Response to Referee #1:

We appreciate the very helpful feedback from the referee. The referee's comments are listed in *italics*, followed by our response in blue. New/modified text in the manuscript is in **bold**.

In this paper, a new retrieval algorithm for temperature and O2 VER is introduced for the O2(1Delta) and O2(1Sigma) bands measured by Sciamachy. O2 VER and temperatures have been derived from these observations before; what is new here is that both are derived simultaneously, and self-absorption is considered in a consistent. The retrieval is applied to one year of data (2010), and temperature data are compared to ACE-FTS and Mipas. The Mipas comparison is particularly useful as Mipas was on the same satellite as Sciamachy, therefore providing close coincidences. The O2 airglow is highly relevant both for the accuracy of greenhouse gas remote sensing products, and for the energy budget of the mesosphere / lower thermosphere, and the data from this new algorithm provide a large step forward compared to previous publications. The paper is also generally very well written. However, I have some questions e.g., regarding the derivation of the prior error and the comparison to Mipas MA/UA data, as well as a few minor points listed below.

Thanks for the comments. Note we change number density notation from n[*] to [*] following the suggestion from referee#2. The description about prior error is clarified, and MIPAS MA/UA/NLC modes are compared with SCIAMACHY $^1\Delta$ band airglow temperature. See the responses below.

Line 249-250: Doesn't this imply an altitude dependent differently strong weighting, as the self-absorption affects the lower levels exponentially stronger?

Self-absorption indeed causes inaccurate results in the linear inversion. It is accounted for in the nonlinear retrieval system in this study. See the following response about the weighting between information from the prior and from observations.

Line 251-252: The statement that a prior error of 100 times the prior value leads to a weak to negligible prior constraint seems not correct in the lower altitudes affected by self-absorption: as there the prior profile is too low, and might be orders-of-magnitude too low, so is the prior error actually quite low. A climatology might be a better estimate of the prior values here, if available.

To the best of our knowledge, there is no available climatology for emission $^1\Delta$ O_2, except results above 60 km based on linear inversion (self-absorption not considered) from Zarboo et al. (2018). Actually, the prior profile is too high, not too low, in the stratosphere where self-absorption is significant. This is because we use a constant value for all altitudes in the prior profile, as indicated in lines 248-249:

"For each retrieval, we first conduct a linear inversion of the spectra and use the vertical mean value of the inverted [O_2*] profile as the prior values for the [O_2*] profile."

Therefore, the prior error is also a constant (100 times larger than the prior value, which is constant at all altitudes) and much larger than any possible [O_2*] values. Figure 3c, figure 5c, and figure 7c also show that the DOFS of [O_2*] are effectively one, supporting negligible prior constraint. To clarify, lines 251-252 are revised to

"**The corresponding prior error is set to be 100 times the prior, a constant value for all altitudes. This effectively gives no prior constraint to the [O_2*] profile and assures its information all comes from observations through near-unity DOFS of retrieved [O_2*].**"

Line 355: These missing points … are these related to high solar zenith angles? As during daytime the dominating formation mechanism is O3 absorption, the O2 airglow varies strongly from daytime to nighttime, and observations with high SZAs would provide very different (lower) values, the signal-to-noise is also low. This should be discussed somewhere, as you don't separate daytime and nighttime observations at high latitudes, and it should also be stressed in discussing your climatology of O2 airglow: it is a climatology covering a whole year of observations, but at a very specific time of day, about 10:00 local solar time.

Yes, these missing points are mostly at high solar zenith angles and the transition zone between day and night. Note we remove the ascending portion and SZA above 100° when averaging individual orbits into the climatology. This removes some ambiguous twilight data in polar regions. The sentences at lines 355-359 are revised to

"**These missing points are generally located at high latitudes and high solar zenith angles. In these transition regions between daytime and nighttime, the horizontal variation of airglow intensity is significant, which violates the homogeneous layer assumption for the retrieval algorithm. Retrieved data are often available for part of the ascending phase of the orbit at the summer hemisphere (most valid data are in the descending phase), leading to some repeated observations at the same latitude, although at different SZAs and potentially in the nighttime. To eliminate such a latitudinal ambiguity and nighttime data, we remove the ascending portion when averaging over multiple days and limit the SZA to within 100°.**"

The last sentence in the introduction is revised to

"**The algorithm is applied to one year of SCIAMACHY of limb observations, including the MLT mode, to construct a climatology of O_2 airglow and upper atmospheric temperature at 10:00 local solar time.**"

The sentence at lines 81-82 is updated to

"**The instrument was launched on board the Envisat satellite which was operational on a sun synchronous orbit with an equator crossing time in the descending node of 10:00 local solar time from March 2002 until April 2012.**"

Line 437: can you provide some idea why the A band has such a stable cold bias compared to the 1Delta?

The systematic difference between temperatures derived from the two bands is more clearly shown in Figure 15a. One possible reason is the propagation of temperature errors from lower altitudes in the A band, where strong self-absorption leads to diminishing observable emission signals. Spectroscopic errors may also play a role, given the slight difference between HITRAN16 and HITRAN20. One sentence is added

"**The low bias in the *A* band temperature is likely caused by error propagated from lower altitude where retrieving temperature from *A* band airglow becomes challenging due to strong self-absorption.**"

Line 468: "Mipas temperature retrieval in 2010 is only available below ~80 km": This statement is factually not correct. A) there are the middle atmosphere / upper atmosphere limb modes of MIPAS which scan up to 120 km respectively 170 km every ten days since 2007. These were coordinated with the Sciamachy MLT mode in such a way that corresponding observations are available every 30 days – about once per months. Observations in the MA/UA modes were carried out also in 2010, and temperatures were retrieved from these modes up to at least 120 km, see e.g., Fig 4 in Sinnhuber et al, JGR, 2022 for an example. Data are available on the MIPAS data server at IMK (https://www.imk-asf.kit.edu/english/308.php), and I am sure the Mipas team (e.g. Bernd Funke or Thomas von Clarmann for the MA/UA modes) would be happy to help in accessing and applying the data. If there are coincidence data between Sciamachy and Mipas for 2010 (and there should be at least 12 days) please do the comparison. B) Just as a caution, the nominal limb mode of Mipas scans up to 68 km, so values above ~70 km are probably dominated by the prior profile.

We add the discussion of the additional MIPAS mode in Section 2.4:

"**The measurement modes of MIPAS used in this study include the nominal measurement mode with an altitude coverage of roughly 6-70 km, the middle atmosphere (MA) mode covering 18-102 km, the upper atmospheric (UA) mode covering 42-172 km, and the noctilucent cloud (NLC) mode covering 39-102 km. The nominal measurement mode makes up the bulk of MIPAS measurements, whereas the MA and UA modes were available every at least 10 days, and the NLC mode only happened on a few days in 2010. We use the nominal temperature profiles from version 8 of MIPAS Level 2 data retrieved by ESA (Dinelli et al., 2021). Version 8 data from the other modes are obtained through the Institute of Meteorology and Climate Research in cooperation with the Instituto de Astrofísica de Andalucía (IMK/IAA) retrieval algorithm (García-Comas et al., 2012; Kiefer et al., 2021). The typical total errors are 0.5-2 K below 70 km and 2-7 K above (for MA, UA, and NLC modes). The typical vertical resolutions in the comparison range of this study are 3-7 km.**"

The sentence at line 468 is removed. Figures 16-17 are changed to only show up to ~70 km, and the text is revised to reflect that

"**To avoid interpolation artifacts, we use the extended profiles in MIPAS Level 2 data, which fill the space above highest retrieval level by a seasonally and diurnally varying climatology. As a result, the comparison should be limited to below the top MIPAS nominal tangent height at 70 km.**"

Comparison between SCIAMACHY $^1\Delta$ band-derived temperature in nominal mode with MIPAS MA, UA, and NLC modes are included at the end of this section. We will leave a more complete intercomparison/validation involving the SCIAMCHY MLT mode to future studies, as this manuscript will not focus on temperature validation.

"**In addition, we compare the O_2 $^1\Delta$ band-derived temperature with the MIPAS IMK/IAA temperature product for the MA, UA, and NLC modes as shown in Fig. 18. The numbers of coincidence with SCIAMACHY for these modes are ~20% of the MIPAS nominal mode, but they provide coverage above 70 km through the top of SCIAMACHY nominal mode retrieval. The SCIAMACHY-MIPAS temperature difference is consistent with the MIPAS nominal mode as in Fig. 16. The absolute temperature difference in the 70--100 km vertical range is generally within ±5 K, except the summer mesopause at northern high latitude.**

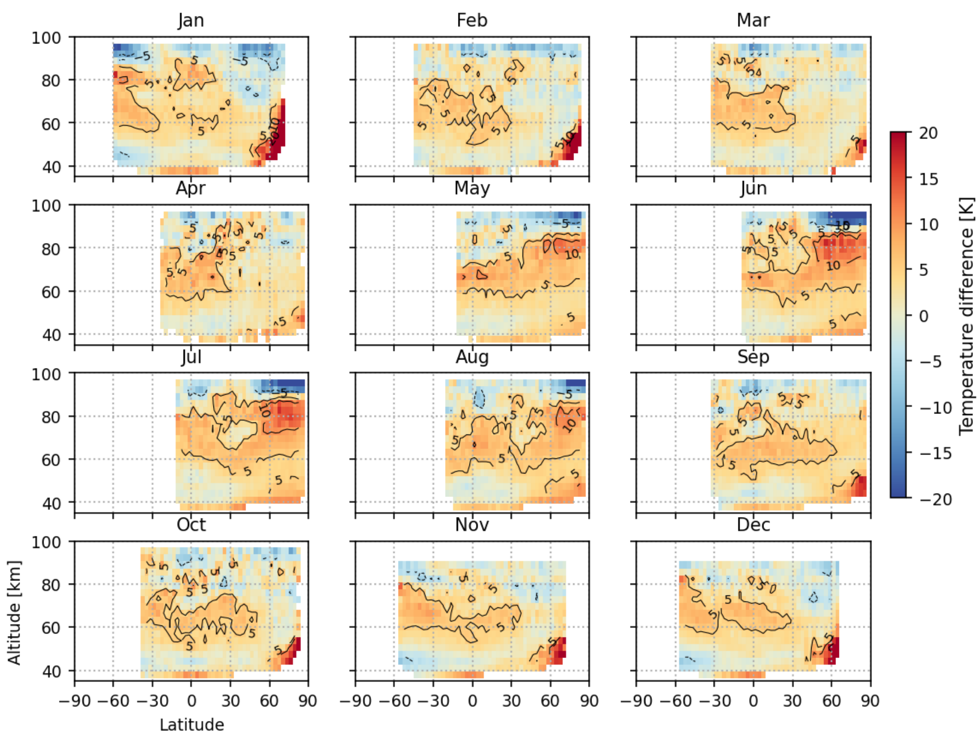

Figure 18. Similar to Fig. 16 but using MIPAS MA, UA, and NLC observation modes instead of nominal mode.

"

Figure 16: Here Mipas temperatures are used up to nearly 100 km – if they are from the nominal mode as you imply, the large differences above 80 km are to be expected, as the nominal mode scans up to 68 km only. It's rather surprising that the region 70-80 km seems to agree fairly well in most month.

We limit the comparison with MIPAS nominal data to below 70 km and add MA/UA/NLC modes to show above 70 km. See the previous response.

Minor points:

Abstract: I know they are commonly used, but nevertheless I found the use of the abbreviations (1Δ and A) for the bands slightly irritating. Could you use the full names (O2(a1Δg), O2(b1Σg+) at least in the abstract?

Thank you for the suggestion. We changed to the full names in the abstract.

Line 5: as the nominal mode only scans up to 93 km in 2010, how do you derive O2(1Delta) in 93-100 km?

93 km is the tangent height, but we retrieve layer properties using tangent height as the lower boundary of each layer. The top layer is assumed to be between the top tangent height and top tangent height plus average layer thickness, which is about 6.6 km. This adds up to about 100 km. The layering scheme is introduced in lines 137-139 in the manuscript:

"**An atmospheric layer bounded by two tangent heights of SCIAMACHY limb observations is the basic spatial resolving unit of this study. We create an additional layer above the outermost tangent height by assuming a layer thickness equal to the average difference between adjacent tangent heights.**"

Lines 9 – 11: please add altitude range where temperatures can be retrieved (~40 – 95 km for nominal mode, 65 – 105 km for the MLT mode?).

We added the valid altitude range of 40-100 km for the nominal and 60-105 km for the MLT mode, combining both bands.

Line 62: Yang et al, SPECTROSCOPY AND SPECTRAL ANALYSIS, 2021 also used the O2 airglow to derive temperatures

It is added to the references.

Line 84: in the nominal limb mode, Sciamachy scanned up to 93 km in 2010. It was slightly higher at the beginning of the Sciamachys operations, but unfortunately this was changed to 93 km already in late 2003.

It is updated to "**In nominal limb mode, SCIAMACHY observed the atmosphere from the surface up to 93 km in 2010.**"

Line 166: where is the number 1.4387760 cm K coming from?

It is the second radiation constant in Planck's law, and its value equals to the produce of Planck's constant, speed of light, and the inverse of the Boltzmann constant (e.g., see https://hitran.org/docs/definitions-and-units/). This sentence is updated to "**where c_0 is a scaling constant, and c_2 is the second radiation constant in Planck's law with a value of 1.4387769 cm K.**"

Line 192: … will also be N. Actually if you formulate it like that, the number should be N-1. It is N in your retrieval because you add an upper bound layer at the top. Can you clarify this?

In fact we create a layer for each tangent height, so the number of layers equals the number of tangent heights and equals N. We clarify this by modifying lines 137-139 to

"**An atmospheric layer bounded by two tangent heights of SCIAMACHY limb observations is the basic spatial resolving unit of this study. We create an additional layer above the outermost tangent height by assuming a layer thickness equal to the average difference between adjacent tangent heights.**"

Line 300-302: "only limited limb views with deeper tangent heights could observe those layers" I am not quite sure I understand this statement. Does it mean only some of the nominal limb scans (which all go down to the surface) provide a good signal-to-noise ratio in these altitudes? This is how I understood this sentence, however I don't understand how it applies to the discussion of a single limb profile as given here. Please clarify.

"Deeper" here is relative in the collection of tangent heights above ~25 km and does not involve limb scans down to the surface. All tangent views can see the top layer, whereas only the bottom tangent view can see the bottom layer. To clarify, item (1) is rewritten:

"**The retrieved [$O_2(a^1\Delta_g)$] becomes increasingly uncertain down to the stratosphere because (1) the lower the layer is, the smaller the number of tangent views can detect them…**"

Line 303: as supported by comparison to results of the MLT mode retrieval

Revised as suggested.

Lines 311-312: This is by design … due to the self-absorption

Revised as suggested.

Lines 317-318: a) the lowest tangent altitude of the MLT mode is around 51 km; b) why is the upper limit set to below 120 km?

a) This particular sounding only goes to 57 km. This relates to our choice of using 8 native across-track soundings, instead of combining adjacent soundings into 4 across-track soundings.

b) This is because above 120 km effectively no $O_2(a^1\Delta_g)$ airglow is observed, and a cut-off saves computing time. Similarly, the A band is cut off at 130 km. This sentence is revised to

"**The lowest tangent height of this particular MLT sounding is at 57 km, and the upper limit is set to below 120 km, above which the airglow is negligible.**"

Line 324: erase the would. They do.

Revised as suggested.

Line 371: the maximum abundance

Revised as suggested.