic" correction applied to all spectra in a scan is derived using differences in vertical gradients of brightness temperature (T_b) between O3 and CO. This is a novel idea and it seems to work well. However, Fig 3 compares T_b profiles for global averages, and one might expect to see differences according to latitude or season since vertical gradients in O3 and CO mixing ratios do change with season and latitude. Whether seasonal or latitudinal variations matter for this "gradient threshold" method is something that needs to be demonstrated. There should also be bars or shading to show the sigma in Fig 3. A panel which represents the fitted values of the slopes together with uncertainties has been added to Figure 3 and is described in Section 3.1. In this panel it can be seen how the difference between the two slopes is bigger than any natural variability. We therefore confirm that the threshold we are using is valid for whatever season and latitude.

3. A second, single-altitude frequency correction seems to be applied using Gaussian fits, but it is not discussed in sufficient detail at the top of p. 6. Why is a Gaussian thought to be the appropriate lineshape (later, thermal broadening is mentioned but what about pressure broadening at the bottom of a scan)? Was the central frequency and half-width fitted independently using some least-squares method? If so, how do the fitted widths compare with those expected? Was this second stage of frequency correction required for both 2001-2003 and post 2004 periods, or only post 2004 as for the broadening correction? The Gaussian fits which are discussed here are not to be confused with the ones performed during inversions to retrieve concentration profiles. Here we are using Gaussian fits exclusively as a method to estimate the center frequency of the line so to assess what is the frequency correction that needs to be applied to the spectrum. The fitted width, in this case, is of no interest. This is done, prior to the inversion process, for all scans. Another method to obtain the same information could have been to locate the position of the maximum peak of the line. But the use of one method or the other didn't present significant differences in the estimated centrum. In fact, even where the line is not Gaussian, the central part can still reasonably be fitted with one and the obtained center will be correct.

4. Figures 2 and 4 should have color bar legends to show the approximate range of altitudes covered. The paper has been updated with legends reporting the colours corresponding to some of the altitudes. Not all altitudes are indicated in the legend for the sake of readability.

5. It is important to show how the frequency shift and broadening varied with time, to give an idea of the magnitudes and temporal variation of both corrections. For example, monthly means of both quantities such as in Fig 7, used for the total number scans, would give the reader this kind of information. As described in Section 3.3, the broadening correction applied was one and the same for the whole period after 8 October 2004. The sigma value chosen for the new response function is the best compromise over this period, since it is the one giving the lowest overall area differences. Thus, the broadening due to the PLL malfunction was assumed to have the same entity over the whole period after 8 October 2004. The period before instead was not affected by any broadening. Regarding frequency shift, it appeared to be random and not follow any trend with time. We also observed no dependency on the satellite temperature. Therefore, we chose not to include any time series reporting the temporal variations of the corrections. This information has been added in Section 3.1.

6. Related to number 5 above, in the dataset there should be quality or error flags (or raw information such as mean scan frequency shift and broadening correction) so that users can further filter or weight these data for scientific study. The paper does not discuss any error estimates in the dataset. ARTS retrieval algorithms are based on the Optimal Estimation Method (OEM), thus statistical errors come from there (information added to the paper in Section 3.2). Moreover, in Section 3.1 we already explain how we discard spectra were CO line is not present. For all the other spectra, the PLL-generated frequency shifts and line broadening have been corrected, so there is no need of data flags.

Minor

1. abstract line 7 "The much of the..." -> "Much of the..." OK
2. p. 2 l. 3 "Schumann-Runge *bands* and continuum..." OK
3. p. 2. l. 27 "This is the first data being part of the..." is confusing Changed to “they are part of the…”
4. caption of Fig 3 "...and in the right place." -> "...and at the expected frequency." OK

5. p. 5 l. 13 "...a value higher than -0.0045 K/m..." Since all slopes are negative, this is confusing. I think what is meant is a slope with a magnitude less than 0.0045 K/m. Clearly, O3 has a larger magnitude in the vertical gradient of T_b, and CO a much smaller vertical gradient in T_b. Referring to slopes in Fig 3 is confusing because CO has a steeper slope, but a smaller vertical gradient. A panel which represents the fitted values of the slopes together with uncertainties has now been added to Figure 3. With the addition of this panel the text is not confusing anymore and it has been left unchanged.

6. p. 6 l. 7 define ARTS at first use OK

7. p. 7 l. 10 "Despite *the fact that* spectra above 40 km tangent altitude are considered..." OK

8. p. 8 l. 6 Is the given sigma FWHM or HWHM? Neither of the two. Sigma is the standard deviation of the Gaussian curve. The relation between sigma and FWHM is given by: FWHM

= 2*sqrt(2*ln(2))*sigma. We have not modified the text since this simply corresponds to the standard definition of a Gaussian function.

9. p. 10 l. 1 High CO vmr due to enhanced descent at the stratopause following an SSW was clearly demonstrated by Manney et al (Atmos. Chem. Phys., 9, 4775–4795, 2009). Reference added.

10. p. 12, l. 5-10 Not enough attention is given to possible biases due to differences in vertical resolution. There can be large vertical gradients in CO that depend on altitude and season. Between 70km and 90 km, where some biases are shown, CO increases by an order of magnitude over 20 km and differences in vertical averaging could be a factor. The mean averaging kernel widths between the various instruments should be given. For example, it is possible that ACE has at least a factor or 2 higher vertical resolution than SMR. Regarding comparisons with MIPAS, the polar summer mesosphere, where there most striking differences occur, is indeed 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 obtained considering data altogether. This information has been added to Section 5.1.

Moreover, we smoothed SMR and comparison instruments profiles using Gaussian filters with various FWHMs. This has been done in order to account for differences in vertical resolution, but the test didn't show any improvement in concentration differences. This suggests that differences in vertical resolution are not the cause of the observed biases. This information has been added to the Section "Conclusion" of the paper.