Dear Professor Lübken,

please see below our replies to the referee report. We marked in bold the modified manuscript text.

With warmest regards,
Peter Panka

*The revised manuscript includes important new analysis relevant to the topic. However, I'm puzzled about the authors' reply to Reviewer#3 regarding the focus of the paper. They say that they "removed any discussions of which model and how well it fits SABER measurements" , dealing now "only with the comparison of various model calculations using WACCM self-consistent inputs of all atmospheric parameters". But then the new title reads "...:Comparison of the CO2(v3) and OH(v) emission models with space and ground based observations". Further, they conclude, e.g., that the new indirect channel provides significant 4.3 um emission enhancement strong enough to *FIT* SABER radiances. This seems not consistent to me with their statement in their reply.*

We thank the referee for pointing out the remaining inconsistencies. We changed the title to make it clear that our main goal is the comparison of models. The manuscript text was also changed correspondingly.

If the focus of the paper was limited to an intercomparison of various models then, in consequence, any comparison to observations should be excluded and the manuscript title and some parts of the manuscript need to be rewritten. In this case, it would be most appropriate to use the (OH-N2 3Q) & (OH-O2 1Q) & R7 model as reference in the comparisons of simulated 4.3 um emissions because its capacity to reproduce SABER observations (4.3, 2.0, and 1.6 um) in a self-consistent manner has been demonstrated in a previous study.

As we have pointed out previously, in the revised paper we focused primarily on the comparison of various models. However, the final goal of any modeling is the interpretation of measurements, we still believe that there is a compelling reason to show comparisons not only among models, but also with various measurements. These comparisons enable us to determine how well each model reproduces basic observed features and thus shed light to the physical processes we are attempting to explain. For instance, we are puzzle by the argument against us showing the comparison between a published model such as the one presented in Lopez-Puertas et al, 2004 and the measurements the model attempts to reproduce. These are published results and for complete scientific sake must be addressed when new results show potential disagreements.

Additionally, we disagree with the reviewer in that we should remove the comparison of models with ground and space observations of OH(v) distributions (Figure 4). This is the compelling point of our study. Without such comparison, the modeled results would represent just an unconstrained theoretical exercise.

However, If they aim at evaluating the various models with observations (as suggested by the present title) it is absolutely necessary to also demonstrate that model inputs and intermediate results are consistent with the measurements. In particular, it then needs to be demonstrated that modeled ABSOLUTE OH densities are consistent with SABER 1.6 and 2.0 um measurements. This was already* *requested by both reviewers in the previous iteration. However, the authors now show RELATIVE populations or VER ratios in comparison with observations. This is a very interesting additional diagnostics, however, it does not allow to judge if one or the other model provides a better description of energy transfer from OH(v>4) to CO2 because the amount of available excited OH is not constrained. In other words, good agreement of modeled and 4.3 um emissions could also be achieved by a combination of a wrong energy transfer mechanism in combination with wrong OH* densities.*

In this study we do not fit any measurements, however, show (on fixed inputs) how various models reproduce specifics of **various measurements**. The total OH density in our calculations is fixed by the WACCM inputs. We show here how various relaxation mechanisms **impact the distribution of the OH(v) vibrational level populations and CO2 pumping**. Regarding the referee's statement that our model "*does not allow to judge if one or the other model provides a better description of energy transfer from OH(v>4) to CO2 because the amount of available excited OH is not constrained":* in our study, where total OH is constrained with WACCM inputs, we show in detail that the new mechanism provides the same enhancement of the 4.3 um emission as the best (Lopez-Puertas et al 2004) model does, even if efficient multi-quantum quenching by O2 (reaction R6) is applied (see also discussion below). Thus we still believe our current comparisons are justified and valid.

This is a major concern, however, it can be easily addressed, either by some minor changes to the text (first case) or by inclusion of an additional row in Fig 3 showing observed and modeled 1.6+2.0 um ABSOLUTE VERs (second case). If in the latter case disagreement between modeled and observed VERs is obtained, a scaling of the WACCM [H][O3] and consequently [O] (because of chemical equilibrium) could be applied in order to achieve agreement.*

In addition to our comments in the previous point, our work represents an 'in between' state of the two scenarios described by the reviewer. We do not fit any signals with WACCM inputs, therefore, we do not see the benefits to show absolute data in the framework of this study. Absolute OH VERs do not explain what is the basic difference between models, which is our main goal of the study. However the VER ratios and relative population distributions displayed in Figure 4 are the appropriate quantities to compare.

Minor comments:

p5 l19-21: The use of pT from WACCM (instead from SABER as in the previous version) could introduce additional uncertainties in the comparison of modeled and observed 4.3 um emissions, which should be discussed. For instance, it is likely that the pronounced differences of both models below 75 km compared to the measurements in Fig 3 (upper panel d) are related to a temperature mismatch of WACCM. Note that, despite of being self-consistent, WACCM output does not necessarily reflect the actual atmospheric conditions location and time because the model is free-running in the mesosphere even in the SD mode. Real and modeled meteorology can thus be quite different.

In the previous version, the reviewers criticized the use of SABER as inputs to our model. Given this criticism and the fact that we do not aim, now and then, to fit any signals, we decided to use self-consistent WACCM inputs and show measurements to only illustrate the performance of various models. We absolutely agree with this referee that compared to (inconsistent) SABER inputs, WACCM inputs "*do not necessarily reflect the actual atmospheric conditions of a given measurement*". However, they appear to be good enough to model specifics of each type of measurements (be these SABER observations for various latitudes/seasons or ground and space observations of OH(v) emissions) and show how well various models account for these specifics.

It is also not clear from the provided references what kind of WACCM simulation has been used. The Marsh et al. reference points to a free-running simulation while the Solomon et al reference points to a SD simulation, however, only for the year 2011.

For all simulations, we use the WACCM model described in Marsh et al. [2013]. We have removed the citation for Solomon et al. [2015], as this is not relevant to our study.

p7 l9-11: There seems to be a misunderstanding of what is stated in Lopez-Puertas et al. regarding the treatment of R6 as multi-or single quantum process. The latter authors adjusted a reference OH_ref(v,z) profile that has been modeled with different implementations of R6 to the observed 1.6 and 2.0 um radiances and used then the adjusted OH(v,z) profile to simulate 4.3 um emissions. Their statement refers to the fact that changes in the vibrational distribution within v=1-5 (used to adjust the 1.6 um measurements) and that within v=6-9 (used to adjust the 2.0 um measurements), caused by the different implementations of R6, had little impact on the simulated 4.3 emission. It is evident that there would have been a significant impact if the OH(v=1-5) and OH(v=6-9) were not adjusted to the SABER 1.6 and 2.0 um radiances. Therefore, the sentence on l9-11 should be removed in order to avoid confusion.

We thank the referee for this detailed explanation of how Lopez-Puertas et al, 2004 dealt with single- and multi-quantum implementation of reaction R6. We now understand that in order to compensate significant OH(v) decay due to multi-quantum quenching by collisions with O2 and keep the transfer of energy to CO2 unchanged *"OH(v=1-5) and OH(v=6-9) were adjusted to the SABER 1.6 and 2.0 um radiances"*, obviously with higher OH(v). This actually may mean that total OH density (or VMR) was increased.

We note here also that neither initial nor final *ABSOLUTE* OH(v) (or total OH densities) obtained in SABER OH signal fittings were shown and discussed by Lopez-Puertas et al, 2004. It, therefore, is not possible to judge how realistic they were, what O3 and H were used, etc.

On the other hand, in this study we demonstrate, based on fixed self-consistent WACCM inputs, that the Sharma mechanism provides efficient energy transfer to CO2, which as oppose to Lopez-Puertas et al, (2004), does not require additional OH adjustment to compensate the multi-quantum O2 quenching. We provided additional text to make this point clearer.