Letter to the Editor.

Dear Editor, all the questions and suggestions by the two referees have been answered and the manuscript has been changed where necessary.

The most important differences are summarized below.

1) The detection and selection scheme based on the v2 algorithm have been illustrated in a figure (figure 1) and the thresholds for detection and selection have been described in detail, adding also a paragraph where the selection based on the optical parameters R and beta_perp has been explicitly formulated
2) The critical comments on the correctness of the statistical comparison has been also addressed in the manuscript, both by indicating the difficulties to obtain “real” coincidences” and by taking a particularly relevant period (July and August 2006) with a good temporal coverage by the two datasets. Furthermore, it has been stated that the purpose of this paper is NOT to perform a validation of the CALIOP data, but an exercise for the statistical comparison of ground-based and satellite borne lidars.
3) Some bugs were found in the software used for the calculation of the vertical distribution of the PSCs (figure 3), thus solving some of the contradictions.
4) The discussion of the diagnostics of CCMs with respect to CALIOP has been organized in a more clear way, and all five models have been discussed in more details, in order to understand their merits and flaws.
5) A statement has been added referring to the latest models using detailed microphysics instead of parametrizations for the formation of PSCs. We feel however, that these are beyond the scope of this paper.
6) The conclusions have been reformulated .

Answers to referee 1

Interactive comment on “Comparison of Antarctic polar stratospheric cloud observations by ground-based and spaceborne lidars and relevance for Chemistry Climate Models” by Marcel Snels et al.

Anonymous Referee #1

Received and published: 18 September 2018

A Review of “Comparison of Antarctic polar stratospheric cloud observations by ground-based and spaceborne lidars and relevance for Chemistry Climate Models” by M. Snels et al.

<General Comments>

This paper describes the comparison between PSC measurements at Antarctic Mc-Murdo Station from ground based lidar and CALIOP satellite measurements. Further-more, the paper tries to extend the comparison of PSC statistics from CALIOP with several CCM model results from CCMVal-2 and CCMI. Although scientific value of this study might be significant, the method of comparison especially with CCM models is not well organized to derive scientifically useful conclusions, as is pointed out below. Also, there are too many typos and careless mistakes in the draft. A major revision is required before this paper will be published in ACP. I recommend that authors should check the draft carefully, including the native check, before submitting the revised draft.

(M1) In Section 3.2, the authors try to compare the PSC statistics from 5 years (2006- 2010) measurements by CALIOP, with the result of 4 CCM models from CCMVal-2, and one CCM model from CCMI. However, the model run type they chose for CCMVal-2 models are REF-B2, which are targeted to be used for future predictions until 2100. The major problem for this comparison is that the result of REF-B2 run contains both inaccuracy in modeled temperatures and imperfectess in PSC schemes which are different in each model. The combination of inaccuracies both in modeled temperature and PSC schemes makes it extremely difficult to understand the nature of PSC in each model. Rather than comparison with CCMVal-2 REF-B2 runs, it is strongly preferred to compare with CCMI outputs with refC1SD runs (which is available from http://badc.nerc.ac.uk/browse/badc/wcrp-ccmi/data/CCMI-1/output), which use nudging with more realistic temperature and wind field, just to test the PSC scheme in each model. Even if the authors stick to the comparison with CCMVal-2 model results, they should at least use the REF-B1 model run results, which are targeted to reproduce the past. In this case, the comparison with CALIOP could be made only for 2006, because REF-B1 run was made only for 1960-2006. Since CCMI refC1SD runs cover until 2010, I strongly recommend making comparisons with CCMI model outputs with CALIPSO Measurements.

ANSWER: (M1) The indication of the REF-B2 run was a typing error, we apologize for that. In this manuscript we evaluate the REF-B1 simulations available for the period 1960–2006. As the reviewer highlights, those simulations were chosen because they have been constructed to include the interannual variabilities of the 11 year solar cycle, the QBO, Sea Surface Temperature (SST), volcanic effects, greenhouse gas (GHG) concentrations, and ozone-depleting substance (ODS) concentrations (Morgenstern et al., 2010). The SST and sea ice evolutions are prescribed using the HadISST1 (Rayner et al., 2003). The variations of the GHGs and the ODSs follow the IPCC SRES A1B scenario and WMO-adjusted scenario A1. To our opinion these free running simulations are the most suitable to be compared with the statistics from available observations.

(M2) In Section 3.1, the authors mention about more sophisticated.0/SD/CARMA model and EMAC/MSBM model, which use more realistic parameterizations for PSCs. It would gain the value of this paper significantly if they could include the comparison of CALIOP PSC statistics with the result of these models.

ANSWER: The more sophisticated models are mentioned in the manuscript because those are, to our knowledge, the most significant advancements in the field of PSC representation in Global Climate Models used for the ozone and climate change studies. The CARMA model is an interactive aerosol and radiation model fully coupled to the WACCM, able to fully simulate advection, diffusion, sedimentation, deposition, coagulation, nucleation and condensational growth of atmospheric aerosols online with the temperature, dynamics and radiation structure simulated by the GCM. This approach is completely different from the parametrizations available in the simulations we are analysing here. A full evaluation of the WACCM/CARMA models in Specified Dynamics runs w.r.t. CALYPSO data are available in literature (in the Zhu et al cited works) and are outside the scope of this intercomparison (where we work with free running simulations). It would be certainly interesting to apply the diagnostics proposed within our analysis to a free-running set of simulations performed with models including interactive aerosols. This could be the objective of a future study, when a set of simulations from new generation models might be available.

M3) In each model, denitrification and dehydration are included as is shown in Table 3. This would change the vertical distribution of HNO_3 and H_2O, which would affect the threshold temperature of NAT and ice PSCs, i.e., T_NAT and T_ice. How-ever, this effect is never mentioned or discussed in the manuscript. Moreover, in many places in the text (especially in Sections 2.6 and 3.4), it is not clearly stated which temperature (MERRA-2, NCEP, or derived T in CCM) is used, and how T_NAT and T_ice are calculated (using HNO_3 and H_2O value from MLS data, modeled value in CCM, or fixed values like 6 ppbv HNO_3 and 4.5 ppmv H_2O). The effect of denitrifi-cation/dehydration in modelled PSC should be discussed in the manuscript.

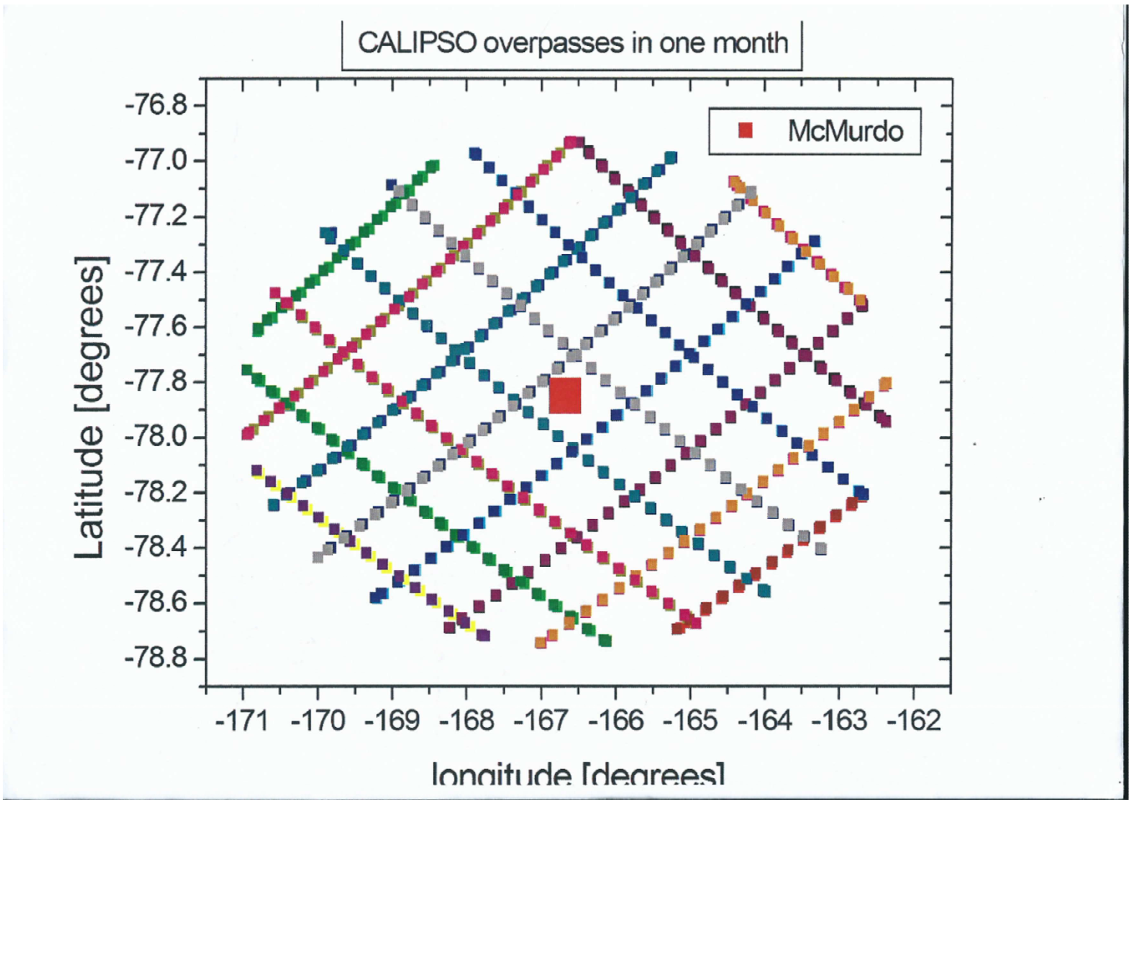
ANSWER: First of all, we use MLS values for HNO3 and H2O concentrations, to calculate the formation temperature of NAT and ice. The temperatures used in this work are taken from MERRA-2. The temperatures used in the CCM models are generated by the models themselves, Tnat AND Tice have been calculated from the HNO3 and H2O taken from GOZCARDS

(M4) For a PSC classes comparison described in Table 1, although the percentage of each PSC class is similar, this does not prove that each one to one PSC is simultaneously observed both by ground-based lidar and by CALIOP. I would recommend authors to add the statistics showing one to one correspondence of comparison of PSC classes observed by tables like the attached tables. Table A shows the statistics when CALIOP measured specific class of PSC, what PSC was observed by Mc-Murdo ground-based lidar, or no PSC was observed. Table B shows the statistics when ground-based lidar measured specific class of PSC, what PSC was observed by CALIOP, or no PSC was observed.

ANSWER: It is not the goal of the article to make a point-to-point comparison for validation purposes. The goal is to verify if the ground-based measurement are representative for a larger area, typically contained in a 7x2 degrees box around McMurdo.

Apart from that a point-to-point analysis presents the following difficulties:

1) None of the overpasses of CALIPSO are sampling the same air mass as the ground based lidar. To illustrate this I show a plot of all overpasses within the 7x2 degrees box, which corresponds roughly to a distance of 100 km from
.

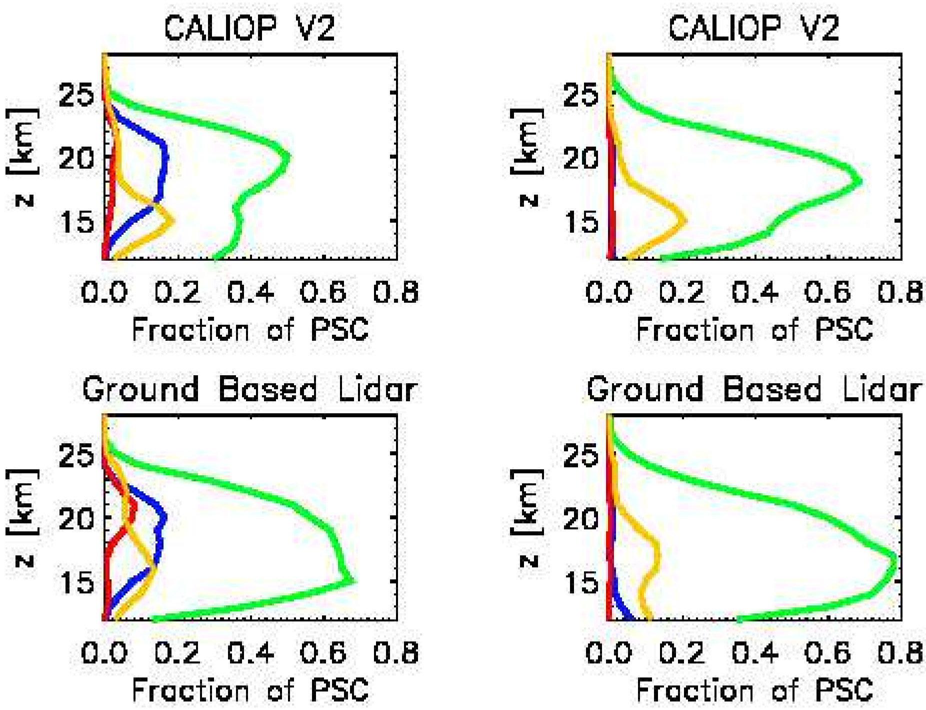


McMurdo. While CALIOP provides a resolution of 5 km (when integration is required due to low signal-to-noise ratio up to 135 km !) the air mass sampled by the ground-based lidar extends to at most 100 m. (30 km * 3 mrad field of view of the telescope).

Another important difference of the two lidars is that a CALIOP overpass occurs in about 30 seconds, while the ground-based data are integrated over 30 minutes.

This implies that the ground-based measurement integrates air masses moving with a wind speed varying from 0 to 50 m/s, depending also on the altitude (the wind speed might be very different at 15, 20 and 25 km), rendering a comparison with an instantaneous profile of CALIOP very questionable.

However, the statistical analysis is only meaningful if the sampling of the two lidars covers the same period of time and if this period of time has a dense coverage. In order to achieve this we concentrate on 2006, having a large number of observations by both lidars with a good coverage (see figure 1 of the manuscript). We then analyse the months July and August and report the statistics in terms of occurrences of PSC classes and dependence on altitude.



Differences and agreement have been discussed in the revised manuscript.

(M5) In Section 3.4, they discuss about the cold pole bias in most CCMVal-2 CCM models. However, when I read the SPARC report No.5 Chapter 4 “Section 4.3.5 Polar stratospheric cloud threshold temperatures” in page 128, there is an explanation that CCM models have warm bias and A_NAT and A_ice show low value compared with ERA-40 temperature. This description totally contradicts with the discussion described in Section 3.4. Please explain why such contradiction occurs.

ANSWER: Looking at figure 4.1 of the Sparc report (page 112) it is evident that all models have a cold temperature bias except for the two UMUKCA models. This is explicitly stated on page 113. Figure 4.15 in "Section 4.3.5 Polar stratospheric cloud threshold temperatures" in page 128, shows that the same two models strongly underestimate the mean PSC Area's which is of course in agreement with the warm bias of these models discussed before. So there is no contradiction.

All the corrections suggested by the referee below have been made.

- ✓ (S1) The numbers in author list are not ordered correctly, i.e., 1, 5, 2, 3, 4. It should be
- ✓ 1, 2, 3, 4, 5.
- ✓ (S2) P1, L3: The abbreviation of CALIOP should be shown also in the abstract.
- ✓ (S3) P1, L9: The meaning of "... and a selection simulations obtained ..." is unclear.
- ✓ (S4) P1, L4: In Pitts et al. (2018, ACP), they use "v2" instead of "V2". Please check if
- ✓ V2 should be changed to v2 throughout the manuscript or not.
- ✓ (S5) P1, L18: The abbreviation of WACCM-CCMI should be shown.
- ✓ (S6) P2, L7: The abbreviation of CALIOP should be shown here, not at P2, L20.
- ✓ (S7) P2, L18: Chemistry Climate Models –> Chemistry Climate Models (CCMs)
- ✓ (S8) P2, L20: clouds and aerosol –> clouds and aerosols
- ✓ (S9) P2, L26: Chemistry Climate Models –> CCMs
- ✓ (S10) P2, L29: The SPARC Report No5 (2010) cannot be found in the reference list.
- ✓ (S11) P2, L30: Chemistry Climate Models –> CCMs
- ✓ (S12) P3, L1: Chemistry Climate Models (CCM) –> CCMs
- ✓ (S13) P3, L14: CALIOP (Cloud Aerosol Lidar with Orthogonal Polarization) –> CALIOP
- ✓ (S14) P3, L14: Details on CALIOP –> Details of CALIOP
- ✓ (S15) P4, L16: Reference (Cairo et al., 1999) should appear at the end of Line 18.
- ✓ (S16) P5, L14: CALIPSO V2.0 data –> CALIPSO v2 data

- ✓ (S17) P5, L15: V2.0 –> v2
- ✓ (S18) P5, L17: V1.0 and V2.0 –> v1 and v2

Answers to Referee 2

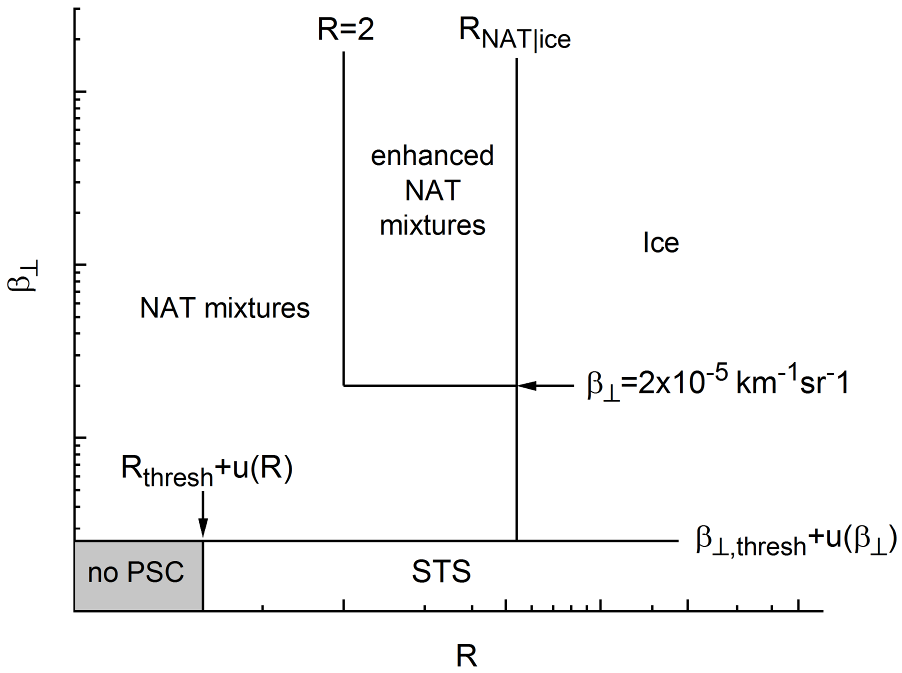
First of all, we want to remark that the referee is referring to a different version of the paper with respect to the one posted on the web-site; acp-2018-589.pdf, probably to the version submitted on 12/06/2018, prior to publication on the website. We thank the referee for his very constructive review, which has surely improved the paper.

General comments

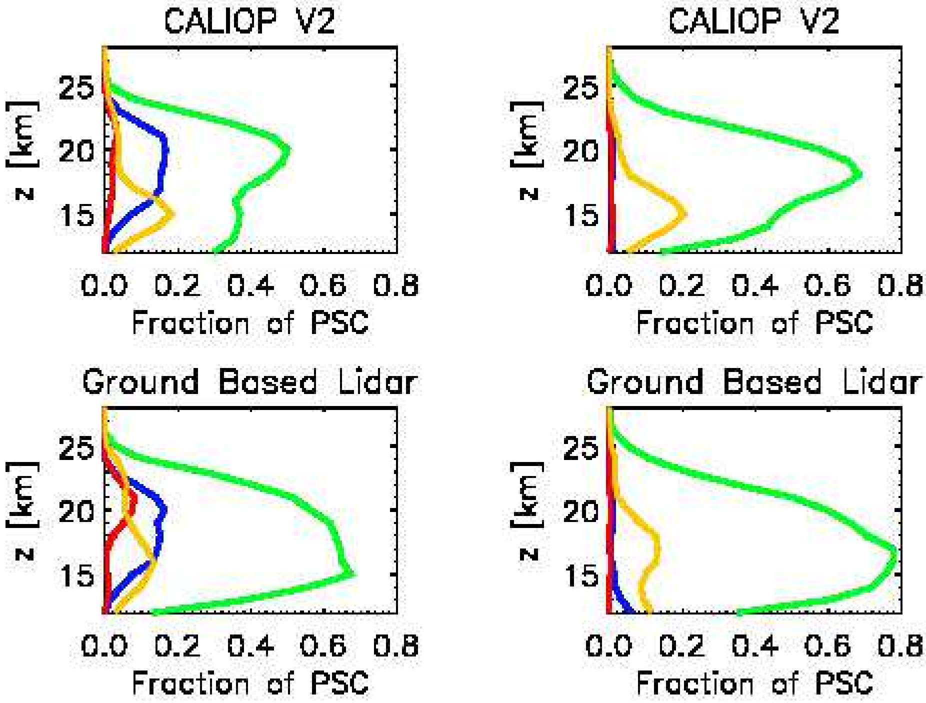
I am not convinced that the way the authors process the CALIPSO and ground-based lidar data is always rigourous and adequate, and this might be a source of many biases and difficulties.

Further, the way to evaluate the agreement between the CALIOP and ground-based datasets, but also the agreement between the different models and CALIOP, look subjective in some cases (e.g. comparison CALIOP-ground-based lidar based on Figure 1, distinction between "rather good agreement above 15 km" and "biased below 15 km" on Figure 2, …..

ANSWER: To answer the referee's lack of confidence in the correctness of the lidar data processing and in order to convince him of the correct treatment of the data, we first state that the CALIOP data have been used as provided by the PI's, using the v2 version of the classified PSCs. The detection and the classification of the ground-based data has been explained in more detail and a new figure has been added to illustrate how the detection and classification algorithm works.



While the value of the confidence indexes provides the confidence in the classification, its value is not used in the classification algorithm, and it provides only a threshold value between two classes. Therefore we've eliminated the confidence indexes from the manuscript and discuss the classification algorithm in terms of threshold values (see figure above). These threshold values have been determined in some cases differently for the two lidars, due to the different nature of the data they produce. This has been discussed in the revised manuscript. For instance, the threshold values for R and βperp correspond with background aerosols, observed in absence of PSCs. These can be easily determined from the CALIOP data, producing daily values, by considering PSC area's on the southern hemisphere at temperatures above 200 K. For the ground-based lidar it is not possible to obtain daily values, and an average has been made of PSC free observations in early June and October.



Several bugs have been found in the normalization of the data, thus producing wrong values for the fraction of the PSC classes. The figure shown above shows the new values for 2006. As a consequence the discussion has been adapted and the distinction between below and above 15 km has been eliminated.

A part of the revised manuscript:

The figure shows that PSCs are observed up to 25 km in July and August. Above 25 km the number of PSC observations is negligible, both for ground-based and CALIOP observations. NAT mixtures are the dominating species in July and August, with a slightly different altitude distribution in July; ground-based occurrences of NAT mixtures are more frequent below 18 km with respect to CALIOP data. The occurrences of ice clouds in July are very similar, while in August some low ice clouds appear in the ground-based data, but are absent in the CALIOP observations. Enhanced NAT mixtures occur mainly in July, and are observed between 17 and 25 km, more abundant in the ground-based observations. The vertical distribution of STS shows a good agreement in July and August.

….general rejection of "outlier" LMDZrepro model although this model scores not so bad following some specific criteria). Concerning the comparison between CCM's and CALIPSO, I find striking that the "best model" giving the best agreement with CALIPSO is highly depending on the methodology used: Based on total PSC frequencies (Table 2), LMDZrepro and WACCM-ccmi are performing the best; based on the SAD histogram, LMDZrepro shows the best agreement based on the range of Log10(SAD); WACCM and CAM3.5 give the closest evolution of the NAT and ice fraction as a function of T-TNAT. Hence, CCSRNIES is the only one of the 5 models considered here that cannot pretend to the status of "best model" following any diagnostic method, although the authors reject overall another model, namely LMDZrepro, and outlier. Overall, I don't see any clear conclusion from this work, and my general feeling is mainly that the way the CALIPSO data ground-based lidar data are processed might present biases or be inadequate, and that the implementation of the different diagnostic methods should be improved.

ANSWER: The reviewer is correct, the previous version of the text was giving the impression of a general scoring of the models, with a final "negative" score for the LMDzRepro or the idea to derive a "best model". This is not the scope of the manuscript. The main focus here is to define diagnostics that permits to compare observations with the "model world" in a consistent way. In order to disentangle, when possible, biases deriving from specific parameterizations that could be attenuated in principle with future improvement, and biases related to the global biases of the model and more difficult to target. For example, when the error is strongly associated to the cold pole bias in stratospheric temperature and therefore attributed to model dynamics, it requires a more structural intervention on the model definition than when bias is associated to the assumptions in the specific parametrization made on the number of particles per cm^3. A future study might imply the development of specific metrics, derived from the diagnostics proposed here, that could allow to define scores and evaluate models. However, as the reviewer correctly remarks, this would not be a straightforward way of proceeding and it is outside of the scope of the present work. We have adjusted the text in relevant sections to illustrate this.

Detailed

Abstract
L. 3-5, p.1: This sentence is particularly difficult to read. Please reword in a more fluent way.
ANSWER: The sentence has been divided in two pieces in order to facilitate the reader.

L. 1 and 6, p.1: The authors repeat partly the same idea. The text could be written more efficiently, or in another way to put the emphasis on the main focus of the sentence.
The sentence has been re-edited. Below follows the new text:
Abstract. A comparison of polar stratospheric clouds (PSCs) occurrence from 2006 to 2010 is presented, as observed from the ground-based station McMurdo (Antarctica) and by the satellite-borne CALIOP lidar (Cloud-Aerosol Lidar with Orthogonal Polarization) measuring over McMurdo. McMurdo (Antarctica) is a primary station in the NDACC (Network for Detection of Atmospheric Climate Change). The ground-based observations have been classified with an algorithm derived from the recent v2 detection and classification scheme, used to classify PSCs observed by CALIOP.
A statistical approach has been used to compare ground-based and satellite based observations, since point-to-point comparison is often troublesome due to the intrinsic differences in the observation geometries and the imperfect overlap of the observed areas.

1. Introduction
L. 7-8, p.2: "Many different schemes…": Do the authors mean that the different schemes use different thresholds for detection and classification ?
ANSWER: The text has been modified; indeed the different schemes often use different thresholds.

Many different schemes using thresholds for detection and classification have been proposed, rendering a comparison difficult.

L.11-12, p.2: “Ground-based lidar observatories… from the early nineties to today”: The authors might be only interested by the period from the early nineties until today, or by a specific location (probably McMurdo), but there exist ground-based lidar time series spanning at least 2 decades more ! (See for instance Jäger, J. Geophys.Res., 2005). Hence, they should be more specific.

ANSWER: We refer to lidar observations in Antarctica. Anyway we now have included also the earliest, up to our knowledge, lidar observations in Antarctica, with references, from 1985 on. Of course there exist ground-based lidar observations much earlier, but not in Antarctica. The Jaeger paper deals with observations in Garmisch-Partenkirchen.

The first lidar observations in Antarctica started in 1985 at Syowa Station. Iwasaka and co-workers (Iwasaka, 1985, 1986) used a polarization sensitive lidar to measure backscatter and depolarization to observe PSCs. Later, in 1987/1988 at the Amundsen-Scott South Pole Station, Fiocco and co-workers (Fiocco et al., 1992) used the elastic backscatter signal from a 20 lidar operating at 532 nm to observe PSCs in relation to the temperature. PSCs have also been observed at Davis, from 2001 to 2004 (Innis and Klekociuk, 2006) and at Rothera (Simpson et al., 2005) from 2002 to 2005.
Long-term observations of PSCs have been performed at McMurdo (Adriani et al., 1992, 1995, 2004; Di Liberto et al., 2014), from 1989 until 2010 and at Dumon D’Urville (Santacesaria et al., 2001; David et al., 1998, 2010), from 1990 until now, both with polarization sensitive lidars. Recently the McMurdo lidar has been transferred to Dome C and is operating there 25 from 2014 on (Snels et al., 2018).

L.12-13, P.2:” A clear issue …”: Do the authors mean that the ground-based time series above Antarctica are not representative enough for climatological studies and model evaluation above Antarctica ? This should require a reference.
ANSWER: The Antarctic lidar stations are few and those with a long term record even fewer (McMurdo, Dumont D’Urville and Dome C). This means that model calculations can be compared at a few locations. It doesn’t mean that they are not representative enough for climatological studies.

2. Comparison of PSC observations by ground-based and satellite based lidars
2.1 CALIPSO observations
2.2 Ground_based PSC observations at McMurdo
L.20, p.3: “Klett algorithm”: This requires a reference.
ANSWER: A reference for the Klett algorithm has been added.

L.2-3, p.4: What do the authors mean by “facilitate” ? Is it about reducing the dataset ? Or having a regular time base ? Or something else ?
ANSWER: It means that we would like to compare data on a daily base, since CALIOP produces at most one overpass per day. Thus we proceed as follows: if more than one ground-based profile is available within a 6 hour time window, only the profile with the smallest time difference with respect to the Calipso overpass is considered. However, this situation is rarely verified. We explained better in the text how we obtain a daily profile for the ground-based data.

2.3 PSC detection and classification
L. 24, p.4-l. 8, p.5: The authors are restarting an overview of the literature, citing the same works as in the overview literature in the introduction. This cares for unnecessary repetitions.

The authors should focus on the message needed at this point of the discussion, without repeating what was said before.
ANSWER: The title of this paragraph justifies a reference to the recent review by Achtert and Tesche. in our opinion. The detection scheme used in this work is based on the CALIOP algorithms, so it is obvious that these are mentioned here.

L.1-2, p.5: These lines include 2 almost similar sentences about the same work ! Please remove what is not necessary.
ANSWER: The sentence has been removed
L. 1-6, p.5: The same reference is cited 3 times during the description of this work. Please remove two of them !
ANSWER: The three references have been removed and we now refer only to Pitts2018, for the V2 classification.

2.4 PSC detection and classification criteria for the CALIPSO V2.0 data

L. 10-12, p.5: Here again, the authors repeat what has been written in the introduction (on ll. 8-10, p.2).
ANSWER: The sentence has been removed and the text has been modified.

L.13, p.5: “below” is actually immediately after the sentence. “As follows” might be more appropriate.
ANSWER: “Below” has been substituted with “as follows” as suggested by the referee

L. 14, 16, p.5: The use of “now” brings some confusion: do the authors mean “in Version 2” or “in the present work” ? Using “In Version 2” (if this is what is meant) might clarify this point.
ANSWER: “now” has been substituted with “in Version 2” as suggested by the referee

L.17-19, p.5: These two sentences are difficult to read. Do the authors mean that there are two criteria, and that a PSC occurrence is assumed if at least one of the criteria are fulfilled ? Writing that two threshold for background aerosols, respectively for the perpendicular backscatter and the scattering ratio, are defined as their median value plus one median deviation, might already clarify the text. Using formulas might also make it more clear. It is also not clear for me what is the relationship between the median deviation and the “unc” quantity. I understand from the text that, in both cases, the effective threshold is the median value+median deviation+ uncertainty. Is it what the authors mean ? Again, an expression using an equation may remove any ambiguity.

ANSWER: Yes, “or” means that it is sufficient if one of the two criteria is fulfilled
We rewrote this section and added a figure to better explain the detection and selection criteria.

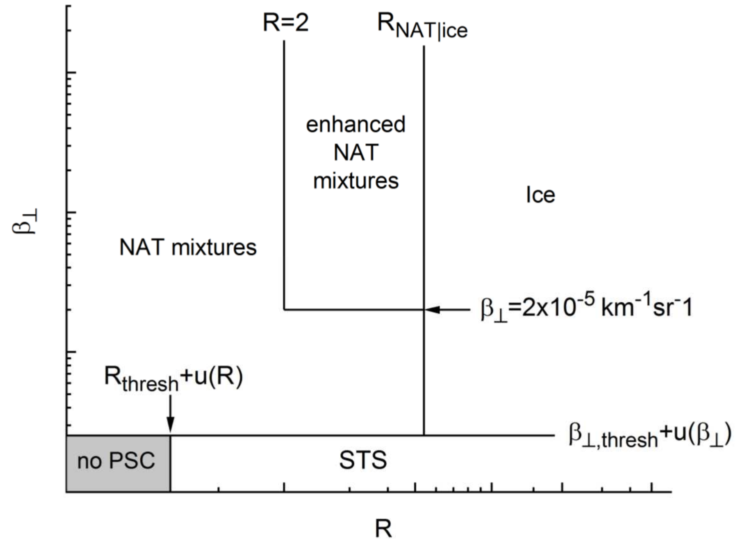


Figure 1. The figure shows the detection and classification criteria of the V2 CALIOP algorithm. The classification as STS, NAT mixtures, enhanced NAT mixtures and ice, requires that threshold conditions for R and/or bperp are satisfied. See the text for details.
The following paragraphs substitute the old ones in the manuscript:
2.4 PSC Detection and classification criteria for the CALIPSO v2 data
The CALIOP v2 PSC detection and composition classification algorithm (Pitts et al., 2018) has been used to create the recently released CALIOP v2 PSC mask database covering the period from June 2006 to October 2017. Here we compare these v2 data with ground-based observations at McMurdo from 2006 to 2010. Major enhancements in the v2 algorithm over earlier versions include daily adjustment of composition boundaries to account for effects of denitrification and dehydration, and estimates of the random uncertainties u(bperp) and u(R) due to shot noise in each data sample, which are used to establish dynamic detection thresholds and composition boundaries. The CALIOP v2 algorithm is represented pictorially in Figure 1 and is described in more detail in the following sections.

2.4.1 PSC detection
PSCs are detected in the CALIOP data as statistical outliers relative to the background stratospheric aerosol population. The v2 background aerosol thresholds bperp_;thresh and Rthresh are calculated as the daily median plus one median deviation of CALIOP data at ambient temperatures above 200 K. PSCs are those data points for which either bperp > bperp;thresh+u(b_perp) or R> Rthresh+u(R). If β_{perp} <= β_{perp_thresh} +u(β_{perp}) and R < Rthresh +u(R), the point is a non-PSC. Noise spikes are eliminated in the CALIOP v2 data by requiring coherence within a running 3-point vertical by 5-point horizontal along-track box.
2.4.2 PSC composition
The PSC composition is determined as follows:

- *If β_{perp} <= β_{perp_thresh} +u(β_{perp}), but R > Rthresh +u(R), the PSC is classified as STS.*

- *A PSC with β_{perp} > β_{perp_thresh} +u(β_{perp}) is assumed to contain non-spherical particles and is classified as NAT (or enhanced NAT) mixture or ice based its value of R. The boundary value separating ice from NAT and enhanced NAT mixtures, RNATjice, is calculated based on the total abundances of HNO3 and H2O vapors as determined* *on a daily basis as a function of altitude and equivalent latitude from nearly coincident cloud-free Aura MLS data*

- *If $\beta_{perp} > \beta_{perp\_thresh} + u(\beta_{perp})$ and R > RNATjice, the PSC is classified as ice.*

- *If 2 < R < RNATjice and $\beta_{perp} > 2\_10^{-5} m^{-1} sr^{-1}$, the PSC is classified an enhanced NAT mixture. All other PSCs with $\beta_{perp} > \beta_{perp_thresh} + u(\beta_{perp})$and R < RNATjice are classified as NAT mixtures.*

The CALIOP v2 data set provides both the grid of classified PSCs according to the v2 algorithm and the associated optical parameters.

L.2, p4; l.17, p.5; l.30, p.6: the time references are confusing. In l.2, p.4, it is indicated that about 1 data point estimated from 30 minute observation is considered every 6h at most; In l.30, P.6, this becomes "1 or 2 measurements occurring per day".
ANSWER:
CALIOP overpasses do not occur every day and at most twice per day. In average we have about 30 CALIOP overpasses per month. Ground-based lidar data are mostly recorded during a CALIOP overpass, but also on days without CALIOP overpasses, usually at the same time that CALIOP overpasses occur and sometimes at different times from the CALIOP overpasses. The latter are not included in this analysis. All other ground-based measurements have been used in the statistical comparison. Generally speaking most of the ground-based profiles have been recorded during a CALIOP overpass, but there might be days with either a ground-based measurement or a CALIOP measurement. So we include all CALIOP measurements falling in a spatial box around McMurdo, and all ground-based data measured in a time frame dictated by CALIOP overpasses, including also the days without overpass.
The text has been adapted accordingly in the revised manuscript.

And in l. 17, p.5, the authors consider a "daily median". On which sampling do they compute the median ? And does the explanation in p.5 mean that a different threshold is considered every day ? An hence that the "background value" is changing every day ? This seems a strange concept of "background value" !

ANSWER: These considerations concern the criteria for the CALIOP data. As said before the CALIOP data were used as supplied by the PIs. The criteria applied by the CALIOP team use a median value of observations above 200 K, i.e in absence of PSCs. The background values are defined as the values of R and □perp in absences of PSCs. Indeed these values can change during the season.

L. 20-31, p.5: Again, all this long description of PSC types would be much more easy to read if they were included in a table and supported by some equations in the text. Also, if the authors find necessary to repeat the change of criteria performed in the CALIPSO dataset, they should at least explain why all these changes are made. Is it a response to the conclusions of the work by (Pitts et al., 2018) explained in ll. 3-6, p.5 ? If yes, the conclusions of (Pitts et al., 2018) might be moved to here.

ANSWER: We inserted a figure showing in a simple way how the detections and classification algorithm uses threshold values . See also the answer given above to the general comments.
The v2 algorithm has also been explained better in the text.

L. 26-29, p.5: I understand that MLS is used to select the PSC type observed by CALIPSO, and that CALIPSO is used to determine the selection criteria. Is there here any problem of snake biting its own tail ? How effective is then this selection ?
ANSWER: Cloud free means that CALIOP did not observe clouds, including PSC clouds of course. All cloud-free MLS data for HNO_3 and H_2O concentrations have been used to determine one of the **selection** (not detection !!!) criteria of Caliop

L. 32, p.5: “the PSC classified grid”: What does it mean ?
ANSWER: This is really confusing, we substituted with “the grid of classified PSCs”

L. 32, p5: Which optical parameters ?

ANSWER: The optical parameters are; backscatter ratio, perpendicular and parallel backscatter coefficient

2.5 PSC detection and classification criteria for the ground-based data
L. 5-9, p.6: Here, the threshold for PSC detection are clearly constant. In which extend are these criteria consistent with the criteria used in ll. 17-19, p.5 ?

ANSWER: The huge number of data acquired by Caliop allow for a very sophisticated statistical elaboration, including the determination of daily means for the threshold. The lidar data are in comparison very few and thus it is very difficult to obtain a reliable daily values. Therefor an average value for the threshold has been adopted, based on previous experiences and also very similar to the average threshold used in the analysis of the Caliop data.

L. 11-13, p.6: I am not sure if this selection occurs in the same way as for the CALIPSO data (See L. 25-26, p.5). Which is the criteria used in that case and how consistent are the selection criteria for the CALIPSO data and the ground-based data ?
ANSWER: The referee probably refers to the phrase “The discrimination between NAT mixtures and enhanced NAT mixtures is made by using the condition R > 2 and bperp > 2_10^{-5} $m^{-1}sr^{-1}$, while the RNAT|ice threshold has been taken from the corresponding CALIOP data, by extrapolating daily values in case of no overpass.
The first part is done in exactly the same way for Caliop and ground-based data. The threshold R(NAT|ice) has been taken from the corresponding CALIOP data, by extrapolating daily values, because it is not always possible to associate a ground-based observation with a coincident Caliop observation.

L.13, p.6: Why do the authors consider here monthly averages while they consider daily averages before ? Isn’t there a lack of coherence in their choices?
ANSWER: This is an error. We extrapolate RNAT|ice from the CALIOP data because Caliop overpasses do not occur on every day within a distance of 100 km from McMurdo. Moreover we are comparing ground based and satellite measurements that are often, but not always, coincident in time.
L. 4-15, p.6: Again, using a table for all the selection criteria could be more readable and make the comparison with equivalent selection criteria applied to CALIPSO more readable.

ANSWER: We inserted a figure for detection and selection criteria

2.6 Comparison of coincident PSC observations at McMurdo from the ground and from CALIPSO during the 5-year observation period

ANSWER: The word coincident is referring to the spatial coincidence, that is considering all measurements of both instruments falling in the box defined as ….We had eliminated the word coincident from the document, in order to avoid confusion, but apparently one escaped our attention, we apologize and substitute coincident by co-located here.

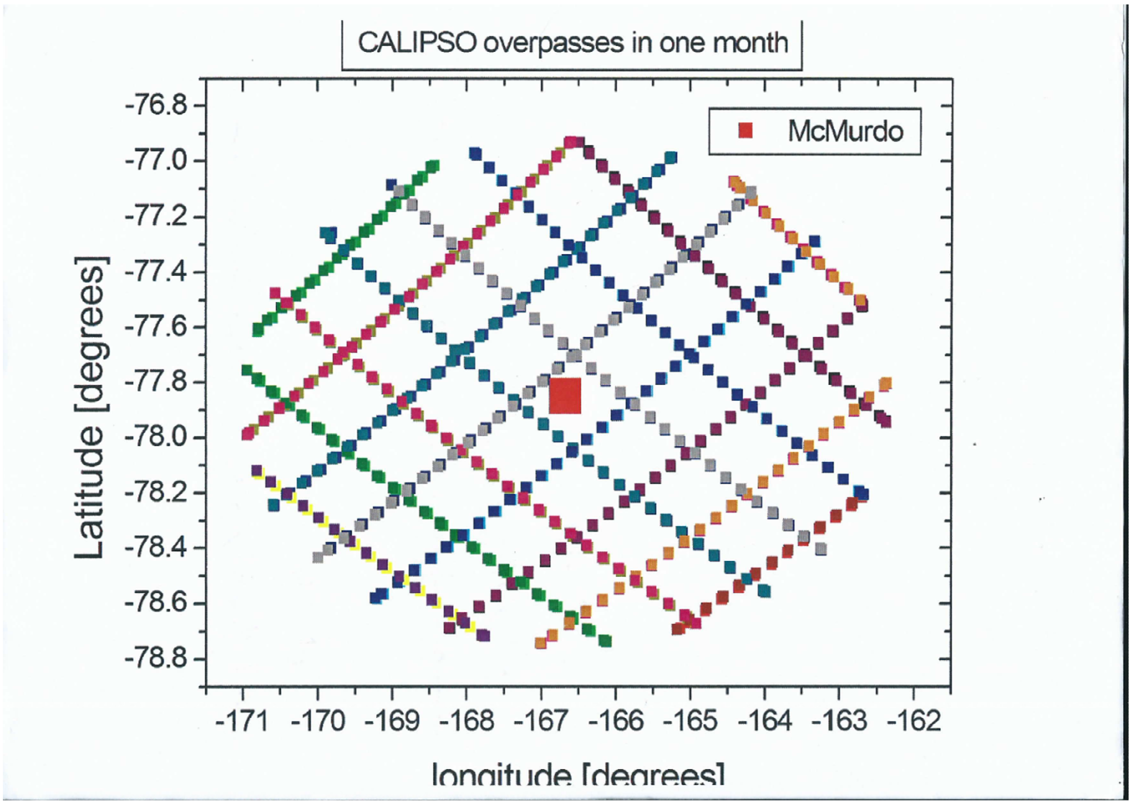
L. 19, p.6: What do the authors mean by “unique definitions” ? Here, the criteria used for ground-based and CALIPSO measurements are different !? This sentence sounds also not very fluent.
ANSWER: The word “unique” has been omitted, since it is not pertinent

L. 3-4, p.7: Does it means that the criteria provided in §2.4, specifically for CALIPSO, are actually not the ones that are really used ? This is quite confusing !
ANSWER: The analysis of the CALIOP data use averaging processes where the signal to noise ratio is low, and varies the threshold on both R and bperp as a function of signal-to-noise ratio. It does not mean that the criteria change, but that other criteria are applied as well, the so-called coherence criteria, taking into account all measured profiles on a piece of the orbit (5-15-45-135 km). It does not influence the analysis of the ground-based data of course.

L. 8, p.7 – l.11, p.9 and Table 1: It is extremely difficult to conclude that the agreement between both plots is good. When focusing on very limited periods showing a clear pattern related to a specific PSC type on one of the plots, the other plot often doesn’t show a similar pattern at the same time and same altitude range. Hence, I cannot agree with the statement in l.6, p.8, that “the overall agreement is rather good”. The authors try to confirm the agreement by providing a statistical comparison over 5 year: this is quite a long time, and I don’t think that the relatively good agreement found between ground-based and CALIPSO for STS, NAT mixtures and ice may provide any real evidence of the agreement between both datasets. I guess it rather gives an overall probability to find a specific PSC type above McMurdo, which is something quite different. For the enhanced NAT mixtures, the situation is even worse since there is about a factor of 2 between the statistics, despite the long time period. Results presented in Figures 2 and 3 are also calculated as averages over a five-year time period, so that they don’t bring more evidence on the agreement between ground-based and CALIOP
measurements. Hence, as suggested by the authors higher in the text, the difference in measurement rate and coverage, different geometry and measurement protocols may induce significant biases in the PSC classification. Did the authors compare directly coincident measurements at specific very limited periods ? Even if, as explained by the authors in l.5-6,
p.7, a point-to-point profile comparison may be unsatisfactory, we should expect that a comparison within a short period shows similar patterns in both plots.
ANSWER:
It is not the goal of the article to make a point-to-point comparison for validation purposes. The goal is to verify if the ground-based measurement are representative for a larger area, typically contained in a 7x2 degrees box around McMurdo.

Apart from that a point-to-point analysis presents the following difficulties:

1) None of the overpasses of CALIPSO are sampling the same air mass as the ground based lidar. To illustrate this I show a plot of all overpasses within the 7x2 degrees box, which corresponds roughly to a distance of 100 km from
.



While CALIOP provides a resolution of 5 km (when integration is required due to low signal-to-noise ratio up to 135 km !) the air mass sampled by the ground-based lidar extends to at most 100 m. (30 km * 3 mrad field of view of the telescope).

Another important difference of the two lidars is that a CALIOP overpass occurs in about 30 seconds, while the ground-based data are integrated over 30 minutes.

This implies that the ground-based measurement integrates air masses moving with a wind speed varying from 0 to 50 m/s, depending also on the altitude (the wind speed might be very different at 15, 20 and 25 km), rendering a comparison with an instantaneous profile of CALIOP very questionable.

However, the statistical analysis is only meaningful if the sampling of the two lidars covers the same period of time and if this period of time has a dense coverage. In order to achieve this we concentrate on 2006, having a large number of observations by both lidars with a good coverage (see figure 1 of the manuscript). We then analyse the months July and August and report the statistics in terms of occurrences of PSC classes and dependence on altitude.

So we follow the suggestion of the referee and analysed short periods with a good time coverage, that is July and August 2006. The referee is correct that an overall statistics covering the 5 year period is not an indication of agreement. We stated that much in the manuscript. See also the answer above to the general comment.

L. 3, p.8: "at the core of the PSC winter season": it might be useful to mention the corresponding period in terms of months.
ANSWER: We added "July and August"

L. 1-5, p.11: I don't see how the different geometries could justify the differences in the results, since Figure 2 presents PSC fractions, and not absolute values. It can be argued that CALIPSO will be more sensitive at high altitude and the ground-based lidars at lower altitude, but I guess this applies to all kinds of PSC. Hence, it is conceivable that the total number of observed events could be affected, but probably not the PSC fractions. Concerning the differences in statistics, how do the authors expect them to influence the agreement between datasets ?
ANSWER: The different observation geometries correspond with different signal to noise ratios at different altitudes. This is valid both for the parallel and perpendicular backscatter coefficient, which constitute the detection and classification thresholds for PSCs. Obviously the PSC class with low values of perpendicular backscatter coefficient (STS) and low values for the parallel backscatter coefficients (NAT) will be more effected by the S/N ratio than ice and enhanced NAT. Since NAT and STS are the most abundant species the S/N ratio has an impact also on the PSC fractions. Moreover, it has been suggested that tropospheric meteorology and cloud cast, which hampers the ground based measurements, may also have an impact on the PSC formation above (On the linkage between tropospheric and Polar Stratospheric cloudsin the Arctic as observed by space–borne lidar, P. Achtert, M. Karlsson Andersson, F. Khosrawi, and J. Gumbel, Atmos. Chem. Phys., 12, 3791–3798, 2012www.atmos-chem-phys.net/12/3791/2012/doi:10.5194/acp-12-3791-2012)

L. 3-4, p.12, Figures 2 and 3: What can explain that the temperature dependence of the NAT fraction may agree quite well between CALIPSO and ground-based measurements (Figure 3), while the same NAT fraction are so different at some altitudes, e.g. around 20-22 km (Figure 2) ? It is unlikely that the number of events is too small at these altitudes to make the estimated fractions statistically not significant.
ANSWER: We found some bugs in the program calculating the fractions. The new results have been discussed in the revised manuscript. (see also answer above to general comments).

L. 8-10, p. 12: I don't understand this conclusion: the differences are manifest on Figure 2.
ANSWER: Differences and agreement have been discussed in the revised manuscript.

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

L. 17-31, p.12: The resolution should be mentioned for the different models and datasets. Resolution aspects play most probably a crucial role in the comparison between models, and with CALIPSO (See also comments on L.4, p.17 and Figure 7).
ANSWER: The resolution is listed in Table 2 of the published manuscript (the referee refers to another older version)

L.14-15, p.13: Which kind of threshold do the authors apply to the SAD when applying the observation operator ? Do the authors mean that they use a mask recording the amount of lidar measurements in every grid cell and putting to zero all grid points that are not covered by any lidar presence ?

ANSWER: A threshold has been defined based on the detection thresholds reported for the v2 detection algorithm of CALIOP. The CALIOP has a very good data coverage and is providing data most of the time, but we might have some grid cells without data. In that case we assume that no PSCs have been observed. This is strictly not correct, but should not affect the overall result, since grid cells without data occur rarely.

L. 16, p.13: The formulation is confusing: is "the sum of all layers" an amount of layers or a distance in km (= amount of layers x 1.5 km) ?
ANSWER: A distance in km.

Caption Figure 4: "the number of km": Please be more specific: does it concern the altitude range ?
ANSWER: YES

L. 6, p.14: What do the authors mean by "NAT-like" ? The ensemble NAT mixtures + enhanced NAT mixture ?
ANSWER : YES in the text we added "NAT plus enhanced NAT"

L.1, p.17: Are there no reasons to think that it is the CALIPSO PSC frequencies that are underestimated with respect to the reality ? I have in mind the way the statistics are processed, the use of monthly means, and the characteristics of the CALIPSO/ground-base station coverage.
ANSWER: The CALIPSO observations are as close to reality as one could wish. The models are surely less "realistic".

L.4, p.17 and Figure 7: "a very large underestimation": with respect to what ? In July, it is very similar to WACCM-cmmi, and very similar to WACCM in August. In September, LMDZrepro is much larger than WACCM. The "very large underestimation" is certainly not general when considering the total PSC frequency. However, it is true when considering the SAD criteria (Figure 7). It has to be noted that LMDZrepro gives overall the closest to CALIPSO in both cases (Total PSC frequency and SAD). Would the similarity with CALIPSO and the outlier character with respect to the other models in the case of the SAD diagnostic be related to the coarser grid resolution of the LMDZrepro model with respect to the other models ?

ANSWER: The sentence should read "The LMDZ model predicts much different NAT (June and July) and ice frequencies (all months) with respect to the other models." We have no reason to assume that the coarser grid of LMDZrepro causes the difference with other models.

L.5, p.17: "The largest biases are found for ice PSCs that tend to be significantly overestimated": Do the authors mean: "underestimated" ? I guess they are still considering the LMDZ model ?
ANSWER: The sentence should read: "The largest biases are found for ice PSCs that tend to be significantly overestimated for all models except for LMDZ, which predicts too small ice frequencies"
L. 7-8, p.17: Taking into account the difference in assumptions, what is the reliability and the robustness of such diagnostic method ? A sensivity study might be needed.
ANSWER: Even if differences in the assumptions on the mean particle size may be critical, all the models have constructed and have tuned their parameterization in order to simulate

a correct PSCs polar chemistry. The aim of this section is to show the variability between the CCMs in their SAD by comparing to realistic estimate of this range derived from the CALIOP observations for NAT and ICE, and not to score them. We propose this diagnostics (the range derived from observations) to be compared with the models in order to derive implications for simulated heterogeneous chemistry. Reviewer is right as a sensitivity study on instantaneous model outputs in Specified Dynamics runs would be needed to tune the proposed diagnostics and turn it into a specific set of metrics. A clarifying sentence has been added in section 3.3.

L. 6, p.18: “This in turn would give less irreversible denitrification processes than in the case of simulation by the models with larger NAT SAD” ?

ANSWER: What we mean here is that a smaller NAT radius would therefore give less irreversible denitrification.

L.4, p.19: occurences of what ? Please be more specific.

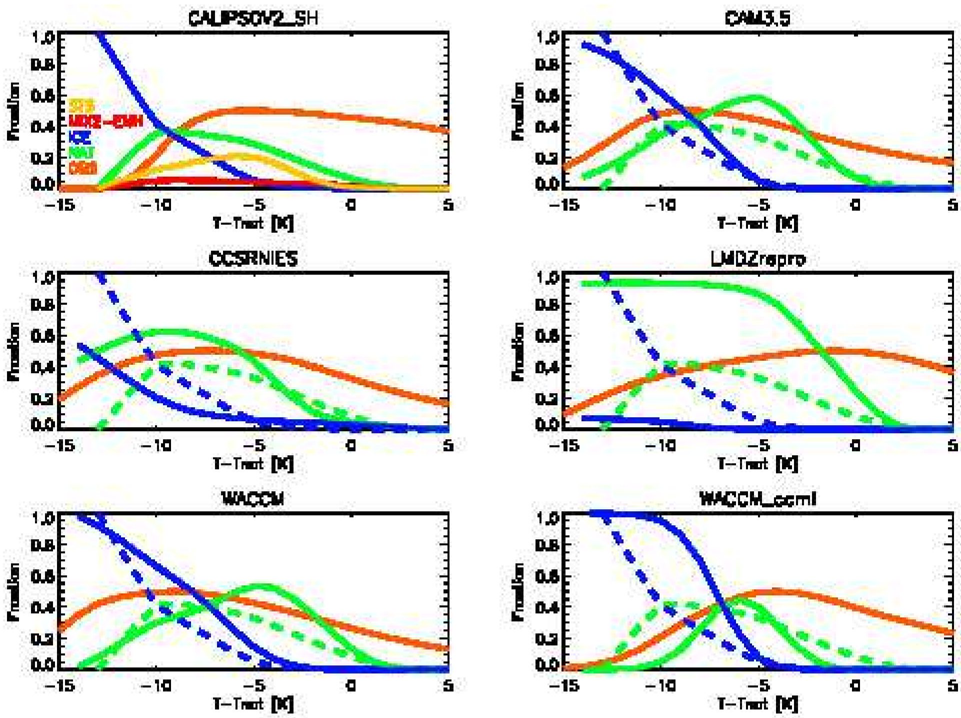
ANSWER: We mean the occurrences of the different PSC types as observed by CALIOP and simulated by the models (NAT and ice only)

L. 6, p. 19: How is the averaging performed ? As a simple mean of all numbers ? Or by weighting by the grid cell area ? Concerning CALIPSO, how do the authors use the monthly means ? By making a mean of means ? Averaging yet averaged values may affect significantly the results.

ANSWER: For the models the grid cells have been summed, for CALIOP the data have been gridded on a horizontal grid of 10x3.5 (lat-lon) degrees, and a vertical resolution of 1.8 km. The averages have been made by summing over all cells and months.

L. 10-12, p.19: “Too slow”, “too fast”: with respect to CALIPSO ? This should be specified. What do the authors mean by “progression for ice/NAT” ?

ANSWER: The expressions “too fast” and “too slow” are with respect to CALIOP. The sentence “progression for ice/NAT” means that the increase of NAT and ice fractions occurs with a stronger temperature (T-TNAT) with respect to CALIOP (dashed lines in the new figure)



L. 1, p.20: “The fraction of data with different PSC”: Please revise the formulation.

ANSWER: The sentence has been reformulated as follows. “The temperature dependence of the fractions of the different PSC types helps in evaluating…..”

L. 3, p.20: the fraction of what ? Please be specific ! “an increase of ice with TTNAT < -5K”: Please revise the formulation: increase with decreasing temperature.
ANSWER: The sentence has been reformulated as follows.”The CALIOP data show a steady increase of the NAT fraction with decreasing T-TNAt up to a value of -10 K, while the increase of the ice fraction shows a higher slope belowe T-TNAT = -10 K.
L.5, p.20: “a sharper increase of the fraction”: fraction of what ?
ANSWER: The sentence has been reformulated as follows.”The increase of NAT and ice fraction for lower temperatures
L. 7, p.20: “while for the other models, the ice…”.

ANSWER to the previous three comments. Figure 8 has been edited to show the dependences of CALIOP also in the graphs of the models as dashed lines. This facilitates the comparison of models with CALIOP. The paragraph has been reformulated.
“The onset of NAT is similar for all models, except for WACCM-ccmi, where NAT starts to form only below Tnat. The onset of the ice formation occurs at T-Tnat = -5 K for all models, except for CCSRNIES. The increase of NAT occurrences with decreasing temperatures is stronger for all models with respect to CALIOP. This is due to the fact that the models consider only the thermodynamic equilibrium conditions for the formation of PSC, and do not allow the existence of supersaturation without PSC formation. The family of models CAM3.5, WACCM and WACCM-ccmi show a faster increase of the ice occurrences with decreasing temperatures with respect to CALIOP. The reason is probably the same as for the NAT behaviour. LMDZ-repro evidently produces much less ice than the other models and CALIOP, and at low temperature NAT is the dominating species, while the other models and CALIOP show a dominant ice occurrence for low temperatures. The CCSRNIES model shows a slower increase of the ice occurrences with respect to CALIOP and the other models.“

4. Conclusions.

L. 12, p.20: A point-to-point comparison is always feasible ! The issue is to know if it is valid and reliable.
ANSWER: The referee is correct in stating that a point-to-point comparison is always feasible, but the point is if it makes much sense to do so. As has been pointed out above, many sources of biases exist and any single comparison of two observations might suffer more or less from one or more biases. So one should perform a statistical analyses on a large number of point-to-point comparison. This is not very different from our approach; we show that for short periods with many co-located observations, in particular July and August 2006. We agree with the referee that the statistics for a five year period does not confirm the agreement between the two datasets, but merely demonstrates that both instruments measure an average occurrence of all PSC types.
The text has been adapted along these lines.
L. 14, p. 20: “very similar”: Based of the results presented in Figure 1, I don’t agree. (See comment above). At least, a statistical indicator and quantitative estimates of the uncertainty should provided.
ANSWER: We agree with the referee that it is preferable to consider only short periods with a good coverage of both instruments.
L. 16, p.20: As already mentioned, I don’t understand the emphasis on “below15 km”. Is it based on Figure 2; If well, this seems very subjective to me.
ANSWER:The discussion about above/below 15 km has been eliminated. It was based on a figure which proved to be wrong, due to several bugs in the normalization of the fractions
ANSWER:
L. 20, p.20: “Models fail to reproduce realistic geographical distributions of PSCs”: I am really not convinced by the demonstration made in this paper. A significant part of the problem

might come from the way the authors implement their different methodologies, and more particularly from the comparison of things that are not really comparable.
ANSWER: The more symmetric distribution of PSCs in the models with respect to CALIOP is probably due to the incorrect temperatures produced by the models, since they don't include temperature fluctuations due to gravity waves.
L. 22, p.20: The more recent WACCMI-ccmi model compared better with CALIOP only for one specific diagnostic method (based on the total PSC frequency). The issues is to understand why: in view of all my previous criticisms, it might be fortuitous.
ANSWER: WACCM-ccmi is really very similar to previous versions. The better agreement s exclusively based on the temperature behaviour.

Technical corrections:
L32, P2. Has been corrected
L11,p3 has been corrected
L2p4, acquisition has been corrected
L18 P4 . done
L14-20, p5, The suggestion of the referee has been followed
L32, P5 done
L4, P6 re-elaborated
L9P6 the sentence has been eliminated because out of place
L13P6 corresponding done
L16P6 5-year done
L18 P6 done
L19 P6 induce OK
L2, 6 P7 signal-to-noise substituted all over the text
L8 P11, corrected
L17, P11 definition TNAT CHECK !!
L1.P12 this is not anymore present in the correct pdf file
L6P13 ok
Caption fig 4 has been corrected

Comparison of Antarctic polar stratospheric cloud observations by ground-based and spaceborne lidars and relevance for Chemistry Climate Models

Marcel Snels[1], Andrea Scoccione[1], Luca Di Liberto[1], Francesco Colao[2], Michael Pitts[3], Lamont Poole[4], Terry Deshler[5], Francesco Cairo[1], Chiara Cagnazzo[1], and Federico Fierli[1]

[1]Istituto di Scienze dell'Atmosfera e del Clima, Via Fosso del Cavaliere 100, 00133 Roma
[2]ENEA, Via Enrico Fermi 45, 00044 Frascati
[3]NASA Langley Research Center, Hampton, Virginia 23681, USA
[4]Science Systems and Applications, Inc., Hampton, Virginia, 23666, USA
[5]University of Wyoming, Laramie, WY 82071, USA

Correspondence to: Marcel Snels (m.snels@isac.cnr.it)

Abstract. A ~~statistical~~ comparison of polar stratospheric clouds (PSCs) occurrence from 2006 to 2010 is presented, as observed from the ground-based ~~station~~ lidar station at McMurdo (Antarctica) ~~, included as a primary station in~~ and by the satellite-borne CALIOP lidar (Cloud-Aerosol Lidar with Orthogonal Polarization) measuring over McMurdo. McMurdo (Antarctica) is a primary lidar station for aerosol measurements of the NDACC (Network for Detection of Atmospheric Climate Change)~~, and by the satellite-borne CALIOP lidar measuring over McMurdo~~. The ground-based observations have been classified with an algorithm derived from the recent ~~V2~~ v2 detection and classification scheme, used to classify PSCs observed by CALIOP.

A statistical approach has been used to compare ground-based and satellite based observations, since point-to-point comparison is often troublesome due to the intrinsic differences in the observation geometries and the imperfect overlap of the observed areas.

~~This~~ A comparison of space-borne ~~, ground-based~~ lidar observations and a selection of simulations obtained from Chemistry Climate Models ~~,~~ has been made by using a series of quantitative diagnostics based on the statistical occurrence of different PSC types. The distribution of PSCs over Antarctica, calculated by several CCMVal-2 and CCMI chemistry climate models has been compared with the PSC coverage observed by the satellite ~~based~~ borne CALIOP lidar. The use of several diagnostic tools, including the temperature dependence of the PSC occurrences, evidences the merits and flaws of the different models. The diagnostic methods have been defined to overcome (at least partially) the possible differences due to the resolution of the models and to identify differences due to microphysics (e.g. the dependence of PSC occurrence ~~from~~ on T-T_{NAT}).

A significant temperature bias of most models has been observed as well as a limited ability to reproduce the longitudinal variations in PSC occurrences observed by CALIOP. In particular a strong temperature bias has been observed in CCMVal-2 models with a strong impact on PSC formation. The WACCM-CCMI (Whole Atmosphere Community Climate Model - Chemistry Climate Model Initiative) model compares rather well with the CALIOP observations, although a temperature bias is still present.

1 Introduction

Lidar observations have been extensively used to characterize the occurrence of PSCs in the polar stratosphere (see e.g.**????**) . The observed optical parameters allow to discriminate different cloud types, such as STS (supercooled ternary solution), NAT (nitric acid trihydrate) and water ice, and external mixtures of the former. Pitts and co-workers ~~(???)~~(**????**), calculated the optical parameters of cloud particles with different size distributions and chemical composition in order to define a PSC classification, which was then applied to the CALIOP data. Achtert and Tesche (**?**) made an assessment of several lidar-based PSC classifications and their impact on the occurrences of the different PSC types. Their conclusion was that the comparison of PSC classifications obtained from different lidar observations is not straightforward and should take into account the measurement technique and classification methodology used. ~~Many different schemes have been proposed with~~ A variety of schemes using different thresholds for detection and classification have been proposed, rendering a comparison difficult. Here we want to compare ground-based and satellite based lidar data, by using a detection and classification scheme for the ground-based data, which closely approaches the new ~~V2~~ v2 classification scheme used for CALIOP (**?**).

Ground-based lidar observatories provide a unique data base, having decadal coverage, albeit with discontinuities, spanning from the ~~early nineties~~ middle eighties to today.

The first lidar observations in Antarctica started in 1985 at Syowa Station. Iwasaka and co-workers (**??**) used a polarization sensitive lidar to measure backscatter and depolarization to observe PSCs. Later, in 1987/1988, at the Amundsen-Scott South Pole Station, Fiocco and co-workers (**?**) used the elastic backscatter signal from a lidar operating at 532 nm to observe PSCs in relation to the temperature. PSCs have also been observed at Davis, from 2001 to 2004 (**?**) and at Rothera (**?**) from 2002 to 2005.

Long-term observations of PSCs have been performed at McMurdo (**????**), from 1989 until 2010 and at Dumon D'Urville (**???**), from 1990 until now, both with polarization sensitive lidars. Recently the McMurdo lidar has been transferred to Dome C and is operating there from 2014 on (**?**).

A clear issue is that the representativeness of ground-based long-term lidar data series of the Antarctic stratosphere might limit their value in climatological studies and model evaluation. Since the long-term ground based lidar observations have been performed only in few locations, the comparison with model simulations and satellite borne instruments is necessarily limited to these locations, which poses a limit to their use. The recent availability of satellite-borne lidar observations provides an almost complete coverage of the globe, and presents the opportunity to test the polar stratospheric cloud scheme of Chemistry Climate Models (CCMs) on synoptic scales. The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) was launched in April 2006 with the primary objective of improving our understanding about the impact of clouds and aerosol on the climate. ~~The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP)~~ CALIOP provides total backscatter and depolarization profiles, allowing classification of the observed clouds and aerosols. The original CALIPSO mission had a minimum time frame of 3 years, but has been ~~extented~~ extended several times and is still active.

Comparison between CALIOP and ground-based observations in the Antarctic stratosphere of PSCs is thus possible from 2006 on and has been pursued in the case of McMurdo Station by performing co-incident measurements with CALIPSO overpasses whenever possible.

Due to their primary role in ozone chemistry, a correct representation of PSCs in ~~Chemistry Climate Models~~ CCMs is needed. Actually, the parametrization of PSC formation in most CCMs depends only on temperature thresholds and on nitric acid and water vapour concentrations for the determination of supersaturation conditions. A rather complete description of the parametrizations used in state-of-the-art CCMs is reported in **?**. The SPARC Report N^{o}5 (2010) Chemistry-Climate Model Validation (CCMVal-2) (**?**) has shown that Chemistry Climate Models can have a biased representation of the stratospheric conditions with colder temperatures that lead to an overestimate of ozone depletion, also due to an unrealistic PSC coverage. Hence PSC simulations show a large uncertainty, as reported in the CCMVal-2 report. Nevertheless, the report presents a preliminary evaluation based on global averages with a subset of CALIOP data.

The most recent ~~Chemistry Climate Models (CCM)~~ CCMs are able to reproduce the denitrification by the formation of STS and NAT and the dehydration through the formation of ice clouds, but use rather approximate schemes based on temperature thresholds for the onset of nucleation, with additional constraints on how much of the available nitric acid is depleted by STS and NAT formation. Although the overall denitrification and dehydration can be represented rather well, the correct description of the formation of STS and NAT, and mixed type PSCs would need a more sophisticated microphysics model.

In the present work we first compare the statistics of occurrence of different PSC classes in the stratosphere over McMurdo ~~station~~Station, as detected by the ground-based lidar operating there and the satellite-borne CALIOP. Subsequently we use the full coverage of the Antarctic CALIOP data to assess the performances of different CCMs in simulating PSC occurrences and PSC distribution over Antarctica.

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

2.1 CALIPSO PSC observations

The CALIPSO satellite was launched in April 2006 as a component of the A-train satellite constellation (**??**). With an orbit inclination of 98.2 ~~^{o}~~°, it provides extensive daily measurement coverage over the polar regions of both hemispheres, up to 82 ~~^{o}~~° in latitude. It hosts the CALIOP ~~(Cloud Aerosol Lidar with Orthogonal Polarization)~~ two wavelength polarization diversity lidar, that measures backscatter at wavelengths of 1064 nm and 532 nm, the latter signal separated into parallel and cross polarization, with respect to the polarization of the outgoing laser beam. Details ~~on~~ of CALIOP can be found in ~~(**??**)~~**?** and **?**. CALIOP data have extensively been used for observing PSCs and improved algorithms for PSC classification have been reported in ~~**???**~~**????**.

2.2 Ground-based PSC observations at McMurdo

A Rayleigh polarization diversity lidar has operated in the Antarctic station of McMurdo since 1991, in the framework of an USA-Italian collaboration (**??**). It measures aerosol backscatter and depolarization profiles from 12 km to 30 km, with a vertical resolution of 30 meters. Aerosol backscattering is retrieved using the Klett algorithm (**?**) and the extinction is calculated according to **?**. The depolarization is calibrated following the method described in **?**. The lidar was operated by science technicians of the National Science Foundation (NSF) during the Antarctic winter, typically from the end of May until the end of September to cover the whole period of PSC occurrence. Potential vorticity reanalysis shows that McMurdo is well within the stratospheric polar vortex from mid-June to the end of September, except for rare events of major vortex perturbation. As a routine, the lidar is operated at the same time every day when meteorological conditions are favorable, or at the earliest chance to do so, for about 30 minutes to render a single profile. When possible, the observations are synchronized with overpasses of the CALIPSO satellite, when its footprint is within 100 km distance from McMurdo. Observations are intensified in coincidence with Optical Particle Counter (OPC) and ~~ozonesondes~~ ozone sondes balloon measurements (**?**). All observations at a wavelength of 532 nm used in the present analysis have been quality checked and the relevant data are publicly available in the NDACC data base (ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/mcmurdo/ames/lidar/).

For the ground-based lidar data a single vertical profile with a vertical resolution of 150 m is obtained by averaging 30 minutes of ~~acquistion. In order to facilitate the statistical analysis of the data, at most one profile is taken for each 6-hours time window.~~ acquisition.

2.3 PSC detection and classification

PSC detection and classification from lidar measurements with orthogonal polarization is usually based on two optical parameters derived from the optical signals with parallel and perpendicular polarization with respect to the laser, the backscatter ratio and the aerosol depolarization. Here we use the backscatter ratio R and the perpendicular backscatter coefficient $\beta_\perp$, in order to be consistent with the v2 detection and classification scheme used for the CALIOP data. The backscatter ratio is defined as

$$R = (\beta^{aer} + \beta^{mol})/\beta^{mol} \tag{1}$$

where β^{aer} is the total aerosol backscatter and β^{mol} is the total molecular backscatter.

~~If we define volume and molecular depolarization as~~

$$\delta^{vol} = \beta_{perp}/\beta_{par}$$

~~and~~

$$\delta^{mol} = \beta^{mol}_{perp}/\beta^{mol}_{par}$$

the aerosol depolarization can be expressed as (?):

$$\delta^{aer} = \beta^{aer}_{perp}/\beta^{aer}_{par} = \frac{(1+\delta^{mol})\delta^{vol}R - (1+\delta^{vol})\delta^{mol}}{(1+\delta^{mol})R - (1+\delta^{vol})}$$

where β^{aer}_{par} and β^{aer}_{perp} are the parallel and perpendicular aerosol backscatter, respectively.

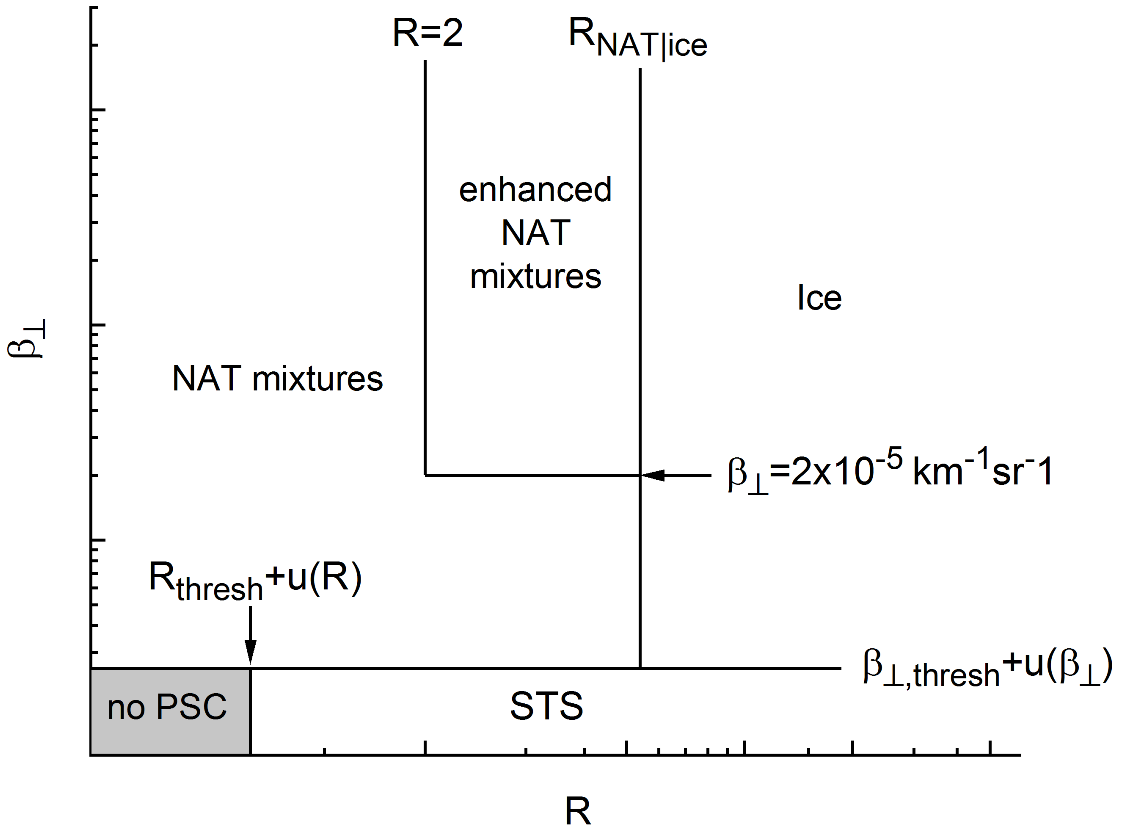
We must bear in mind that for all lidar measurements the optical parameters represent an average value of the microscopic properties of an ensemble of of many particles in a large air volume which may belong to different composition classes. Only rarely the observation of an air volume can be totally attributed to a single class of particles, except for particular cases where the temperature conditions exclude the co-existence of particles with different compositions. Thus the resulting macroscopic optical parameters are mostly due to external mixtures and are dominated by the species with the largest relative abundance and/or the largest optical parameters. When classifying the PSCs, the classifications indicate the dominant species or the mixtures of species.

Recently an overview of different detection and classification procedures has been reported by ?, showing how the different algorithms applied to

2.4 PSC Detection and classification criteria for the CALIPSO v2 data

The CALIOP v2 PSC detection and composition classification algorithm (?) has been used to create the recently released CALIOP v2 PSC mask database covering the period from June 2006 to October 2017. Here we compare these v2 data with ground-based observations at McMurdo from 2006 to 2010. Major enhancements in the same data produce a variety of classifications. Here we limit the discussion to the data obtained by the CALIOP lidar and the McMurdo lidar observatory. Pitts and co-workers proposed a detection and classification algorithm for PSCs in 2009 (?), which has been slightly modified a few years later (?). This algorithm used a classification of NAT mixtures in three classes, according to the different presence of NAT (Mix1, Mix2, Mix2-Enhanced). In 2013 an assessment has been performed on the PSC classification methods used (?) . In particular they examined the boundaries drawn between the composition classes and the causes of minor misclassifications and discussed the "cross-talk", meaning an overlap between composition classes in optical space, due to measurement noise . They observed that 5–6 % of PSCs classified as STS might actually be NAT mixtures, whereas only 1–2 % of PSCs classified as NAT mixtures might actually be STS. In this case the cross-talk was mainly due to the measurement noise for small values of β_{perp}. Their conclusion was that little cross-talk occurred between NAT-mixtures and STS, and that the ice assignment was very robust. On the other hand the separation of the NAT mixtures in Mix1, Mix2 and Mix2-enhanced classes was less reliable. This depends on the measurement uncertainties in R and δ^{aer} on one side and also on the location of the boundary between ice and NAT mixtures, which depends on the degree on denitrification (?)v2 algorithm over earlier versions include daily adjustment of composition boundaries to account for effects of denitrification and dehydration, and estimates of the random uncertainties u($\beta_\perp$) and u(R) due to shot noise in each data sample, which are used to establish dynamic detection thresholds and composition boundaries. The CALIOP v2 algorithm is represented pictorially in Figure 1 and is described in more detail in the following sections.

This classification method, with the exception of the mountain wave ice class has also been applied to a previous analysis of the McMurdo ground-based lidar data (?).



2.5 PSC Detection and classification criteria for the CALIPSO V2.0 data

Figure 1. The figure shows the detection and classification criteria of the V2 CALIOP algorithm. The classification as STS, NAT mixtures, enhanced NAT mixtures and ice, requires that threshold conditions for R and/or $\beta_\perp$ are satisfied. See the text for details.

A new version V2.0 of the CALIPSO data based on a different algorithm has been made available recently (?) and CALIOP data analyzed with this algorithm have been released as V2 CALIOP data. Here we compare these V2 data with ground-based observations at McMurdo from 2006 to 2010. The differences between V1.0 and V2.0 CALIOP PSC algorithms are listed below. Data pre-processing - Level 1 CALIOP data are now corrected for (minor) cross-talk between the parallel and perpendicular polarization channels. Uncertainties (“unc”) are calculated for all quantities based on the noise-scale factor (NSF) embedded in Level 1 data.

2.4.1 PSC detection

PSC detection - The thresholds for the background aerosol are now defined PSCs are detected in the CALIOP data as statistical outliers relative to the background stratospheric aerosol population. The v2 background aerosol thresholds $\beta_{\perp,thresh}$ and R_{thresh} are calculated as the daily median plus one median deviation of 532-nm perpendicular backscatter (β_{perp}) and scattering ratio (R) of CALIOP data at ambient temperatures above 200 K. PSCs are those data points for which β_{perp} or R exceeds the respective background threshold by more than unc(perp) or unc(R).

PSC composition - The former Mix1 and Mix2 classes have been combined into a single NAT mixtures class. The former Mix2-enhanced class has been renamed enhanced NAT mixtures and it now includes only those NAT mixtures with R either $\beta_{\perp}$ > 2 and β_{perp} $\beta_{\perp,thresh}$+u($\beta_{\perp}$) or R> $2\cdot10^{-5}$ $m^{-1}sr^{-1}$. The wave ice class remains the same as in Version 1.0, i.e. ice PSCs with R>50. Each point detected as a PSC is assigned a non-spherical particle confidence index CI(NS) = β_{perp} – unc(perp)/ unc(perp) and an STS confidence index CI(STS) = R– unc(R)/ unc(R). R_{thresh} +u(R). If CI(NS)≤ 1 and CI(STS)$\beta_{\perp} \leq \beta_{\perp,thresh}$ +u($\beta_{\perp}$) and $R \leq R_{thresh}$ +u(R), the point is a non-PSC. Noise spikes are eliminated in the CALIOP v2 data by requiring coherence within a running 3-point vertical by 5-point horizontal along-track box.

2.4.2 PSC composition

The PSC composition is determined as follows:

- If $\beta_{\perp} \leq \beta_{\perp,thresh}$ +u($\beta_{\perp}$), but R > 1R_{thresh} +u(R), the PSC is classified as STS. If CI(NS)
- A PSC with $\beta_{\perp}$ > 1, the PSC is presumed $\beta_{\perp,thresh}$+u($\beta_{\perp}$) is assumed to contain non-spherical particles and is classified as a NAT (or enhanced NAT) mixture or ice based on its value of R. The position of the boundary R. The boundary value separating ice from NAT (and enhanced NAT) mixtures, $R_{NAT|ice}$, is now mixtures, $R_{NAT|ice}$, is calculated based on the total abundances of HNO_3 and H_2O vapors . These are as determined on a daily basis as a function of altitude and equivalent latitude from nearly coincident cloud-free Aura MLS data. For estimating the total abundances of HNO_3 and H_2O, one should avoid MLS measurements that are impacted by uptake of vapour by PSC particles, since MLS only observes the gas-phase abundances. Since the MLS and CALIOP measurements are essentially co-located, all MLS profiles where CALIOP detects the presence of clouds, have been ignored. So for each day, the mean HNO_3 and H_2O abundances have been calculated as a function of altitude and equivalent latitude only on MLS profiles in clear (cloud-free) air. Then each point containing non-spherical particles is assigned an NAT|ice confidence index CI(NAT|ice) = R – $R_{NAT|ice}$/ unc(R). For points classified as ice or wave ice, CI(NAT |ice)
- If $\beta_{\perp}$ > 0. For NAT (or enhanced NAT) mixtures, CI(NAT|ice) ≤ 0. $\beta_{\perp,thresh}$+u($\beta_{\perp}$) and $R > R_{NAT|ice}$, the PSC is classified as ice.

- If $2 < R < R_{NAT|ice}$ and $\beta_{\perp} > 2\cdot10^{-5}$ $m^{-1}sr^{-1}$, the PSC is classified an enhanced NAT mixture. All other PSCs with $\beta_{\perp} > \beta\perp_{thresh}$ +u($\beta_{\perp}$) and $R < R_{NAT|ice}$ are classified as NAT mixtures.

The CALIOP ~~V2 data set provide both the PSC classified grid~~ v2 data set provides both the grid of classified PSCs according to the ~~V2~~ v2 algorithm and the associated optical parameters. ~~From here on we will call the first CALIOP V2 PSC product and the latter CALIOP V2 data, for clarity.~~

2.5 PSC Detection and classification criteria for the ground-based data

In order to compare the ground based lidar data to the CALIOP data we have adopted a new algorithm which follows the same approach and uses the same optical parameters as the ~~V2 CALIOP algorithm~~ v2 CALIOP algorithm (see Figure 1).

2.5.1 PSC detection

The ground-based raw data have been ~~reelaborated~~ re-elaborated to produce the backscatter ratio ~~R~~ R and the perpendicular backscatter coefficient ~~β_{perp}. The detection algorithm uses an average threshold of the backscatter ratio R_{thr}~~ $\beta_{\perp}$. While the determination of the background aerosol thresholds for the CALIOP data uses a very large number of observations, the quantity of ground-based lidar data is much smaller and does not allow a similar treatment. Instead of using daily medians we calculated a median value from all ground-based data in the 5-year period without PSCs (typically before 15 June or after 1 October) or in obvious clear sky conditions. Thus the background aerosol thresholds were determined as the median values plus one standard deviation about the median. In this way we obtained fixed background thresholds for the backscatter ratio R_{thres} =~~1.15, derived from the average signal to noise ratio of the lidar signals, and a threshold for β_{perp}, $\beta_{perp,thr}$ of 5~~1.15, and also for $\beta_{\perp}$ =1·10~~$^{-7}$~~$^{-6}$ $m^{-1}sr^{-1}$~~, estimated from observations without the presence of PSCs. All data with R~~. While most PSC detection schemes for ground-based lidar data use a threshold only for R (**?**), the scheme used here is more permissive and allows all data with R >1.15 ~~or β_{perp}~~ +u(R) or $\beta_{\perp,thresh}$>~~5~~1·10~~$^{-7}$~~$^{-6}$ $m^{-1}sr^{-1}$~~, and~~ +u($\beta_{\perp}$), where u(R) and u($\beta_{\perp}$) are the running standard deviations over altitude, and a local temperature below 200 K in a range between 12 and 30 km ~~are considered~~ to be detected as PSCs. Note that this procedure is ~~essentially the same as the V2 algorithm for CALIOP~~ very similar to the v2 CALIOP algorithm, except that we use fixed background thresholds and different estimates of the uncertainties in the data. Finally~~a coherence criterion has been applied, in order to avoid to detect isolated "spikes "~~, to mimic the CALIOP coherence criteria, we require continuity along the vertical profile to avoid identifying isolated noise spikes as PSCs. ~~This coherence criterion requires a continuity on the profile , i.e. in the vertical dimension, while the coherence criterion for CALIOP takes into account profiles along the flight track.~~

~~The PSCs are successively classified as STS, NAT mixtures, enhanced NAT mixtures and ice, by using similar criteria as reported for the CALIOP V2 algoritm. The discrimination between NAT mixtures and enhanced NAT mixtures is made by using the condition R > 2 and β_{perp} > $2\cdot10^{-5}$ $m^{-1}sr^{-1}$, while the NATIice confidence index CI(NATlice) has been taken from the corresponding CALIOP data, since it is based on the total abundances of HNO_3 and H_2O vapors, and thus valid for CALIOP and~~

2.5.2 PSC composition

Composition classification for ground-based ~~data. We used monthly averages , since the variations during one month were not significant, and it largely simplified the procedure. The non-spherical particle confidence index CI(NS) = β_{perp} – unc(perp)/ unc(perp) has been calculated from~~ PSCs is nearly identical to the CALIOP v2 procedure, the ~~standard deviation of the observed β_{perp} parameter. This procedure approaches the V2 algorithm applied to the CALIOP data as good as possible, and should provide a solid base for comparison~~exception being that we use monthly averages for $R_{NAT|ice}$ computed from daily values included in the v2 CALIOP data files.

2.6 Comparison of ~~coincident~~ co-located PSC observations at McMurdo from the ground and from CALIPSO during the ~~5 year~~ 5-year observation period

Here we compare PSC statistics from ground-based and satellite-borne lidars, with the goal to assess if ~~their different measurement procedures induces~~ the differing measurement procedures used for each of them, induce a bias in the PSC classification, which might hamper the ~~unique~~ definitions of useful common diagnostics for assessing the performance of regional and global models.

A comparison is made using ~~264~~ 248 profiles acquired by the ground-based lidar and ~~8082 profiles~~ 585 overpasses extracted from the CALIOP data base within a 7~~^{o}x1^o~~$^\circ$x2° longitude-latitude box centered on the McMurdo site for the years 2006-2010. The choice of the box dimension is dictated by the need to have a minimal latitudinal range (to avoid the inclusion of data in different vortex regimes for stations close to the polar vortex edge) and to have a significant number of observations (a larger longitudinal range assuming a local uniform distribution around the site) and corresponds roughly with a distance of 100 km from McMurdo.

~~The comparison~~ CALIOP overpasses do not occur every day and at most twice per day. In average we have up to 40 CALIOP overpasses per month. Ground-based lidar data are mostly recorded during a CALIOP overpass, but also on days without CALIOP overpasses, usually at the same time that CALIOP overpasses occur and sometimes at different times from the CALIOP overpasses. The latter are not included in this analysis. All other ground-based measurements have been used in the statistical comparison. Generally speaking most of the ground-based profiles have been recorded during a CALIOP overpass, but there might be days with either a ground-based measurement or a CALIOP measurement. So we include all CALIOP measurements falling in a spatial box around McMurdo, and all ground-based data measured in a time frame dictated by CALIOP overpasses, including also the days without overpass.

The comparison between data obtained by space-borne and ground-based instruments is not straightforward. Lidars on satellites provide altitude resolved PSC observations on a synoptic scale, with fixed revisit times on the ground spot, and their observations in the stratosphere are unaffected by tropospheric visibility. Ground-based observations are limited by the weather conditions and become prohibitive in case of heavy cloud cover. Moreover the measurements occur once or twice per day, possibly in co-incidence with satellite overpasses. Sometimes they are conditioned by other activities such as intensive measurement campaigns of other instruments. The different geometry and measurement protocols might induce a bias in PSC statistics of ground-based and satellite-based lidar observations.

The ground-based lidar observes ~~from~~ at distances up to 30 km from the ground, while the satellite based lidar is in orbit at 705 km and observes backscattering from distances around 700 km. This implies that the ~~signal to noise~~ signal-to-noise ratio of CALIOP is in general lower than that of the ground-based lidar. Therefore the CALIOP data use averaging processes where the ~~signal to noise~~ signal-to-noise ratio is low, and varies the threshold on both ~~R and β_{perp}~~ R and $\beta_{\perp}$ as a function of signal-to-noise ratio.

For these reasons, a point-to-point profile comparison of these data bases may not be sufficient to evaluate whether or not the instruments provide a compatible information of PSCs coverage and partition in different classes, which, at the end is the information needed to evaluate models and provide a climatic survey of the polar stratosphere.

The purpose of this analysis is not to perform a validation of the satellite-borne instrument, but to verify if the two instruments provide compatible information in terms of occurrences of the different PSC classes around McMurdo.

In order to illustrate how ground-based and space-borne lidar observations of PSCs compare, we show as an example the height-time evolution of PSC classes for CALIPSO and McMurdo data bases for the year 2006 (see Figure ~~??~~2), having the best temporal coverage with respect to the other years (2007-2010).

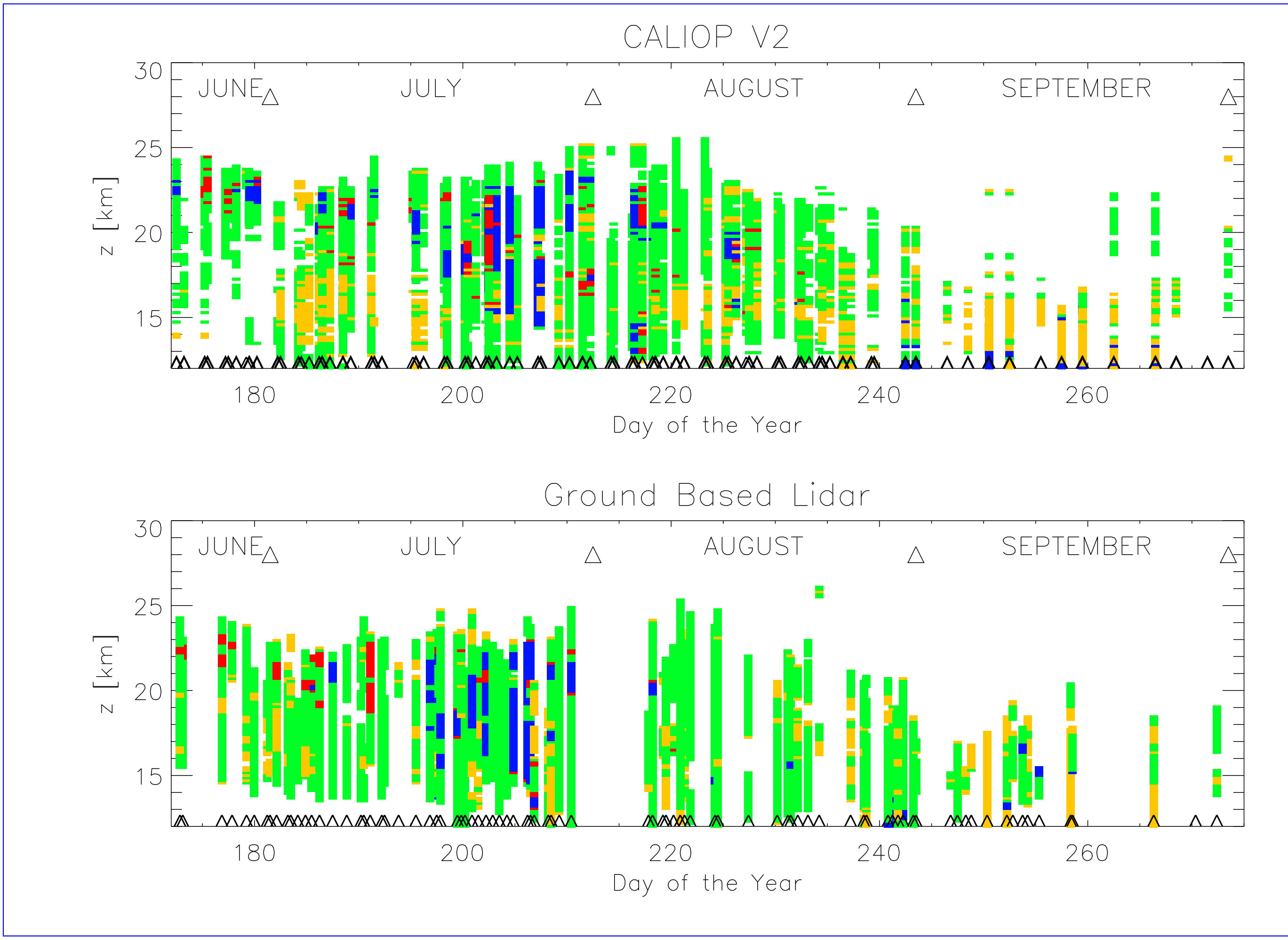


Figure 2. PSC observations recorded in 2006 above McMurdo. Upper panel: CALIOP around McMurdo from ~~V2~~ v2 product. Lower panel: Ground-based lidar data. The PSC classes are represented by colors; green = NAT mixtures, orange = STS, blue = ice, red = enhanced NAT mixtures. Triangles on the x-axis indicate the day when at least one observation was available.

Both the CALIOP PSC product, and the classification of the ground-based lidar optical parameters, ~~classified with the V2~~ obtained with the v2 algorithm adapted for ground-based data, provide a similar view for this winter with a dominance of NAT mixtures with isolated periods of ice PSCs ~~at the core of the PSC winter season~~in July. Enhanced NAT mixtures appear ~~jointly to ice clouds while STS are~~ mostly in June and July, around and above 20 km, while STS has been observed in the lower layers ~~at the beginning and at the end of the season~~throughout the season, being the major species in September. These results are not directly comparable with the analysis previously reported (**?**), where a ~~more classical~~ different classification scheme for ground-based data was adopted and different PSC classes were assigned. Although the overall agreement with CALIOP is ~~rather good~~acceptable, many small differences are evident, and confirm that a point-to-point comparison of these data is not straightforward.

	2006		2006-2010	
PSC classes	~~McMurdo ground-based %~~CALIOP	~~McMurdo CALIOP%~~gr.based	CALIOP	gr.based
STS	~~25.5~~15.5	~~24.2~~14.6	22.4	13.8
NAT mixtures	~~59.4~~73.6	~~57.9~~76.0	60.1	71.6
enhanced NAT mixtures	~~1.5~~2.3	2.4	2.5	2.6
ice	~~13.6~~8.5	~~15.3~~7.1	15.0	12.0
overpasses/observations	128	75	615	248

Table 1. Frequency of occurrence (in %) of ~~PSCclasses during June-July-August-September between 12~~ PSC classes for 2006 and ~~30~~ for 2006-2010. The last line represents the number of overpasses in the McMurdo box for CALIOP and the number of ground-based observations. Observations in the 12-30 km ~~height~~interval have been considered. The ice-class for CALIOP includes also mountain wave ice.

For this reason a statistical comparison, including ~~five years of measurements (2006-2010)~~ all measurements of a specific Antarctic winter, from 2006 to 2010, has been pursued. This statistical comparison is meaningful as long there is a good coverage in time. Here we show the results only for 2006, being the year with the best coverage in the ground-based data set. Table 1 shows the number of occurrences for each PSC species for 2006 and the full period of 5 years. A point-to-point comparison might be approximated by a statistical analysis of the data for the shortest possible period. For 2006 the number of profiles for July and August might be sufficient to perform a month by month comparison, although with a larger uncertainty due to the smaller number of observations. In table 2 the occurrences for the PSC classes have been calculated per month for 2006. The agreement is reasonable for STS and NAT mixtures, which account for 80 to 90 % of the observed PScs. The sum of ice and enhanced NAT mixtures shows also a good agreement, while the repartition between the two classes shows some differences. This might be due to the fact that the value of $R_{NAT|ice}$, which separates the two classes might be different for CALIOP and ground-based data set. The extrapolation of $R_{NAT|ice}$ from the CALIOP dataset might be poor, because of the distance of the overpass track with respect to the location of the ground station. The maximum deviation in occurrences between CALIOP and ground-based observations is in the order of 5 %, which is acceptable, considering all possible biases.

~~Table 1 shows the relative abundance of occurrences for all four PSC classesas classified by CALIOPand ground-based lidar.Overall there is a very good agreement for ice and STS, while for the ground-based lidar there is a small shift from enhanced NAT mixturesto NAT mixtureswith respect to CALIOP. We should bear in mind, however, that this agreement, in terms of occurrences of the different PSC classes in a five year period at McMurdo, might be in part fortuitous, since significant differences might exist during the winter season and also as a function of altitude. In order to explore these possible differences, we firstly~~

We also compare the PSC occurrences as a function of altitude during the winter season, by accumulating all PSC observations ~~for the five year period (2006-2010) for each month (June through September~~each month of 2006 (July and August)

	2006		July 2006		August 2006	
PSC classes	CALIOP	gr.based	CALIOP	gr.based	CALIOP	gr.based
STS	15.5	14.6	12.8	11.1	15.1	11.4
NAT mixtures	73.6	76.0	67.5	72.2	83.1	86.6
enhanced NAT mixtures	2.3	2.4	2.8	3.8	0.8	0.2
ice	8.5	7.1	16.9	12.9	1.0	1.8
overpasses/observations	128	75	35	31	37	22

Table 2. Frequency of occurrence (in %) of PSC classes for 2006 and for July and August separately. The last line represents the number of overpasses in the McMurdo box for CALIOP and the number of ground-based observations. Observations in the 12-30 km interval have been considered. The ice-class for CALIOP includes also mountain wave ice.

between 12 and 30 km. In figure ~~??~~ 3 the vertical profiles of monthly PSC occurrence for ~~the years 2006-2010~~ 2006 are reported. Occurrence is calculated as the fraction of observations where a determined class of PSC occurs. The upper row displays the CALIOP PSC product, while the lower row shows the PSC classification obtained by applying the approximate algorithm to the ground-based data.

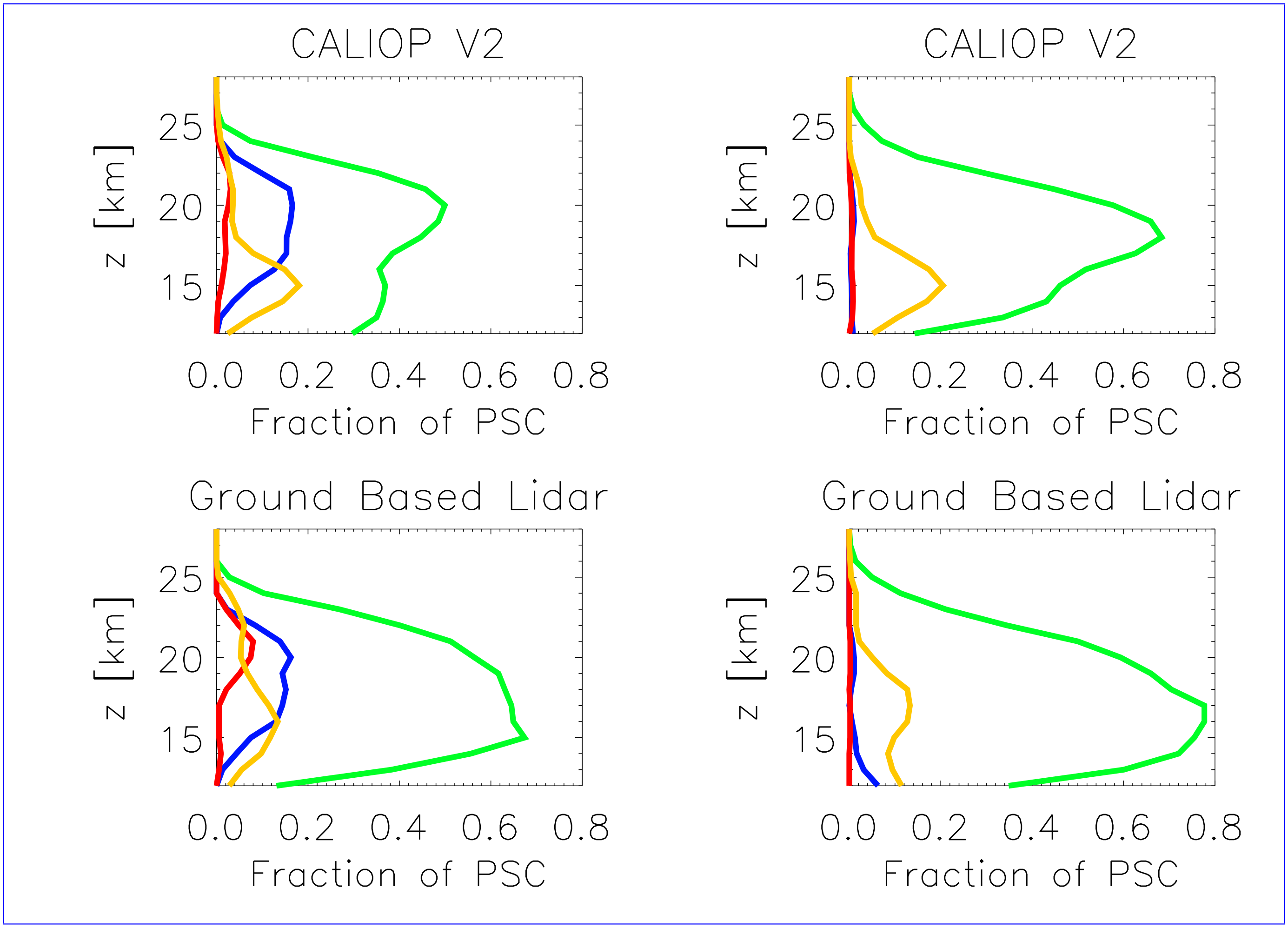


Figure 3. ~~Winter 2006-2010~~ The PSC vertical distribution for the 2006 winter as a fraction of the total observations for the four PSC classes (orange = STS, green = NAT mixtures, red = enhanced NAT mixtures, blue = ice), the ~~four~~ three columns indicate the months ~~June to September~~ July and August (from left to right). Upper row: CALIOP ~~V2~~ v2 product. Lower row: ground-based lidar at McMurdo.

The figure shows that PSCs are observed up to 25 km in ~~June,~~ July and August~~, while they are descending below 20 km in September~~. Above 25 km the number of PSC observations is negligible, both for ground-based and CALIOP observations. ~~There is a reasonable agreement between CALIOP and ground-based observations for ice, enhanced NAT mixtures and STS. NAT mixtures~~ NAT mixtures are the dominating species ~~during the winter,~~ with a slightly different altitude ditribution in July~~and August; CALIOP observations show a maximum for NAT mixtures around 20 km , while~~ ; ground-based occurrences of NAT mixtures are more frequent below 18 km with respect to CALIOP data.

The occurrences of ice clouds in July are very similar, while in August some low ice clouds appear in the ground-based data~~show a flatter distribution. Ice and enhanced~~ , but are absent in the CALIOP observations. Enhanced NAT mixtures occur mainly in July, ~~the enhanced NAT mixtures~~ and are observed between 17 and 25 km, ~~while ice persists also at lower~~

altitudes. The overall agreement is satisfactory however, considering the different observation geometries and statistics. The main discrepancies appear at lower altitudes, where the though more abundant in the ground-based lidar observes more ice in most months with respect to CALIOP, and also the observations. The vertical distribution of NAT is more shifted to lower altitudes wrt CALIOP. STS shows a good agreement in July and August.

Another way to compare the statistical distribution of PSCs as observed by both instruments is to use the temperature dependence. The temperature dependence of the occurrence of different PSC classes has been studied intensively with in-situ and remote data with the goal to to confirm hypotheses on microphysical mechanisms of PSC formation (**?**). In this context we want to use it as another tool to investigate a possible bias when comparing ground-based and satellite based observations centered on McMurdo. The temperature data base used for the data analysis of CALIOP is MERRA-2 (Modern Era Retrospective analysis for Research and Applications) which uses the GEOS-5 analysis. In a previous analysis of the McMurdo ground-based lidar data (**?**), the temperature was obtained from radiosoundings and, where these were not available, from NCEP. For the present analysis, however, we choose to use the same MERRA-2 temperature data for the ground-based data, in order to avoid a temperature bias while comparing with CALIOP data. The ice formation temperature $T_{ice\ NAT}$ has been obtained from daily values of the EOS MLS retrieved data for HNO_3 and H_2O number densities.

The probability density functions of the different species are reported in figure ?? 4 as the ratio of the occurrence of each species and the total number of observations at the specific temperature T-T_{NAT}. The total number of observations is reported as well, in arbitrary units, to indicate the variation of the number of observations with temperature. The ratio of occurrence has not been displayed when the number of observations at a specific temperature is too low to be statistically valid (less than 5 % of the total observations.

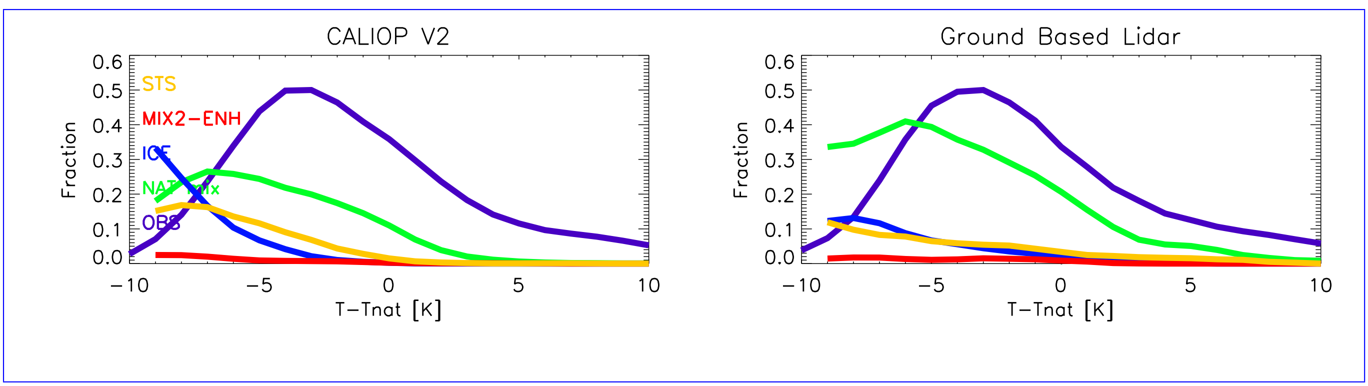


Figure 4. Fraction of PSC observations in 2006-2010 centered at McMurdo (calculated as the ratio of the number of data points for each PSC class and the total number of data points) as a function of the difference between the temperature and the equilibrium temperature for NAT. PSC classes are reported in different colors. The purple line indicates the total number of observations at a specific temperature in arbitrary units.

One can observe that the relative number of occurrences as observed by spaceborne and ground-based lidar at McMurdo vary in a similar way with the local temperature. The total number of observations have a very similar temperature distribution, which indicates that the two instruments statistically sample air masses with a similar temperature distribution. The temperature dependence of the NAT and STS PSCs is very similar, although the peak for NAT is slightly shifted to lower temperatures. The onset for ice is the same, although it the ice fraction at lower temperatures appears to be more present at lower temperatures larger for CALIOP with respect to the ground-based data. This is probably due to the fact that ice is not frequently observed around McMurdo (**??**) and that the few observations occur at different altitudes as can be seen also in figure **??**. 3.

As a conclusion, the statistical agreement between CALIOP and ground-based data is rather good above 15 km, and is biased below, probably due to a a different rejection of isolated PSC observations as performed with a different spatial coherence criterion.

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

The coupling of stratospheric chemical models with climate models has led to a new generation of models. These coupled Chemistry-Climate Models (CCMs) are CCMs have been used within the Chemistry-Climate Model Validation activity 2 (CCMVal-2) Eyring et al., 2008(**?**) and represent both stratospheric chemistry and atmospheric climate. CCMVal-2 models do not include a representation of stratospheric aerosol physics and chemistry, but use parametrizations to take into account the formation of PSCs. There are large differences among CCMs for their treatments, regarding their formation mechanisms, types, and sizes (**?**). All CCMs involved in the CCMVal-2 experiment include water-ice PSCs; all except CMAM also include $HNO_3 \cdot 3\,H_2O$ (nitric acid trihydrate , (NAT). Most CCMs furthermore treat sulfate aerosols, e.g. in the form of supercooled ternary solutions (STS) of sulfuric acid (H_2SO_4), nitric acid (HNO_3), and water (**?**).

Evaluating the ability of CCMs to reproduce ice and NAT PSCs is a key factor to interpret simulated stratospheric polar ozone changes. The comparison of space-borne PSC observations with CCM simulations requires adequate diagnostical diagnostic methods. Here we assess the ability of models to simulate PSCs taking into account diagnostics that mostly focus on microphysical factors, such as the NAT and ice surface area densities and diagnostics that are sensitive to the coupling of those with the simulation of polar vortex variability and its mean state.

3.1 Overview of the models

Here we consider 4 CCMs involved in the CCMVal-2 experiment, CAM3.5 (Community Atmosphere Model 3.5) (**?**) and WACCM (Whole-Atmosphere Chemistry-Climate Model) (**?**) both developed at NCAR, CCSRNIES (Center for Climate System Research/National Institute for Environmental Studies, Japan) (**?**), and LMDZrepro (Laboratoire de Météorologie Dynamique Zoom- REPROBUS) (**?**), developed at IPSL (Institut Pierre-Simon Laplace), and one CCM included in the Chemistry–Climate Model Initiative (CCMI), WACCM-CCMI (**??**).

Some general features such as the horizontal resolution and vertical levels have been displayed in Table 3.

CCM	Years	Horizontal resolution	vertical grid	References
CAM3.5	1991-1999	2.5^{o} $^{\circ}$ x 1.9^{o} $^{\circ}$	L26	?
CCSRNIES	1991-2005	2.8^{o} $^{\circ}$ x 2.8^{o} $^{\circ}$	L34	?
LMDZrepro	1991-2005	3.75^{o} $^{\circ}$ x 2.5^{o} $^{\circ}$	L50	?
WACCM	1995-2005	2.5^{o} $^{\circ}$ x 1.9^{o} $^{\circ}$	L66	?
WACCM-CMMI WACCM-CCMI	1960-2010	2.5^{o} $^{\circ}$ x 1.9^{o} $^{\circ}$	L66/88	??

Table 3. Horizontal resolution and number of levels for the CCMs used. The output of the models has been taken for the years indicated in the second column.

All models include water-ice PSCs as well as nitric-acid-trihydrate (NAT)NAT. They also treat sulfate aerosols in different forms, such as supercooled ternary solutions of sulfuric acid, water and nitric acid (STS) STS (CAM3.5, WACCM and CCSRNIES), or liquid aerosol (LMDZrepro).

The conditions at which PSCs condense and evaporate vary, not only for water-ice PSCs but also for NAT and STS, between CCMs (?). The simplest assumption is Most CCMVal-2 models use a thermodynamic equilibrium assumption that PSCs are formed at the saturation points of HNO_3 over NAT and H_2O over water-ice. This assumption is made in most CCMVal-2 CCMs.

The microphysical processes of condensation and evaporation of the PSCs vary among the different models. Most models use a thermodynamical equilibrium assumption that PSCs are formed at the saturation conditions for nitric acid over NAT and water over ice. CAM3.5 and WACCM allow for saturation of up to 10 times saturation (?). Table 4 illustrates how the CCMs considered here use different formation processes and sedimentation velocities.

CCM	Thermodynamics	particles	NAT/Ice Sedimentation
CAM3.5	NAT: HY; ice:EQ	NAT/ice/STS	NAT / ice but not STSradius dependent
CCSRNIES	EQ	NAT/ice/STS	NAT/ice dep. on mode radius radius dependent
LMDZrepro	EQ	NAT/iice/LA	
WACCM	NAT: HY; ice:EQ	NAT/ice/STS	NAT / ice but not STSradius dependent
WACCM-CMMI WACCM-CCMI	NAT: HY; ice:EQ	NAT/ice/STS	NAT / ice but not STSradius dependent

Table 4. Main features of simulation and of the microphysics of polar stratospheric clouds. EQ =thermodynamic equilibrium with gaseous HNO3 HNO_3 / H2SO4 H_2SO_4 / H2O H_2O assumed. HY = non-equilibrium / hysteresis considered. LA=liquid aerosol , SAD = sulphuric acid dihydrate (adapted from CCMVal-2 report (2010)).

Note that the equilibrium assumption allows to determine the total mass of condensed PSCs, and that a size distribution needs to be postulated in order to derive surface area densities (SAD). Since the sedimentation velocity depends on the size of the particles, the size distribution assumed has a significant impact on denitrification and dehydration processes through sedimentation of PSCs.

Some differences between WACCM and WACCM-CCMI should be mentioned here. While the CCMVal-2 version of WACCM simulated Southern Hemisphere winter and spring temperatures that were too cold compared with observations, in the CCMI-1 simulations this problem was addressed by introducing additional mechanical forcing of the circulation via ~~parameterized~~ parametrized gravity waves(**?**). Also the polar heterogeneous chemistry was recently updated (**?**) and further evaluated by ~~(**?**)~~**?**.

Recently Zhu and co-workers introduced a new PSC model ~~(**???**)~~ (**???**) within the CESM1 (Community Earth System Model) Whole Atmosphere Community Climate Model version 4.0 (WACCM 4.0), with Specified Dynamics (SD) coupled with the Community Aerosol and Radiation Model for Atmospheres (CARMA) model. This new model takes into account detailed microphysical processes for the formation of NAT and STS, instead of the parametrizations used in the CCMVal-2 and CCMI-1 models. An evaluation study on ~~EMAC~~ the ECHAM5/MESSy Atmosperic Chemistry (EMAC) model has been reported (**?**), using MSBM (multi-phase stratospheric box model) for the processes related to ~~PSCS~~ PSCs (**?**). The submodel MSBM uses two parametrizations for the NAT formation, one based on the heterogeneous formation on ice, the second for the homogeneous formation of NAT. The model simulations for the Arctic winter 2009/2010 and 2010/2011 showed that simulated PSC volumes are smaller than those observed and that the simulations do not produce PSCs as high as they are observed.

These models are, to our knowledge, the most significant advancements in the field of PSC representation in Global Climate Models used for ozone and climate change studies. The CARMA model is an interactive aerosol and radiation model fully coupled to the WACCM, able to simulate advection, diffusion, sedimentation, deposition, coagulation, nucleation and condensational growth of atmospheric aerosols online with the temperature, dynamics and radiation structure simulated by the GCM. This approach is completely different from the parametrizations available in the simulations we are analysing here. A full evaluation of the WACCM/CARMA models in Specified Dynamics runs with respect to CALIPSO data is available in literature (**???**) but is beyond the scope of this intercomparison, where free running simulations are used.

Here we limit our analysis to simulations produced by four models from CCMVal-2 and one model from ~~CCM1~~CCMI. One of the goals is to use different diagnostics to test the model simulations versus the CALIOP observations. Recent studies of ~~problems in simulating~~ model simulations of PSCs can be found in (**????**).

3.2 Comparison based on the PSC vertical extent

Presently, the evaluation of CCMs for what concerns stratospheric aerosol and in particular PSCs is still incomplete. The SPARC report (**?**) includes a model inter-comparison of PSC surface area densities (SAD) concluding that more work is needed to evaluate NAT and ice aerosols and that a comparison with observations is clearly needed, since currently no global data sets ~~that can be used~~ are available to evaluate these constituents~~, are available~~. The CCMVal-2 data base includes the surface area density (SAD) for NAT and ice clouds and sulphates. Here we restrict the analysis to the reference simulations

(~~REF-B2~~REF-B1), the transient set of simulations aiming at reproducing the past 1960-2006 conditions where all forcings are taken from observations (Eyring et al., 2010).)

Both model and CALIPSO observations in the latitude range [~~83°~~82°,60~~°~~°] are binned in a 3.5~~°~~°x7~~°~~° grid and on 15 vertical levels with a resolution of 1.5 km. CCMVal-2 data south of ~~83°~~82°S are excluded to fit with CALIPSO latitudinal coverage. In this study, we use two different WACCM versions, with the same PSC scheme, used within CCMVal-2 and CCMI.

To be able to compare with the CALIOP lidar observations, we have to derive the mean PSC layer vertical extent and the frequency of occurrence as a function of height and of temperature for the models from the PSC surface area density (SAD) spatial distribution. To do so, it is necessary to apply a simplified observation operator to the model output (i.e. identify the model grid points where a lidar would have observed NAT or ice clouds by defining a threshold for the SAD values produced by the models). We firstly define a vertical extent of PSCs as the sum of all layers (in km) containing a specific class of PSC. In order to study seasonal and geographical variations, we construct maps of monthly means by accumulating all observations~~from 2006 to 2010. Figure **??**~~.

Figure 5 shows maps of the monthly mean vertical extent (in km) for ice, NAT mixtures, enhanced NAT mixtures and STS PSCs as observed by CALIOP from 2006 to 2010.

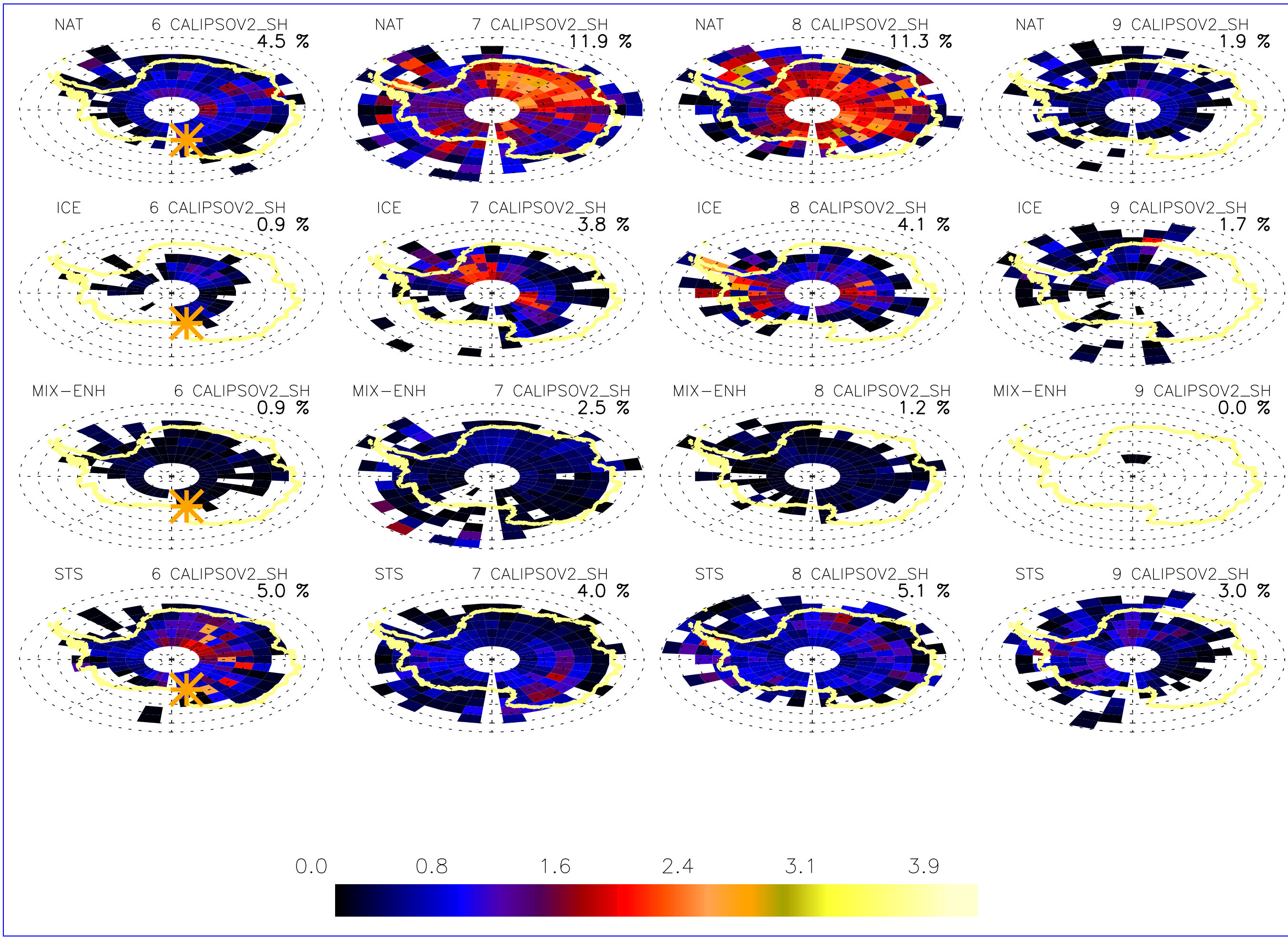


Figure 5. The vertical extent for NAT mixtures, ice, enhanced NAT mixtures and STS (from top to bottom) obtained from CALIOP observations, averaged over 5 years for June, July, August, September (left to right) is displayed. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60°S) occupied by different PSC classes for each month is reported in the top right corner. The colour scale indicates the number of km occupied by PSCs between 12 and 30 km.

The ice PSC distribution has a clear non-zonal longitudinal distribution with a maximum in the 90° W - 0° longitude sector. This appears as ~~a clear~~ an indication that mountain waves play a major role in ice cloud formation on the lee side of the Transantarctic chain~~(that crosses~~, crossing the continent as an ideal prolongation of the Antarctic Peninsula~~)~~. This has previously been reported by **?** and by **?** based on the combination of CALIPSO and COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate) GPS-RO (Global Positioning System Radio Occultation) data. The latter reports an analysis based on a single winter data set showing that mountain wave generation is a regular feature influencing ice PSC distribution. NAT-like (NAT plus enhanced NAT) PSCs have a maximum in the 0~~º~~° - 90~~º~~ °E longitude quadrant. **?** suggested that mountain waves may be responsible for the non-zonal NAT distribution that were indeed observed closer to the Transantarctic chain while **?** also consider that NAT formation can be related to the outflow of ice clouds. **?** pointed out that increased convection due to orographic triggering in the lee of the Transantarctic chain is related to the occurrence of enhanced NAT mixtures. Enhanced NAT mixtures have a minor vertical extent with respect to NAT mixtures and form in the inner vortex (where colder temperatures occur) with a zonal distribution similar to NAT mixtures. STS are observed predominantly in June, again with a clear majority in the same region of NAT and enhanced NAT mixtures formation. The McMurdo site is characterized by a majority of NAT-like PSCs (also visible in the time-series reported in figure ~~**??**~~2).

Here we compare maps of NAT and ice PSC occurrences, produced by the ~~two WACCM~~ five models, showing the geographical distribution of NAT and ice in the southern hemisphere (south of 60~~º~~°S) for the winter season, from June to September ~~,~~ (see figures 6, 7, 8 9 and 10) with CALIOP observations (figure ~~**??**~~5).

The ~~first one is sufficiently representative of the bias observed in all CCMVal-2 simulations analyzed here. The second simulation is shown to give a qualitative indication of the improvement of PSC distribution with a more reliable temperature and Antarctic stratosphere dynamics. The~~ vertical extent for the models is estimated analogously to the observations. The horizontal resolution applied to estimate the occurrence is the same among models and CALIOP data. ~~Effect of vertical resolution differences between model~~ The effect of the differences of vertical resolution among models and observations is reduced by ~~estimating~~ calculating a total aggregate vertical occurrence.

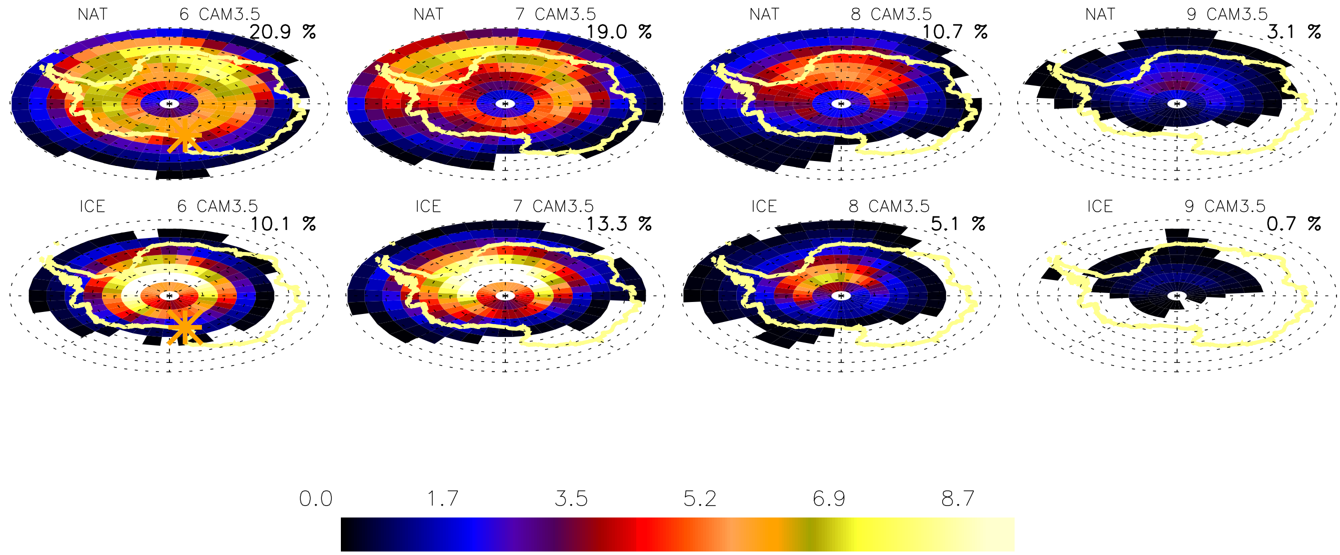


Figure 6. ~~WACCM~~ CAM3.5 PSCs vertical extent for NAT and ice, averaged over five years during the months of June, July, August, September (left to right). Please note that the color scale is different from ~~figures ?? and 10~~the other maps. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60~~o~~°S) occupied by different PSC classes for each month is reported in the top right corner.

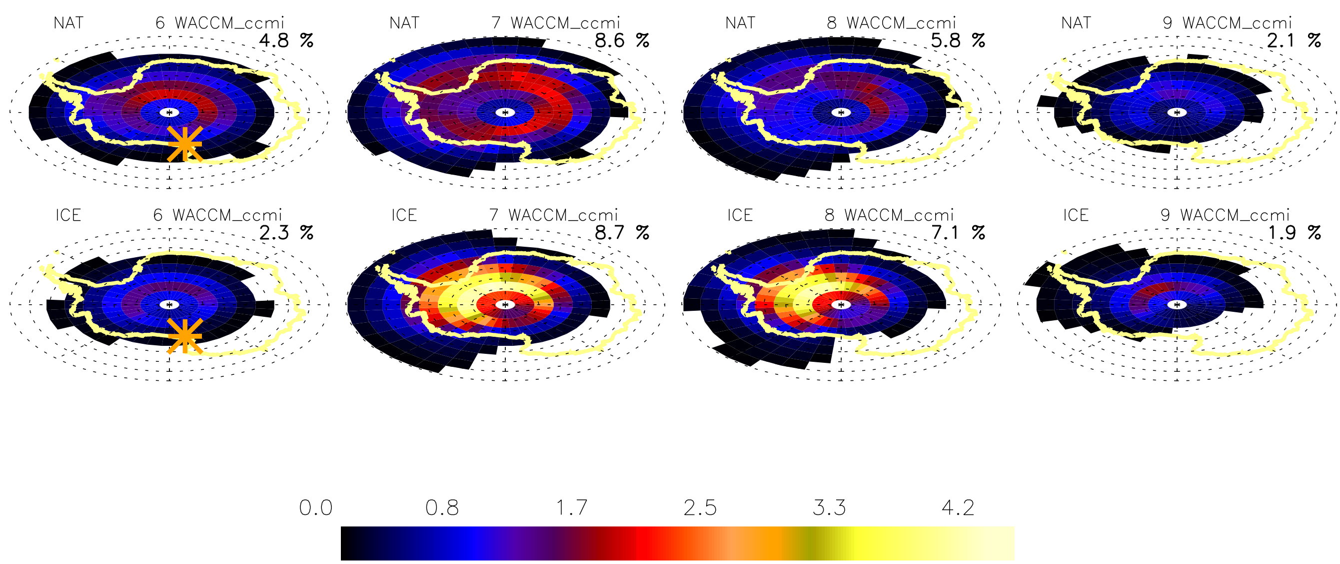


Figure 10. WACCM-CCMI PSCs vertical extent for NAT and ice, averaged over five years during the months of June, July, August, September (left to right). Please note that the color scale is different from ~~figures ?? and 9~~the other maps. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60~~o~~°S) occupied by different PSC classes for each month is reported in the top right corner.

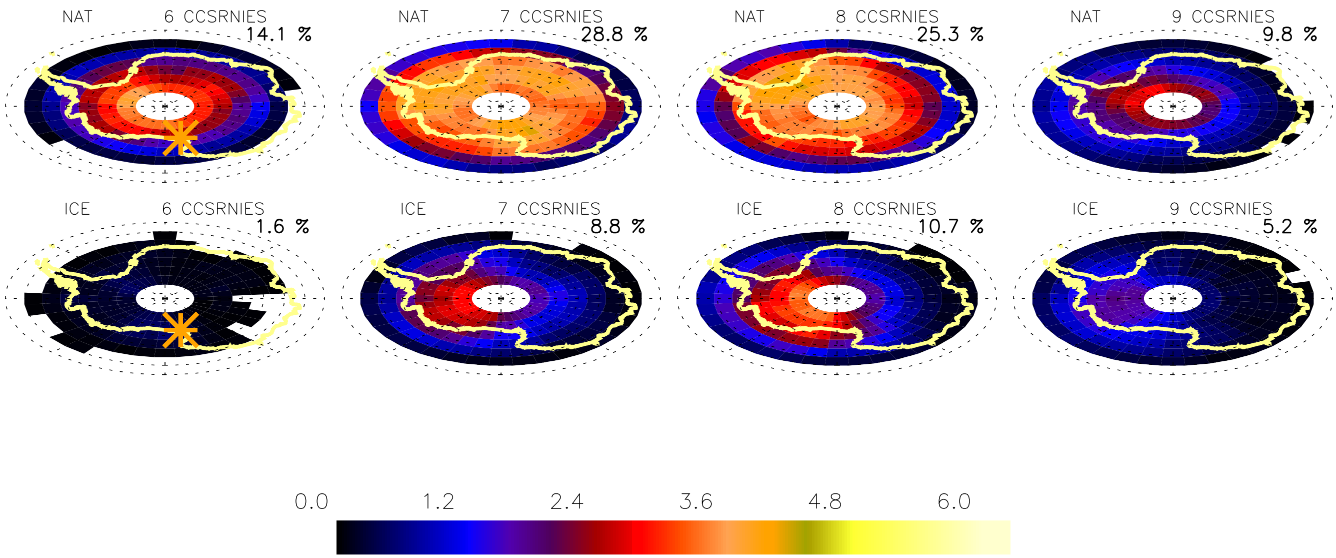


Figure 7. CCSRNIES PSCs vertical extent for NAT and ice, averaged over five years during the months of June, July, August, September (left to right). Please note that the color scale is different from the other maps. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60°S) occupied by different PSC classes for each month is reported in the top right corner.

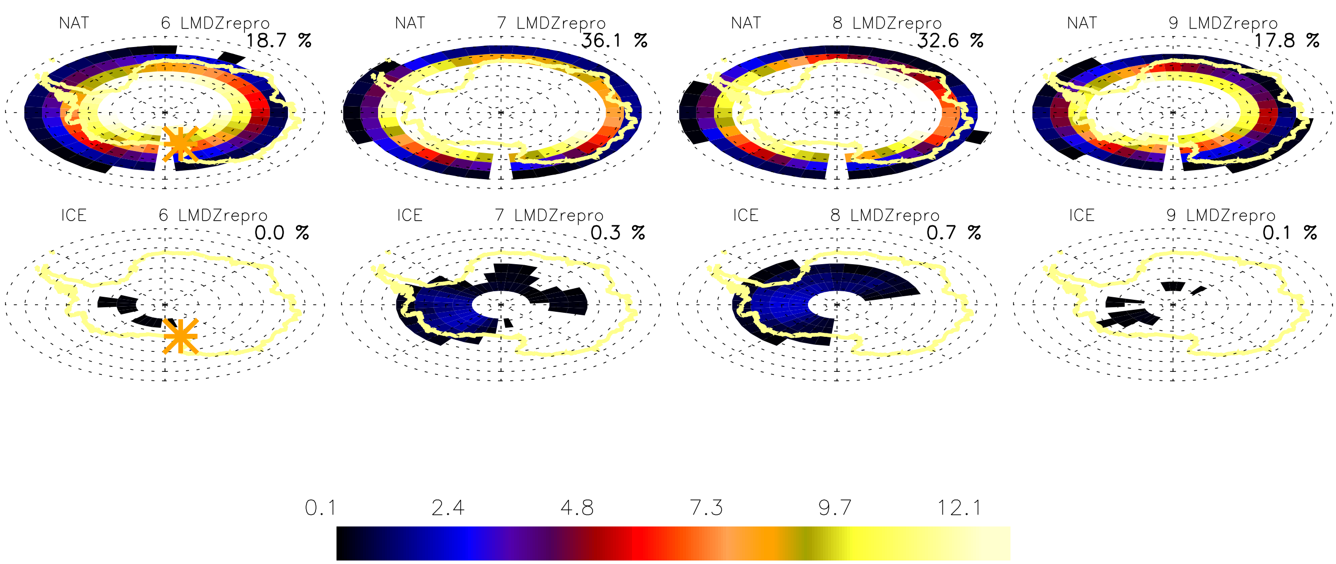


Figure 8. LMDZrepro PSCs vertical extent for NAT and ice, averaged over five years during the months of June, July, August, September (left to right). Please note that the color scale is different from the other maps. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60°S) occupied by different PSC classes for each month is reported in the top right corner.

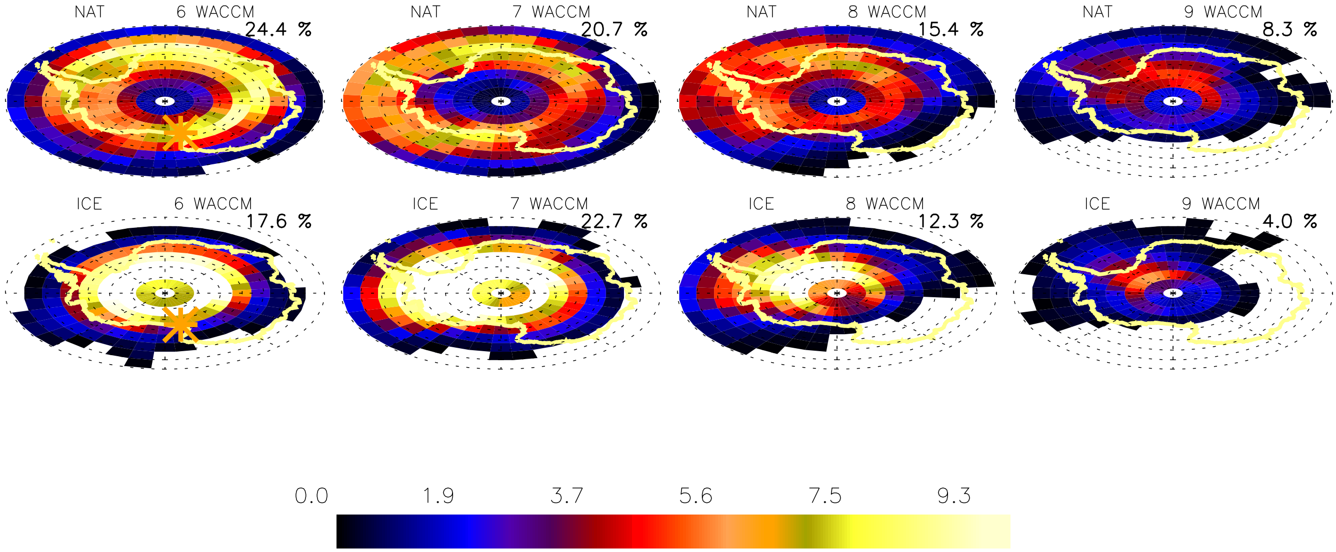


Figure 9. WACCM PSCs vertical extent for NAT and ice, averaged over five years during the months of June, July, August, September (left to right). Please note that the color scale is different from the other maps. In the left column the location of McMurdo is indicated with a yellow asterisk. The fraction of the overall air volume (between 12 and 30 km height south of 60°S) occupied by different PSC classes for each month is reported in the top right corner.

Several differences are evident. First of all the geographical distribution of the PSCs in the simulations appears to be different from the CALIOP observations, although the small numbers of observations, in particular for ice in June and September makes any comparisons with models speculative. The NAT occurrences as observed by CALIOP in July and August are mainly concentrated in East Antarctica,while WACCM predicts more NAT towards the Antarctic peninsula and the Weddell Sea. The WACCM model shows too large occurrences of both NAT and ice PSCs with respect to the the more recent WACCM-CMMI model and with respect to observations. The onset of PSCs as predicted by the models is also anticipated with respect to observations. The WACCM-CMMI model compares better with observations, for what concerns the seasonal behaviour, the occurrences and geographical distribution. The reduction of the cold bias in the WACCM-CCMI version (Doug Kinnison personal communication) may be the most relevant factor leading to PSC distribution improvement in the new model version. Table 5 reports the total PSC vertically integrated frequencies of occurrence for the five models and for CALIPSO from June to September as already indicated in figures 5,6, 7, 8, 9 and 10.

	NAT mixtures				ice			
	Jun	July	Aug	Sep	Jun	July	Aug	Sept
~~CALIPSO*~~ CALIPSO	4.5	11.9	11.3	1.9	0.9	3.8	4.1	1.7
CAM3.5	~~25.9~~ 20.9	~~26.0~~ 19.0	~~22.8~~ 10.7	3.1	10.1	~~23.1~~ 13.3	~~28.5~~ 5.1	~~16.11.9~~ 0.7
CCSRNIES	~~14.9~~ 14.1	~~36.8~~ 28.8	~~38.6~~ 25.3	~~20.2~~ 9.8	1.6	~~17.3~~ 8.8	~~23.2~~ 10.7	~~8.9~~ 5.2
LMDZrepro	~~2.0~~ 18.7	~~5.9~~ 36.1	~~14.4~~ 32.6	17.8	0.0	~~1.2~~ 0.3	~~1.9~~ 0.7	0.1
WACCM	24.4	20.7	15.4	8.3	17.6	22.7	12.3	4.0
~~WACCM-cmmi~~ WACCM-CCMI	4.8	~~5.6~~ 8.6	5.8	2.1	2.3	8.7	7.1	1.9

Table 5. Total PSC frequencies (in %) in the 13-25 km height layer for NAT and ice clouds for June-July- August-September for the observations and models. Fractions below 1% are not reported in the table. Note that CALIPSO NAT includes the enhanced NAT mixtures class

~~Table 5 reports the total PSC vertically integrated frequencies of occurrence for the five models and for CALIPSO over June to September~~The differences between the simulations obtained from the CCMs and CALIOP observations are discussed in terms of geographical distribution, onset and decline of PSCs during polar winter and total vertical extent for NAT and ice.

The CAM3.5 model overestimates NAT and ice throughout the winter and shows an early onset of PSCs in June and also an early decline in August, with respect to CALIOP observations.

Also CCSRNIES shows a too strong presence of NAT and ice, with respect to CALIOP, in particular in September, but shows a correct seasonality, with July and August being the months with the largest presence of PSCs.

The LMDZrepro model produces a correct onset and decline of the PSC formation, but shows the largest NAT frequency and the lowest ice frequency of all models.

WACCM is similar to CAM3.5, but with a larger NAT and ice frequency. The onset of PSC formation is early, as for CAM3.5. ~~Models in general significantly overestimate PSCs in June~~

The simulations produced by WACCM-CMMI follow the same trend for both NAT and ice, as observed by CALIOP, although the NAT frequency in July ad August is underestimated and the ice PSCs are overestimated in July, August and September, with respect to ~~observations, as confirmed by~~ CALIOP.

In discussing the geographical distribution of PSCs, it should be noticed that the small numbers of observations in some cases makes any comparisons among models and CALIOP difficult. The NAT occurrences as observed by CALIOP in July and August are mainly concentrated in East Antarctica, while ice is more manifest towards the 90°W direction.

The geographical distribution of NAT and ice in the CAM3.5, WACCM and WACCM-CMMI simulations is similar, with a dominant presence in the ~~CCMVal report (2010) showing a high occurrence from May onward with a maximum often occurring in June. The LMDZ model is a clear outlier with a very large underestimation of both NATand ice PSCs all throughout the season. The largest biases are found for ice PSCs that tend to be significantly overestimated~~90°W - 0° sector. CCSRNIES shows a more symmetric distribution of NAT, with a slight increase in the 90°W - 0° sector in particular in July and August. LMDZrepro shows a very strong presence of NAT, with a preference of the 90°W - 0° sector, while ice is present in very small amounts, mostly around 90°W.

As a conclusion the WACCM-CCMI model compares better with observations, for what concerns the seasonal behaviour and the occurrences. The reduction of the cold bias in the WACCM-CCMI version (Doug Kinnison personal communication) may be the most relevant factor leading to a better agreement with observations with respect to the older versions of the model (WACCM and CAM3.5).

All other models overestimate the NAT occurrences, most probably due to the cold temperature bias. Also ice is much overestimated, with the exception of LMDZ-repro which underestimates the ice occurrences with respect to CALIOP.

3.3 Comparison based on SAD

Another diagnostic method consists of comparing the ~~surface area density (SAD)~~ SAD for CCMs and CALIOP. A range of SAD values can be obtained for NAT and ice for each model. ~~Surface~~ The surface area density for the CCMVal-2 is estimated based on a semi-empirical relation between mass and mean surface areas given ~~to~~ by the model providers and reported in the CCMVal-2 report. We must be aware, however, that SAD is a derived variable and depends on the assumptions on the mean particle size for each model (as detailed in the ~~CCMVAl-2~~ CCMVal-2 report, 2010). When models predict both NAT and ice clouds, we assigned the SAD to ice if the SAD for ice is larger by a factor of 3 than the one for NAT. The SADs for CALIOP have been evaluated by using an empirical relationship derived from coincident lidar and size distributions observations (**?**). Figure 11 shows the histograms of ice and NAT values for SAD for each model together with the range of SAD reported in ~~(?)~~ **?**. The fraction is normalized to the total number of model grid points in order to identify the differences in PSC occurrence among models and between classes.

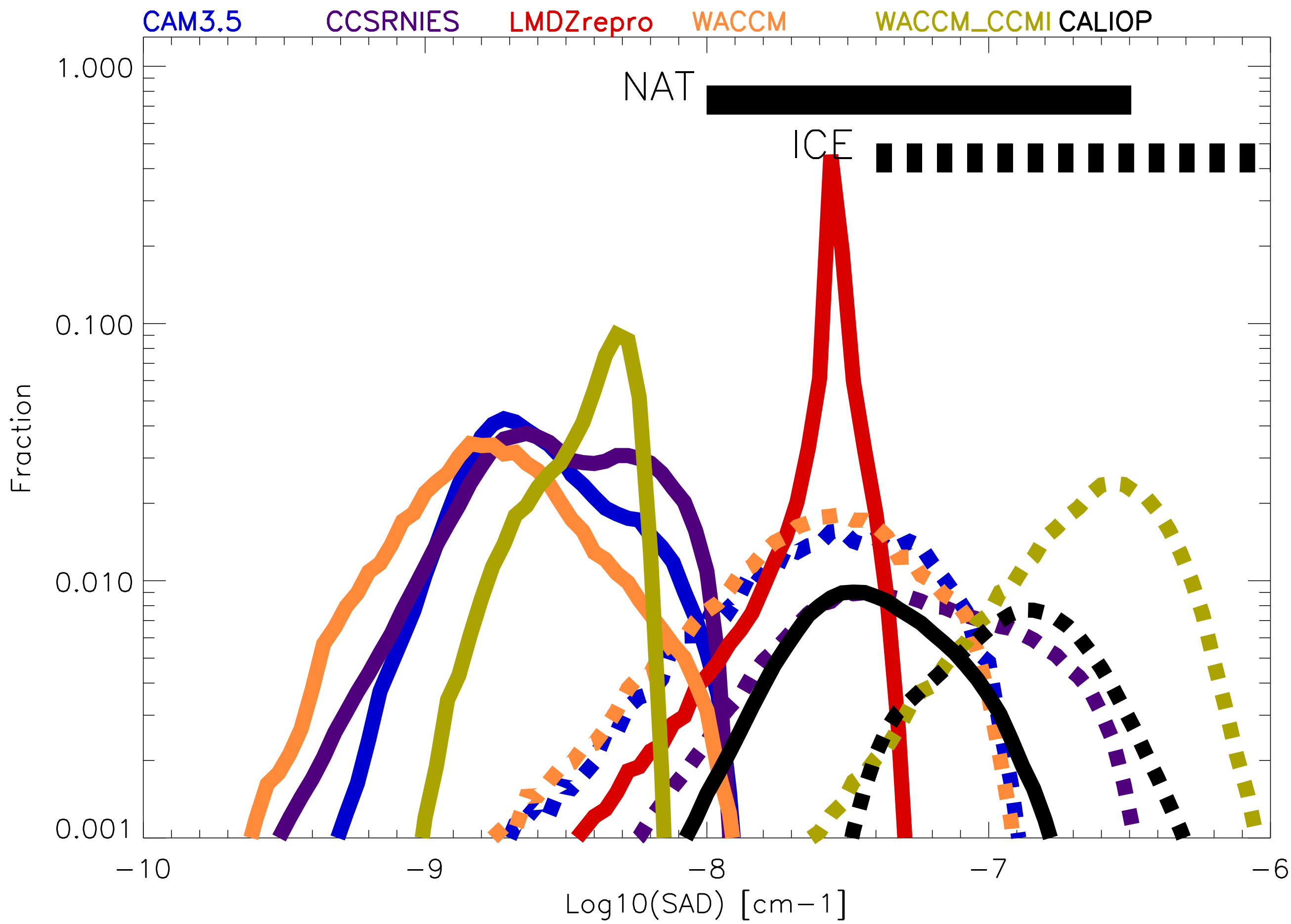


Figure 11. Histogram of the NAT (solid lines) and ice (dashed lines) ~~Surface Area Densities~~ SADs for some CCMVal models and for CALIOP are displayed. The histograms for the model data have been truncated and represent 93% of the total SAD. The straight lines at the top of the figure indicate the range of SAD values for NAT and ice "observed" by ground-based lidars and are taken from ~~(?)~~?.

We observe that for most of the models NAT PSCs have SAD ranging between $3{\cdot}10^{-10}$ and 10^{-8} cm^{-1} except for LMDZrepro that has larger SAD for NAT PSCs and is clearly an outlier. In general all models produce SADs for NAT that are smaller by one order of magnitude than the SAD calculated from CALIOP data, except for LMDZ-repro. The variability among models for the NAT SAD may be related to the assumptions made on the number of particles per cm^{-3}. The narrow peak at larger NAT SAD values for the LMDz model could be consistent with the use of much larger particle number density and smaller particle radius in the simulation. This in turn would give less irreversible denitrification processes simulated by the models with larger NAT SAD (~~CCMVAl-2~~ CCMVal-2 report, 2010, Chapter 6). Most of the models have ice PSCs in a SAD range between $2{\cdot}10^{-9}$ and 10^{-6} cm^{-1} and are generally a factor of 2-3 smaller than CALIOP values, except for the WACCM-CCMI simulations, which predict a larger value than that derived from CALIOP observations.

3.4 Comparison based on PSC occurrences

The comparison between CALIOP and CCMs can also be made by using the occurrences as a function of T-T$_{NAT}$, similarly to what has been done above for the comparison between ground-based and satellite-borne lidars above McMurdo. In figure 12 the PSC occurrences as predicted by the models and as observed by CALIPSO between 60~~°S and 83°~~°S and 82°S averaged over the 2006-2010 period have been displayed as a function of T-T$_{NAT}$, where T$_{NAT}$ has been calculated from HNO_3 and H_2O number densities. Note that the models produce only NAT and ice occurrences.

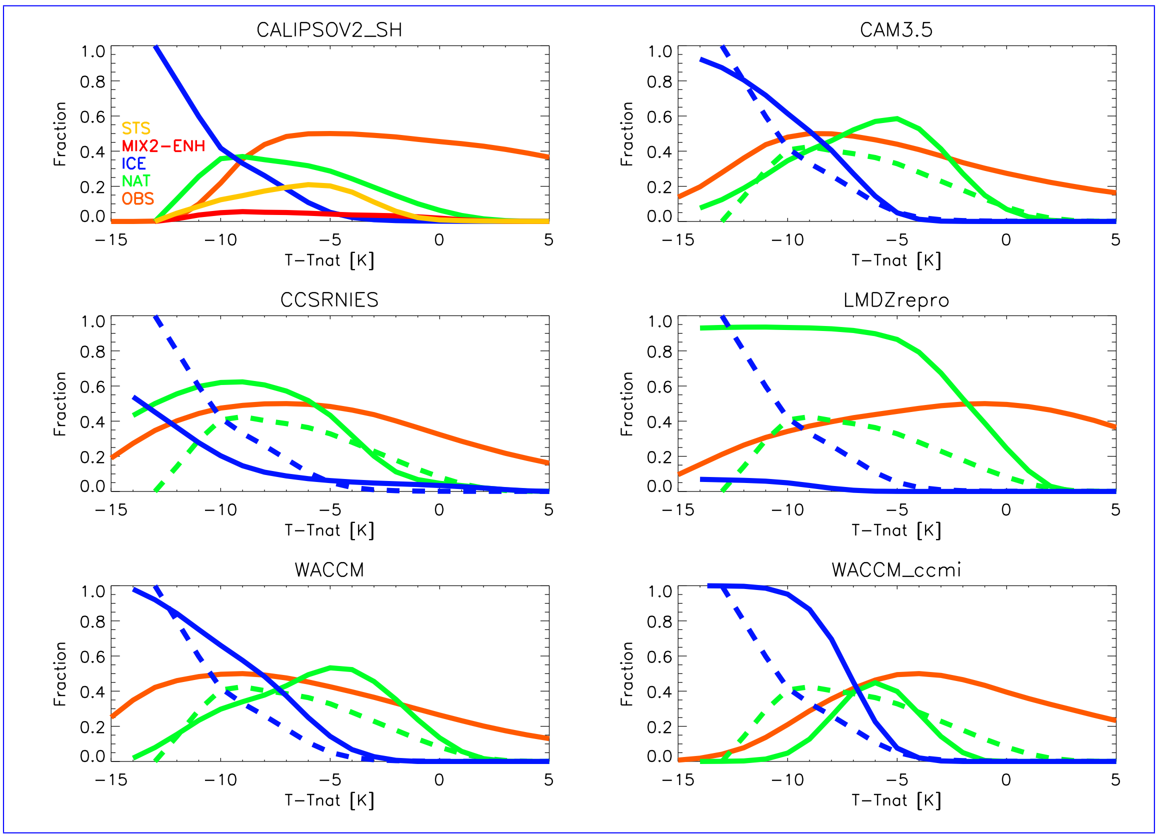


Figure 12. As Figure ~~??~~ 4 but for ~~60-83°~~60-82°S CALIOP ~~V2~~ v2 observations and CCMs data. CALIOP data are reported as dashed lines on each model as reference. The orange curve (OBS) is a histogram for the temperature distribution of all observations (CALIOP) or simulations (models).

As reported in the CCMVal-2 report, most models show a well-known cold pole bias in stratospheric temperature. The bias is in general attributed to model dynamics, as in (**?**) that identifies a lack of westward wave forcing resulting in a more intense and persistent polar vortex. A clear improvement is ~~in fact~~ obtained with an improvement in the gravity waves scheme as in (**?**), resulting in more realistic temperatures in the WACCM-CCMI simulation as described above.

The onset for NAT is similar for all models, except for WACCM-cmmi, where no NAT is observed above T_{NAT}. The onset for ice is occurring for T-T_{NAT} = -5 K for all models, except for CCSRNIES. LMDZ-repro has a too slow formation for ice and a too fast formation for NAT. The family CAM3.5, WACCM and WACCM-CCMI all have a too fast progression for ice and for NAT.

The fraction of data with different PSC classes helps in evaluating how realistic the microphysical scheme is, since this variable is normalized to the number of observations and in principle independent from the possible biases. CALIPSO sees a progressive increase of the fraction with temperature decreasing and an increase of ice PSC with T-T_{NAT} < -5 K that is close to the ice formation temperature. The total fraction of ice PSCs increases steadily from temperatures below T_{NAT} reaching 0.8 The onset of NAT is similar for all models, except for WACCM-CCMI, where NAT starts to form only below T_{NAT}. The onset of the ice formation occurs at T-T_{NAT} = -12 K . In general all models(except LMDZrepro) show a sharper -5 K for all models, except for CCSRNIES. The increase of NAT occurrences with decreasing temperatures is stronger for all models with respect to CALIOP. This is due to the fact that the models consider only the thermodynamic equilibrium conditions for the formation of PSC, and do not allow the existence of supersaturation without PSC formation. The family of models CAM3.5, WACCM and WACCM-CCMI show a faster increase of the fraction at T-T_{NAT} < -5 K with respect to CALIPSO but with different partitioning between NAT and ice. LMDZrepro shows an unrealistically low ice content, while the for other models ice is dominant ice occurrences with decreasing temperatures with respect to CALIOP. The reason is probably the same as for the NAT behaviour. LMDZ-repro evidently produces much less ice than the other models and CALIOP, and at low temperature NAT is the dominating species, while the other models and CALIOP show a dominant ice occurrence for low temperatures. The CCSRNIES model shows a slower increase of the ice occurrences with respect to CALIOP and the other models.

In general CAM3.5 and WACCM that share the same microphysical scheme have a more than satisfactory agreement, notwithstanding the cold bias that generates an excessive PSC coverage. On the other hand, WACCM-CCMI has a more realistic PSC coverage but a likely too efficient ice PSC generation due to the new scheme. So, even if the overall skills f of the model are largely improved, this kind of diagnostics (the slopes of curves in figure 12 and the "onset" PSC temperature) suggest the need to explore the ability of a single component of the model system such as the microphysical scheme.

4 Conclusions

4 Conclusions

A statistical comparison has been made of five years of proposed for PSC observations at McMurdo, obtained from ground-based and satellite-borne lidar measurements. The analysis of the ground-based data has been performed by using a detection and classification algorithm which closely follows the V2 v2 algorithm applied to CALIOP data, in order to avoid a bias due to different classification schemes. In favorable circumstances, however, a point-to-point comparison of ground-based and satellite-borne lidar observations is feasible, as can be seen in figure ??. The relative occurrences of the four PSC classes, STS, NAT mixtures, enhanced NAT mixtures and ice, averaged over five years are very similar for ground-based and CALIOP observationsResults have been shown for July and August 2006, being the months with the best temporal coverage.

A comparison of PSC occurrences as a function of time and height in 2006, shows that both datasets capture the general features of the PSC season, in terms of occurrence of each species throughout the winter. The vertical distribution and the temperature dependence of the occurrences of the different PSC classes show some discrepancies, in particular below 15 there are noticeable differences in the height distribution of NAT around 20 km. As a conclusion, the statistical agreement between CALIOP and ground-based data is rather good above 15 km, and is biased below, probably due to a a different rejection of isolated PSC observations as performed with a different spatial coherence criterionacceptable, considering the different observation geometry and other possible biases.

From CALIOP data we have derived a A set of diagnostics , useful to evaluate if biasesin Chemistry Climate Models are related to their PSCs microphysical schemeshas been proposed to compare the PSC simulations from CCMs with respect to CALIOP, with the goal to evaluate possible biases. The diagnostics are based on spatial (vertical and horizontal) SAD distribution of ice and NAT particles together with their temperature distributions. Those diagnostics are here applied to a subset of CCM simulations form CCMVal2 from CCMVal-2 and to a more recent version of WACCM from CCMI. Models fail to reproduce realistic The geographical distributions of PSCs within in the polar vortex observed by CALIOP is not well reproduced by most of the models. Moreover the model SADs are generally lower than those observed for NAT and NAT frequency is overestimated, with respect to CALIOP for all models, except for WACCM-CCMI. The onset of PSC formation is anticipated in the CAM3.5 and WACCM models, with respect to CALIOP, while CCRSNIES and LMDZrepro show a too efficient PSC production at low temperatures. While these discrepancies are evident for the older models, the more recent strong presence of NAT in June and September with respect to July and August. LMDZrepro has the largest amount of NAT and the smallest amount of ice PSCs. WACCM-CCMI shows the best agreement with CALIOP, both for onset and decline and for absolute values, although NAT is slightly underestimated in July and August and ice is overestimated in the same months. As a conclusion the WACCM-CCMI model compares better with CALIOP observations for ice and NAT, due to additional forcings applied in order to eliminate the cold temperature bias.

Acknowledgements. The authors acknowledge the financial support by PNRA in the frame work of the projects 2004/2.09 and 2009/B.08. We also acknowledge the support of the ISSI-PSC initiative project. Logistical and winter-time technical support was provided by the US National Science Foundation through NSF awards 0538679 and 0839124 to the University of Wyoming.