Response to the comments of Reviewer #2 (Artem Feofilov)

We thank the reviewer for his very useful comments and suggestions and for making himself known. The manuscript has been considerably improved with these comments. Detailed responses to his comments follow.

Summary

Carbone dioxide is the main radiative cooler of the mesosphere and lower thermosphere (MLT), taking energy from the heat reservoir in the course of collisions with atomic oxygen and emitting it to space. Correspondingly, knowing its concentration, trends, and variability is of great importance for the atmospheric physics community. The present manuscript seeks to provide high-quality data retrieved from spectrally resolved radiances measured by MIPAS instrument in 4.3um CO2 band in limb observation mode, but its main purpose is to explain and justify the retrieval methodology and to demonstrate its high accuracy. The latter is stressed several times in the manuscript and this sets the aim high, but it also makes the requirements to quality tougher. In my review, I try to explain why I believe that currently presented methodology does not fit these requirements and what should be done to convince the community in robustness of the algorithm and, therefore, quality of the retrieved data. We think the methodology does fit the requirements. See the responses and the reasons given below.

Major comments

Most points of my concern have already been addressed in the detailed analysis of the first reviewer and they have to be addressed in the revised version of the manuscript.

All the points made by Reviewer 1 have been addressed. Please see the responses (more than 30 pages) to Reviewer 1 and also the revised manuscript.

Here, I would like to draw more attention to three issues, which I believe are crucial not only for the atmospheric researchers and potential users of the CO2 product, but also for the ones who are just learning atmospheric retrieval since overlooking these issues in the future might be costly for them and for the community.

1) For me, the major issue in the suggested retrieval methodology is the regularization approach the authors have chosen. This method is known to give stable results, which correlate well with currently accepted model of this or that parameter. On the other hand, this method is also known as the one, using which one can easily “throw the baby with the bathwater”, so it should be applied with great caution and the reasons for using a regularization approach should be explained along with regularization parameters (strength, matrix form, L-curve, residuals). We have expanded and gave more details in the description of the retrieval method in Sec. 3. As we have used this method for more than 30 species retrieved from MIPAS spectra and have been published so many papers on this we assumed it was very well known and no further description was necessary. Have in mind that this is not the first time we apply this method, and even have applied it to the very difficult problem of the joint retrieval of T and NO in the thermosphere (Bermejo-Pantaleón et al., 2011). The fact that we minimized the details does not mean that we are not doing a very careful and rigorous work. All the information required is
now given explicitly in the revised version. We have included a figure for the regularization; we have described the type of the regularization (matrix), and many more details (see new Sec. 3). The L-curve is not applicable to height-dependent regularization, as used here. We have included new figures with residuals (new Figs. 8a,b) and also another figure (new Fig. 9a) showing a statistical analysis of the Chi2 of the retrieval (a quantity which is informative about the fitting of the spectra in the inversion) for the 2 years of processed measurements covering all latitudes and seasons.

We should not fool ourselves with the results of self-consistency study performed with very mild conditions, which is used in the manuscript as a proof of concept. First of all, as the authors know, the problem is non-local and non-linear, Nonlinear and non-local with respect to the de-excitation rates, rad. transfer for the non-LTE populations, etc., but in the micro windows chosen, the problem is fairly linear in vmr (CO2), at least in the small range where the CO2 vmr is expected to vary. Of course the problem is “non-local” in a sense that not only the air at the tangent point contributes but also air along the entire ray path. This, however, is accounted by our retrieval, which inverts the full Jacobian and makes NO assumption that the signal comes from the tangent point only. Have in mind that because of the high spectral resolution of MIPAS the problem here is very different from that of SABER (for the case of H2O mentioned below).

Mild conditions: Reviewer 1 has also raised this point. We insert here the response to he/she. We have made two additional tests with rather unusual (extreme) CO2 profiles. One is the CO2 profile that was found by Rezac et al. (2015) in the inversion of SABER radiances (see the top-right panel of their Fig. 12), with very low CO2 vmr between 70 and 85 km and a pronounced peak near 90 km. The other case was chosen to check if our algorithm would be able to retrieve the expected effect of waves propagation on the CO2 vmr profile in the upper mesosphere and lower thermosphere. For this test, we looked at the variability of CO2 vmr in the WACCM model at all latitudes, longitudes and local times during a whole year of data co-located with MIPAS measurements and extracted the profile with the largest oscillation (see Fig. 4d). The results are included and discussed in the revised version (see Figs 4a-d and Sec. 4). They show that our algorithm is indeed able to retrieve such extreme CO2 profiles and if they are present in the real atmosphere we will retrieve them.

and the perturbation of the atmospheric profile should be performed in a more sophisticated manner than in the single-channel self-consistency tests. In 2008, two of the authors of the manuscript participated in a joint work dedicated to H2O retrieval in the MLT [Feofilov et al., 2009], where two research and one operational codes were compared, and the robustness of the retrieval scheme was tested. Even in this single-channel retrieval, the self-consistency test was performed for the conditions much harsher than those used in the manuscript, which is supposed to address much more complex scheme. In real atmosphere, the temperature and trace gas profiles in the MLT are perturbed by a mixture of gravity waves with amplitudes in temperature reaching 20-30 K. But, in the case of rapidly increasing (or decreasing) volume mixing ratios (VMRs) of trace gases, the perturbations of their vertical distribution should be also taken into account. Last, but not least is the modification of pressure in accordance with hydrostatic law.
We are fully convinced that our calculations are self-consistent and robust. We think our tests (particularly after including the tests for the new CO2 profiles) are already rather extreme (see above). Temperature is NOT a retrieved quantity but was taken from MIPAS simultaneous measurements, so there is no need to perturb it (only for estimating the propagated errors). See the analysis of the CO2 variability in WACCM simulations described above. WACCM simulations incorporate GWs rather realistically and we have not found extreme “wavy” CO2 profiles (maximum amplitude of 20%).

Another proof of the robustness of our algorithm is the very low percentage of not-converged profiles, less than 0.6%. And that cannot be attributed to a very strong regularization since the differences between the a priori (WACCM) and retrieved CO2 are very large (see Figs. 1, 2 about the WACCM-MIPAS differences included in the reply to Reviewer 1).

Last, but not least is the modification of pressure in accordance with hydrostatic law. We are not retrieving temperature, but CO2 vmr. Of course the p-T retrieved and used in the retrieval of CO2 was hydrostatically adjusted.

Correspondingly, a correct setup for the self-consistency test for the retrieval scheme suggested by authors should look as follows (see also two-channel self-consistency test in [Rezac et al., 2015]):

(1) atmospheric profile is taken from the model and is perturbed by some realistic gravity wave above certain altitude (i.e. 80 km). The new profile is pressure-adjusted and the vertical VMR changes are taken into account properly;
(2) This profile is called a “reference” one and is used in forward radiance calculations with the help of non-LTE model described by the authors;
(3) either another perturbation is applied to the initial profile or a smooth profile is taken and is sent to the input of the retrieval algorithm along with simulated radiances obtained at the second step;
(4) retrieved CO2 profile is compared with the one obtained at the first step and relative values of average radiance discrepancy are shown for each height.

The discrepancies obtained in this test will give an idea of the robustness of the method itself and about the errors introduced by applying the regularization scheme. To address the uncertainty of the final product, the same test has to be repeated for the cases, when the atmospheric profile is perturbed below 80 km, down to the ground. At this stage, the uncertainties of surface pressure need to be taken into account.

Essentially this is the procedure we followed except that we do not introduce a GW perturbation. Note that we are NOT retrieving temperature but CO2. Nevertheless we have additionally tested the retrieval with two rather extreme CO2 profiles (see above), and the algorithm is able to retrieve practically all of their structure.

2) The second point of my concern is the 90km feature reported in [Rezac, 2015], which is not present in the smooth vertical distributions shown in the manuscript. This requires special attention since SABER overlapped with MIPAS and the aforementioned work considered all possible mechanisms, which could lead to an enhancement of 4.3 um radiance. With the information at hand, one can suggest several explanations for this mismatch:
(a) experimental artifacts in SABER;
(b) experimental artifacts in MIPAS;
(c) smoothing of MIPAS radiance profiles;
(d) using large regularization parameter in the MIPAS retrieval.

This is where real physics starts into play and this is what I meant under “baby and bathwater” example above. If we exclude (a) and (b) and assume that modern non-LTE models are not capable of explaining this feature then one has to concentrate on the missing pumping mechanism. If the authors prove that neither (c) nor (d) is responsible for a lack of this feature in their data then one has to understand the fundamental difference between integrated 4.3 um band signal and spectrally resolved radiances of the main isotopologue picked by the authors for the retrieval, taking into account that the main contribution to broadband SABER channel still comes from the main isotopologue. To compare apples to apples, it would be interesting to see the comparison of overlapping SABER vertical profiles of 4.3 um radiance and MIPAS spectra convolved with the SABER instrumental function. For the same reason, it would be also good to have an access to spectral information from MIPAS.

Summarizing this part, I would say that the manuscript addressing CO2 concentration in the MLT and not addressing in details the 90km enhancement is not complete.

Artem, about this point there are actually two aspects. One that really concerns this paper, that is your points c) and d) (which are actually the same); and the other aspect, the 90 km feature in the retrieval of CO2 from SABER data reported by Rezac et al. (2015). We agree that understanding the peak at 90 km in SABER data is very important, and to know if it is a “real” peak or a SABER measurement artefact (or, why not, a SABER retrieval artefact) is a different aspect that is clearly beyond the scope of this paper. What we can assure, however, is that, if such peak exists in the real atmosphere, our algorithm applied to MIPAS spectra is able to retrieve it (see Fig. 4c in the revised manuscript); it is not smoothed out by the regularization. We have already mentioned above the significant differences between the WACCM a priori and the MIPAS inverted CO2 profiles (see Figs. 1 and 2 in the reply to Reviewer 1), and the retrieved CO2 profiles in MIPAS do show a significant variability (e.g., they are not completely smoothed, see the two new figures (Figs. 13a,b) showing all the profiles retrieved for April and December 2010. Also, MIPAS has, near 90 km, a vertical resolution of about 5km. We do not know how wide the SABER peak is but if it were wider we would have retrieved it. Opposite, peaks narrower than 5 km would not be fully resolved by MIPAS.

Having said that, we have some hints about this problem. In the earlier stages of the MIPAS retrievals, when we were using the previous “current” non-LTE model collisional rates (e.g., those before the inversion from MIPAS data), we did found very large peaks in the retrieved CO2 around 90-95 km in the polar summer, with values exceeding 500 ppmv. However, after the retrieval of the new collisional rates described by Jurado-Navarro et al. (2015), in particular, the Kvv rate affecting the 2.7 µm levels, those peaks nearly disappeared (see new Figs. 13a,b). Hence we suggest that the large peaks in SABER data might be caused by an incorrect Kvv collisional rate.

Note: MIPAS spectra are freely and readily available from ESA and we already made available some spectra to Ladi Rezac a couple of years ago.
3) The third ideological issue I see in this manuscript is an attempt of retrieving N variables from less than N equations. In the previous work [Jurado-Navarro et al., 2015], an optimized set of rate coefficients was obtained. I call it “optimized” and not the “true” one because the authors themselves write (p8, lines 4-8) that “the retrieved rates are expected to correlate with the errors caused by model parameter uncertainties”. In principle, this “error compensation” approach is widely used in this field (the most obvious example is kVT(CO2-O) rate coefficient, which is estimated from atmospheric observations and is used to interpret atmospheric observations, but its laboratory value is 2-3 times lower than the “atmospheric” one).

It seems there is here an important misunderstanding: MIPAS is a spectrometer, not a radiometer. We do not only have N equations, but, roughly speaking, M*N equations, where M is the number of spectral grid points per tangent altitude. This makes the retrieval formally over-determined. Possible ill-posedness of the retrieval caused by the fact that some of these equations are almost linearly dependent is fought by the regularization, which reduces the degrees of freedom of the retrieval.

About calling the collisional rates “optimized” or “true”, it is not really very important. The key point is that the errors of the retrieved rates are given. Thus, we prefer to call them “retrieved” rates. From the physical point of view it does not have much sense to talk about “true” rates.

Note that we DO include the errors of the retrieved collisional parameters in the CO2 error budget. We just do it in a way (see Eq. 2 in the revised version) that avoids the double counting of the contribution of those errors to the CO2 error budget, one, indirectly, through the collisional rates and the other directly through the non-LTE model in the forward calculation.

The only problem I see here is that kVT estimates have been performed with other profiles fixed, while in the present case we are dealing with simultaneous fitting of the radiance and CO2 profile, so one has to choose, what is known and what is unknown. If both the rates and CO2 profile are treated as unknowns, I doubt that the solution is unique since minimization of radiance discrepancy can have several minima.

It seems there has been also a misunderstanding here. In the Jurado-Navarro et al (2015) paper we aimed at retrieving the collisional rates only. However, we retrieved the CO2 vmr simultaneously with the collisional rates in order minimize the error of CO2 (avoid an systematic error of CO2) on the retrieved rates. MIPAS spectra have a very high resolution and therefore contain information on both the rates and CO2. In those retrievals we used very wide spectral regions (many more spectral points than in the current CO2 retrieval) in order to have information on both, on the collisional rates and on the CO2 vmr. Those retrievals are very computationally expensive, because of the large number of spectral points and henceforth it was not feasible (and not necessary) to use the MIPAS spectra for all years of measurements. The collisional rates should be "universal", i.e., independent of the atmospheric conditions except for their temperature dependence. For that reason we choose 4 days, 2 for solstice conditions (to cover the summer in both hemispheres) and 2 for mid-latitude (to cover spring/fall in both hemispheres). In this way we assure to cover all expected atmospheric temperature variability of the collisional rates. CO2 could have been fixed, e.g. taken from coincident geo-located ACE measurements. However we prefer to retrieve it since we have additional information on it. In order to be sure there was not a significant cross-talk between
the retrieved rates and the retrieved CO2, we compared the retrieved CO2 with the geo-located CO2 from ACE (see Fig. 7 in Jurado-Navarro et al., 2015). There you can see the nearly perfect agreement between MIPAS and ACE for equinox conditions (lower two panels). For solstice conditions (polar summer) the agreement is not so good. However, the retrieved CO2, simultaneously with the retrieved collisional rates that are consistent for all conditions (i.e., for the 4 days of measurements), is the one that gives the better overall fit to the spectra (smaller residuals) for all conditions.

In this paper we do the retrieval of CO2 and LOS but the collisional rates previous retrieved are kept fixed. The microwindows in this case are much narrower. In any case, we have checked (as required by Reviewer 1) that the retrieved CO2 (using the previously retrieved collisional rates) give also a very good fit (small residuals) in the whole spectral range, i.e. not only were CO2 is retrieved but also at frequencies where the collisional rates were previously retrieved (see new Figs. 8a,b). This therefore gives us confidence in the retrieved CO2 and confirms the validity of the retrieved collisional rates.

**Action:** We have clarified all these aspects in the revised version of the manuscript.

Moreover, the minimum of a multi-parametric function might be washed out. In the aforementioned joint work, it was just on the edge of detectability with the trace gas profile constrained by occultation measurements, and here we have another degree of freedom. In addition, the set of rate coefficients obtained in tropical region in this approach may be inconsistent with the one obtained in the polar region. In my opinion, the non-LTE rate coefficient retrieval should be based on a same approach as was used in previous works with maximum number of parameters is fixed and for CO2 I would use vertical profiles coming from ACE-FTS co-located with MIPAS since occultation measurements are not affected by non-LTE. In this case, the set of rates would still be an “optimized” one, but at least the solution would be more reliable and there will be more confidence in CO2 retrieved with these rate coefficients.

The only problem I see here is that kVT estimates have been performed with other profiles fixed, while in the present case we are dealing with simultaneous fitting of the radiance and CO2 profile, so one has to choose, what is known and what is unknown.

See comments above. In summary: 1) In MIPAS we have more information than in SABER (MIPAS is a spectrometer not a radiometer) so we have information not only on the rates but also on CO2. 2) For some conditions the retrieved CO2 coincides with the ACE CO2. Hence there would be no difference on the collisional rate if we retrieved CO2 or keep it fixed as ACE CO2. We cannot expect to have a perfect agreement between MIPAS and ACE for all conditions (as, for example, there is not between SABER and ACE, Rezac et al., 2015b) using a unique set of collisional rates, since the rates are not the only source of systematic errors in MIPAS and also because they are different instruments and using different techniques. Additionally, the single measurement collisional rates retrieved by Jurado-Navarro et al. (2015) do NOT show any significant dependence on latitude and season. Therefore differences in the CO2 in polar summer should not be attributed to different rates.

A comparison of the three datasets would be very valuable which we plan to perform in a near future.

**Minor comments**
- The LOS approach needs to be described in more details – in its current form, the main idea is not explained.

We agree. This has been described in more detail in the revised version (see Sec. 3).

- Horizontal inhomogeneities affecting the retrievals are mentioned and the methods of their reducing is suggested, but the explanation of “excluding opaque spectral lines” is not convincing. The tests show that the neighboring layers perturbed by gravity waves can affect tangent point retrievals even in transparent media. A sensitivity study is needed here, which would show the upper estimate of the horizontal inhomogeneity effects on the retrieval uncertainty.

The forward model calculations include gradients along the line of sight (LOS) in the non-LTE populations of the emitting levels caused by both kinetic temperature gradients as well as by different solar illumination conditions along the LOS (variable solar zenith angle). We have included this information in the revised version (was not clear in the submitted version). Therefore we expect the errors due to the actual inhomogeneities along the line of sight to be very small.

This information has been included in Sec. 3 of the revised version.

Temperature was retrieved from the same spectra in a preceding step and the non-LTE model is used online for the measured kinetic temperature. So local perturbations of the tangent point temperature should be included.

We understand that the problem of horizontal inhomogeneity is not a small-scale wave-structure but horizontal gradients or discontinuities which act over long distances; and we have accounted for them by including LOS gradient in $T$ and in the non-LTE populations (also caused by the variation of SZA along LOS). If we were sensitive to perturbation by gravity waves, this would mean that we could detect them, but we do not think so because we guess any effect over the whole LOS would average out. Maybe in the nadir could be detected.

- Neither hot oxygen pumping CO2 nor electron pumping of N2 is discussed. Do the authors believe that these mechanisms can be neglected at thermospheric heights?

We do not think so. See below the reply to Reviewer 1.

The hot O atoms was postulated by Feofilov et al. (2012) as an additional source of the excitation of CO2(v2) (15 μm) in order to understand the differences between the collisional rates of CO2-O measured in the laboratory and those derived from atmospheric measurements. Sharma (2015) found, however, that “The chance of a “hot” atom colliding with CO2 is therefore virtually nil.” in the MLT region. For the excitation of CO2(v3) near 4.3 μm would be even much less probable since v3 is more energetic than v2. Thus, there is no evidence of such excitation mechanism for CO2(v3). Also, this mechanism was not included in the recent retrieval of CO2 from SABER measurements (Rezac et al. 2015a).

About the auroral electrons: The excitation of CO2 via VV transfer from N2(v) excited by auroral electrons is more than two orders of magnitude smaller than the solar excitation in the 4.3 and 2.7 μm bands. See, for instance, the CO2 4.3 μm radiances measured by SABER. Also, no sign of such excitation has been observed in the analysis of the MIPAS CO2 4.3 μm spectra.
- Vertical resolution of 5-7 km and up to 10-20 km is stated. At the same time, CO2 falls off rapidly in the MLT and the stated uncertainty of its retrieval is much smaller than CO2 change at 5 km distance not speaking about 20km offset.

It is common (and suggested) practice to report uncertainty and altitude resolution separately. Our uncertainty estimates do not refer to a “point” in the atmosphere (where the mixing ratio is undefined anyway, see von Clarmann (2014)), but they refer to a result at a given resolution. It is not conclusive to demand that the variability within an air parcel described by a MIPAS data point shall be included in the uncertainty estimate of the characterization of the air parcel.

How do the authors explain this mismatch

See above: our error estimates characterize the uncertainty at a given resolution and can be compared only to the truth represented at the same resolution. Any assumption of a “true atmosphere of infinitesimal resolution” would lead to absurdities (c.f. von Clarmann (2014)).

and what are the real estimates of the retrieval uncertainty associated with vertical resolution?

One of the co-authors has shown (von Clarmann, 2014) that the smoothing error concept leads to absurdities. Thus we intentionally do not include the effect of smoothing in the uncertainty. The error due to smoothing involves either a “true” atmosphere of arbitrary resolution or a “true” atmosphere of infinitesimal resolution that is undefined. The only error estimate, which can reasonably be reported, is the uncertainty of the retrieval at the reported resolution, which characterizes the expected deviation between the retrieval and the true atmosphere degraded to the same resolution.