Referee Report on “Comparison of Antarctic polar stratospheric clouds observations by ground- and satellite based lidars and relevance for Chemistry Climate Models” by Snels et al. – Second revision

General comments

The clarity of the paper has significantly improved, both with the inclusion of a new figure (Figure 1) and with the revision of the text. I am also happy that the revision of Figure 2 (now, Figure 3) focussing on the single year 2006 instead of a cumulative statistics over the period 2006-2010, possibly with the correction of some bug in the algorithm, gives a much more convincing agreement between ground-based and CALIOP observations. Together with both tables giving a detailed overview of the statistics on the various PSC classes, this gives much more confidence in the statistical analysis, and in the relevance of the paper.

In the second part of the paper (comparisons between models and CALIPSO) however, the authors come back to the comparison of results covering the 5 years (2006-2010). This might be unfortunate in the sense that they loose again what they gained by restricting themselves to the year 2006 in the comparison between CALIPSO and ground-based lidar measurements above McMurdo. So, we don’t know what is the effect of interannual variability on Figure 5-12, although the differences found while considering either only 2006, or 2006-2010, are significant (See Table 1), as is, consequently, the importance of interannual variability.

Further, the authors explained very well in the reply to referee that the aim of the second part of the paper is not to attribute (bad or good) scores to the different models considered here, but to propose useful diagnostic tools for the comparison between models and observations, applied here on five models. Overall , the way they revised the text reflects well this aim. However, the conclusion follows clearly the other way, electing WACCM-CCMI as “best choice”. I think the authors could go beyond their current conclusion to show which kind of diagnostic they can provide on their reference model.

As a conclusion, although this paper is very interesting and, in my opinion, absolutely worth to be published as is, I think it could be improved and come to even more convincing conclusions by following the suggestions give here above.

The references to the manuscript (line and page numbers) refer to the pdf version acp-2018-589-manuscript-version5.pdf and are mentioned as indicated in this file, although the numbering suffers, from p. 12, of obvious imperfections (e.g., the first line on p. 12 is given the line number 5).

Thanks to the authors for the extended discussions provided in the reply to referee.

Detailed comments

Abstract

- L.12-13 and L. 16-20, p.1: If the aim of this work explained in the reply to the general comment ("The main focus is to define diagnostics that permit to compare observations with the "model world" in a consistent way. (…)") is now very clearly explained in Section 2.6 (L. 12-14, p. 7 and L. 10-11, p.8), the sentence in the abstract still mentions that the aim is to assess the performances of the different CCMs in simulating PSC occurrences and PSC distribution over Antarctica. It might be useful to add the clarification they provided in their response in some way in this abstract.

2. Comparison of PSC observations by ground-based and satellite based lidars

- The revision of Section 2.4 and the use of Figure 1 makes this section much more clear.
- L. 24-27, p.7: These two sentences are just a repetition of what is written in L. 21-23, p.7, and can thus be removed.
- L. 25-26, p.7: "in a spatial box" is not very informative; the authors could usefully repeat the dimension (I guess 7° longitude x 2° latitude). Also, "centered on" (if this is what the authors mean) could be more precise that "around".
- L. 7, p.9: 18 km might be an estimate that better reconciles the ground-based and CALIOP cases than 20 km for the lower limit of enhanced NAT mixture occurrences.
- Caption Figure 3: I guess the authors mean "the two columns".
- Caption Figure 4: "a specific temperature in arbitrary units" is surprizing. Rearranging the sentence or adding suitable punctuation might be useful to remove the confusion.
- L. 8-10, p.13: I think this sentence ("This is probably due (…) as can be seen also in figure 3") should be removed or revised. The new Figure 3 shows an ice fraction of about 20% in July with a remarkable agreement between the CALIOP and ground-based cases and cannot be used to justify the differences between both cases in Figure 4. In fact, Figure 4 shows again cumulated statistics over the years 2006-2010. It would be much better, for a better coherence and an easier and more correct comparison between the different diagnostic tools, to focus on the year 2006, also in Figure 4. This would imply, of course, that the same choice is made for the following of the paper, including the plot of Figures 5-12.

3. Comparison of CALIOP PSC observations in the Southern Hemisphere with CCM simulations

- L. 5 (as indicated in pdf version), p.14: "CAM3.5 and WACCM allow for saturation of up to 10 times saturation": I don't understand what the authors mean by this sentence.

- L. 10-11, p.16: “Recent studies of model simulations” is particularly vague and should be precised. Do the authors refer to “an overview of PSC simulations [by WACCM])”, or to the new version WACCM/CARMA not considered here ?
- L. 8-9 (two last lines), p. 18: Do the authors take into account CALIPSO averaging kernels for this exercise ?
- L. 10-11 (as indicated in pdf version), p. 21; Figures 5-10, Table 5: There is an inconsistency between the altitude range mentioned in the figure captions on the one hand, and the Table 5 caption on the other hand.
- Caption Table 5: I don’t see why the authors write that “fractions below 1% are not reported in the table”. E.g. LMDZrepro estimate for September is 0.1%.
- Table 5, Table 2: Isn’t it strange that the estimates provided by Table 2 for July and August 2006 are so different from the ones provided by Table 5 for July and August 2006-2010 ? In particular, neither the values nor the trend (increasing/decreasing) agree in the case of ice. Further, is there any issue in considering, also in this case only the year 2006 ? This would allow a much more detailed and interesting comparison with all results of Section 2.
- Figure 11, p. 23: Is the period considered here also 2006-2010 ? Please specify in the caption.

4. Conclusions

- Second paragraph: In their previous reply to referee, the authors insist on the fact that it is not the scope of the paper to give general scores to the various models, but rather to present useful diagnostic tools for the comparison between model results and a set of observations. However, scoring the 5 test-models against CALIPSO is basically what is done here, with as final conclusion that WACCM-CCMI is the “best model”. In order to follow their objective explained in the reply to referee, it might be useful, in this conclusion, to go a step forward with respect to the observation of over/underestimation of NAT frequency, anticipated or delayed onset of PSC formation etc., to try drawing some (preliminary) conclusions about the performances of the model (microphysical scheme, efficiency of the dynamics, ability to describe dynamical effects such a s mountain effects, etc.) as what is done in Section 3.4.

Technical corrections

- L. 11, p.2: “A variety of (…) has been proposed”.
- L. 15, p.12: Duplicate “to”.
- L. 27, p.12: Missing parenthesis.
- L. 8, p.13: “larger than” instead of “more larger with respect to”.
- Table 4: “ice” instead of “iice”.
- Caption Table 5: Is a part missing ? There is no final punctuation.