We would like to thank the anonymous referee 2 for the comments on our manuscript. We are grateful for advice helping us to improve the manuscript and take all points seriously into account. After a careful consideration of the points mentioned in this review, we however think that some of the points are not applicable to our manuscript. This will be discussed in detail point for point below. Generally, our impression is, that a misunderstanding concerning the scope and the focus of the paper might be the reason for the different opinions. This misunderstanding might be introduced by the word 'retrieval' in the title. We understand this word describing the action of performing a retrieval with a particular setup and the presentation of the respective results rather than an analysis on which setup is favorable for which circumstances.

1 Answers to specific comments

- 1) (discussion about which retrieval approach was formerly used and which vertical sensitivity was achieved): This comment seems to go in line with comment 3, where it is recommended to test several retrieval approaches with our dataset and to discuss the achieved vertical resolution. Here the respective discussion of former studies is demanded.

  We have already mentioned some previous studies in the introduction, but we agree that we might add additional information on former studies in the retrieval section. However, an 'extensive' review-like discussion is far beyond the scope and focus of the paper, which is already comprehensive. Furthermore, we doubt, that a discussion of the different studies using different retrieval approaches will be conclusive with respect to the achieved vertical resolution, since -aside from the chosen retrieval approach- too many parameters are different among the individual studies to be able to achieve comparability between those studies.

  E.g. the study by [Bevilacqua et al., 1985] mentioned in the review; they achieved a vertical resolution of 7 to 10 km (we did not find the 12 to 15 km stated in the review), which is better than the 16 to 22 km achieved in our study. But aside of the retrieval approach, basic properties of the analyzed spectra are different, which are expected to have a large influence on the vertical resolution (integration time is much longer, 10 h vs. 1h; the receiver temperature is much lower, 130 K vs. 1800 K; the frequency resolution is higher, they measure a different transition of CO, 115 GHz vs. 230 GHz; the retrieval is only performed up to 100 km, thus the ambiguity of linewidth introduced by the region above 100 km is not considered, etc.). Furthermore, [Bevilacqua et al., 1985] measure the vertical resolution using the weighting functions and not averaging kernels, so that the numbers of the vertical resolution are not directly comparable themselves. Similar arguments can be found for the other studies. E.g. [Aellig et al., 1995] (2 to 3 independent layers between 50 and 80 km) and [Forkman et al., 2003] (20 km) find a similar vertical resolution as we do, but likewise their underlying measurements are not directly comparable to the KIMRA spectra we used.

We would like to mention that we generally agree that it would be interesting to study the performance of different instruments and approaches, but this has to be done using comparable datasets, measured with different instruments but at the same location and conditions as e.g. recently done for H2O by [Straub et al., 2011]. But this is not the goal of the present manuscript and cannot be achieved using the available data basis.

Therefore, we will address your comment by putting a note on the difficulty of comparing the different studies conclusively in the retrieval section and will give a brief idea of the vertical resolutions achieved so far.

- 2) (Further information on regularization matrix): We agree, that the regularization matrix was not adequately described in the manuscript. We will add further information on this in the revised paper: The a priori covariance matrix $\mathbf{S}_{\mathrm{a}}$ was initially based on the respective WACCM CO standard deviation, but was then empirically modified in a way that the retrieval has enough freedom to fit the spectra of the whole course of the winter sufficiently. Major changes have thereby been made above 80 km altitude, where the available information on the CO statistics is least certain. Non-diagonal elements are zero, thus correlation between vertical layers are not considered. Like the a priori profiles, $\mathbf{S}_{\mathrm{a}}$ is also left constant for the complete retrieval run. It is defined in fractions of the a priori on the vertical retrieval grid. Fig. 3 shows the diagonal elements converted to absolute vmr ($\sqrt{S_{\mathrm{a}}^{ii}} \cdot x_{\mathrm{a}}^{i}$). The caption of Fig.3 states imprecisely that $\sqrt{S_{\mathrm{a}}^{ii}}$ is shown, which could be the reason for the confusion. This will be changed in the revised manuscript.

- 3) (Please include an analysis how the vertical sensitivity depends on the chosen regularization approach.) In general, we agree again that a comparison of the performance of different retrieval approaches might be interesting. But we do not think, that such an analysis fits into the present manuscript, which presents the results and characteristics of one particular retrieval run. The manuscript is in this sense similar to previous studies, but in many points already more comprehensive without this extension.

  To address the proposed extension seriously, an extensive discussion on the different retrieval approaches would have to be included, as well as a discussion of the results for every individual approach. Since this cannot be achieved in a single subsection, the focus of the manuscript (presentation of one particular dataset and its characteristics) would be strongly changed by this extension. The scope of the paper might become unclear so that the reader might even be confused.

  Furthermore, well-working retrieval setups, which are able to fit the measured spectrum adequately, should result in a similar information content of the retrieval result. This has been confirmed by [Eriksson, 2000] for ozone. As stated in the answer to comment 1), we agree that it might be interesting to proof this also for CO in the context of a general intercomparison of different CO microwave measurements, but this cannot be achieved within the scope of the present manuscript.

- 4) (A_{frac} vs. A_{vmr} in Sect. 3.4): We disagree with the statement, that the complete section 3.4 is based on the fractional AVK matrix A_{frac} instead of the vmr AVK matrix A_{vmr}. Both representations are mentioned and discussed in every single subsection and are included in all relevant figures equally. It is true, that we start the respective discussion with A_{frac}, which makes sense, since our direct retrieval result are profiles in fractions of the apriori, as explained in the manuscript. Furthermore, we already acknowledge in the paper, that also the vmr AVKs are important for comparisons. Thus, this representation including the transformation between both is discussed extensively. Therefore we cannot see, that here is something incorrect and has to be changed.

- 5) (Calculate and add smoothing error to error estimation): We disagree with this point. [Rodgers, 2000, p.48] states "the retrieval can either be regarded as an estimate of a state smoothed by the AVKs rather than an estimate of the true state, or as an estimate of the true state, but with an error contribution due to smoothing". We consistently refer to the first interpretation in the manuscript, thus treat the retrieval result as an estimate of a smoothed state, which has always to be seen with respect to the AVKs. It would therefore be misleading and wrong to include the smoothing error, since it would cause a doubled consideration of the smoothing effect; first by the use of the AVKs in comparisons and second by an additional error contribution.

  Furthermore, even if it would generally be possible to calculate the smoothing error from the WACCM statistics, the value of such an calculation is questionable. [Rodgers, 2000, p.49] states with respect to this point "To estimate it [the smoothing error] correctly, the actual statistics of the fine structure must be known. It is not enough to simply use some ad hoc matrix that has been constructed as a reasonable a priori constraint in the retrieval. If the covariance is not available, it may be better to abandon the estimation of the smoothing error, and consider the retrieval as an estimate of a smoothed version of the state, rather than an estimate of the complete state". This refers again to the first interpretation of the retrieval result and is actually the reason why we use this interpretation.

  And even if the relation between the WACCM statistics and the real world statistics could be established, it is still questionable to base the error estimation on a model result, since a major application of the presented time series is the verification of model data, thus this could lead to a circular reasoning.

  We have already mentioned the need the consider the AVKs in the respective parts of the document, but will address your comment by including this statement also in the conclusions section to underline its importance in the context of the used interpretation of the state vector.

- 6) (Comparison of the presented dataset to the FTIR data from the Kiruna station): We agree, that the coincident ground-based FTIR data should also be compared to the KIMRA measurements. The dataset presented by [Velazco et al., 2007] can of course not be compared directly to our data, but only the extension of this time series to later years including the KIMRA period. For the following three reasons (in connection with the fact that the paper is already comprehensive) we decided to focus on the satellite comparison in this paper, having in mind that the FTIR-KIMRA comparison should be done in future. First, the FTIR measures a strato-mesospheric column of CO, whereas the microwave radiometer provides altitude resolved information like the satellites. The focus of the paper is therefore an altitude resolved comparison of vmr profiles. Second, the FTIR measurements depend on sunlight, whereas the main interest of the microwave measurements is the polar winter night, which decreases the potential number of coincidences. Third, due to the spatial extend of the satellite data, the KIMRA-satellite relationship of the CO profiles is more important for future studies.

2 Answers to major technical corrections

The requested major technical corrections a) and b) refer to the structure of the document. Our impression is that this touches the area of personal preferences and can hardly be objectively decided. This is also demonstrated by the fact, that referee 1 states that the document is “generally well written and well organized”. Nevertheless, we will consider the individual points as stated below

- a) subsection 3.4.5 (“Summary of retrieval performance”) will be deleted and parts of the content redistributed

  subsection 3.5.3 (“Results”) will be renamed to “Results of the error estimation” to reduce the generality of the title. The contents are, however, not repetitive, thus should remain in the revised manuscript.

  subsection 5.5 (“Discussion of satellite comparison”) will be shortened by restricting the subsection to a discussion of the profile shape deviation without repetitions, therefore it will also be renamed to “Discussion of the profile shape deviation”

- b) We agree, that the three mentioned subsections 4.1, 4.2, and 4.3 of section 4 are not necessary, they will be deleted in the revised manuscript, the contents will be put together without any subsections in section 4.

  We do, however, not think that manuscript should completely be restructured. We are not aware of any objective rules on the maximum depth or number of subsections in a paper, except that the content should be structured logically using these elements. From our point of view, this is better achieved in the way we prepared the manuscript, than by a flat structure.

- c) We also considered combining the figures 13, 14, 15, and 16 into one single figure. However, only the profiles plotted now together in one of the individual figures are comparable, since the collocation criteria change from figure to figure.

In particular the number of days that go into the average changes among the figures, so that even the KIMRA profile is not the same in all the individual figures. Thus, a combination would result in a lot of curves in one figure, of which only subsets are comparable. This is difficult to visualize in a way, that it can easily and unambiguously interpreted by the reader.

The overall appearance of such a figure would be similar to e.g. figure 12, a figure for which referee 1 demanded a simplification. Thus, we decided against combining the figures 13 to 16.

3 Answers to minor technical corrections

- page 4248, table 1: caption will be changed and the row 'Mean number on measurements per day' will be included in the revised manuscript.
- page 4248, table 2: column 'references' will be added in the revised manuscript.
- page 4252, table 5: The column 'number of coincident measurements' was intentionally not included and replaced by the column 'days' with coincident measurements. This is because the number of coincident measurements is different for the individual instruments in an experiment, thus one column per instrument would have to be inserted, which makes the table confusingly large. Therefore, we decided to leave the table as it is.
- page 4213, line 21: will be corrected in the revised manuscript.
- page 4216, line 3: will be changed in the revised manuscript.
- page 4211, line 5: will be deleted in the revised manuscript.
- page 4211, line 11: Since we generally refer to the middle atmosphere, we intentionally stated, that 'the meridional circulation is directed from the equator to the winter pole'. Only the mesospheric circulation is 'directed from the summer pole to the winter pole', but this is not contradicted by 'from the equator to the winter pole'. However, the statement 'directed from the summer pole to the winter pole' is not true for the stratosphere, so that this phrase cannot be used generally for the middle atmosphere. We suggest rewriting the phrase to 'the meridional circulation is directed toward the winter pole' to avoid further confusions.
- page 4262, figure 11: the legend will be extended in the revised manuscript. For reasons of consistency we will also extend the legend of figure 12.

References

[Aellig et al., 1995] Aellig, C. P., Kämpfer, N., and Hauchecorne, A. (1995). Variability of mesospheric CO in the fall and winter as observed with ground-based microwave radiometry at 115 GHz. *J. Geophys. Res.*, 100(D7):14125–14130.

[Bevilacqua et al., 1985] Bevilacqua, R. M., Stark, A. A., and Schwartz, P. R. (1985). The variability of carbon monoxide in the terrestrial mesosphere as determined from Ground-Based observations of the $j = 1 \rightarrow 0$ emission line. *Journal of Geophysical Research*, 90(D3):5777–5782.

[Eriksson, 2000] Eriksson, P. (2000). Analysis and comparison of two linear regularization methods for passive atmospheric observations. *Journal of Geophysical Research*, 105:18157–67.

[Forkman et al., 2003] Forkman, P., Eriksson, P., Winnberg, A., Garcia, R. R., and Kinnison, D. (2003). Longest continuous ground-based measurements of mesospheric CO. *Geophysical Research Letters*, 30:1532.

[Rodgers, 2000] Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding.* World Scientific Publishing.

[Straub et al., 2011] Straub, C., Murk, A., Kämpfer, N., Golchert, S. H. W., Hochschild, G., Hallgren, K., and Hartogh, P. (2011). Aris-campaign: intercomparison of three ground based 22 ghz radiometers for middle atmospheric water vapor at the zugspitze in winter 2009. *Atmospheric Measurement Techniques Discussions*, 4(3):3359–3400.

[Velazco et al., 2007] Velazco, V., Wood, S. W., Sinnhuber, M., Kramer, I., Jones, N. B., Kasai, Y., Notholt, J., Warneke, T., Blumenstock, T., Hase, F., Murcray, F. J., and Schrems, O. (2007). Annual variation of strato-mesospheric carbon monoxide measured by ground-based fourier transform infrared spectrometry. *Atmospheric Chemistry and Physics*, 7(5):1305–1312.