

Replies to Dr. John King

May 18, 2020

Performance of the higher-resolution subset of the CMIP6 ensemble

- Reviewer comment: *A somewhat surprising finding is that the higher-resolution subset of the CMIP6 ensemble appears to perform worse than the lower-resolution simulations over regions of complex orography such as the Antarctic Peninsula. This clearly warrants further investigation, but is outside the scope of the current paper.*
- Reply: We fully agree that this point is of particular interest. It will require a special analysis in a future dedicated work since this is not the focus of this article.

Representativity of the CloudSat measurements

- Reviewer comment: *The CloudSat dataset covers a relatively short time period (4 years), which raises questions about its representativity. However, the authors use a Monte Carlo approach to demonstrate that this is not a problem. With a little more work it might be possible to use this same approach to make a quantitative assessment of the statistical significance of the differences between models and observations.*
- Reply: A statistical analysis has been added in Appendix D to distinguish the regions where the comparison with the CloudSat dataset can be trusted with a good confidence level. It points out that the best reliability is on coastal regions, ice-shelves and peninsula, whereas there is a significant difference between the CloudSat snowfall means and re-analyses and model results on the interior of the continent. In addition, we carried out the same test with various time-series (whole temporal coverage and selected or random 4-years). The conclusion is that this choice has no significant influence on the result of the statistical test

for the CMIP (5 & 6) experiments whereas it plays a major role on ERA5 performances.

Other remarks

The number of stations per region has been added to the Table B1 in Appendix B. Missing acronyms have been detailed. The color palette has been modified for all figures to match the ones used by the IPCC for the CMIP (5 & 6) experiments.

Replies to Reviewer 2

May 18, 2020

Scope and format of this Brief Communication

- Reviewer comment: *While this is relevant work and a potentially interesting work, I am not sure why the authors chose to turn it into a ‘brief communication’ paper, given that I think many results are somewhat unsubstantiated and/or incomplete, and important details are missing.*
- Reply: This Brief Communication should be considered as an update to the article "Evaluation of current and projected Antarctic precipitation in CMIP5 models" by Palerme et al. (*Climate Dynamics*, 2017, doi:10.1007/s00382-016-3071-1), which provides an in-depth evaluation of CMIP5 model output against CloudSat and ERA-Interim data. With CMIP6 (and ERA5) output now available and the IPCC AR6 in preparation, this update is timely. This is why we chose the Brief Communication format: We build heavily on previously published and well-recognized work (56 citations in peer-reviewed literature, according to the Web of Science on April 30, 2020), which provides the in-depth background discussion the reviewer seems to be calling for. We clarified this at the end of the introduction in the following sentence:

Using new reanalyses and output of the most recent CMIP exercise, this work provides a brief update of the analysis by Palerme et al. (2017), which focused on CMIP5 and ERA-Interim.

Uncertainties of the CloudSat dataset over some regions

- Reviewer comment: *While the authors present CloudSat as a benchmark data set, its performance over the Antarctic interior is highly doubtful.*
- Reply: To assess the statistical significance of the difference between the observations from CloudSat and the reanalyses and the models,

we performed a Welch t-test that has been added in Appendix D. As it was mentioned, it attests that the difference between the snowfall means of the CloudSat dataset and ERA5 and the CMIP (5 & 6) ones are significantly different in some regions. In particular, comparisons have to be taken with caution on the Plateau and the Low West regions where there may be a significant underestimation of the surface snowfall means by CloudSat - due to various microphysical processes occurring at lower levels. We choose to keep the results of the CloudSat climatology as it is the only source of observations covering most of the Antarctic continent but we highlighted that there is a low confidence in these areas.

Link between temperature and precipitation

- Reviewer comment: *As the authors surely can confirm, there are many levels of complexity involved with precipitation formation, and only focusing on near-surface temperature definitely understates these complexities.*
- Reply: Of course, as the reviewer rightly presumes, we did not intend to imply that near-surface temperature was the only parameter or process that determines Antarctic precipitation rates, even though there is a strong link between temperature and precipitation changes over the Antarctic Ice Sheet on long time scales and larger spatial scales, as ample literature on this issue shows. But of course we agree that this is at best a very first order effect, with many much more subtle physical processes involved in precipitation formation. These processes complicate the detailed picture as soon as one goes beyond the very first order. To reduce the risk of such misunderstandings among our readers, we decided to 1) remove the description of the temperature station data from the main text and to move it into the supplementary material) and, more importantly, to 2) rewrite the short paragraph mentioning the reduced temperature errors:

We note that although there is no progress in the representation of large-scale mean precipitation and of its seasonality from CMIP5 to CMIP6, there is a concomitant measurable progress in the representation of surface air temperature. Regional-scale multi-model median root-mean square errors are reduced by typically 5 to 10% between these successive CMIP generations (see Figure D1 in the annex). This indicates that in spite of a clear physical link between temperature and precipitation changes on long time scales (e.g., Krinner et al., 2008;

Frieler et al., 2015), precipitation errors in current-generation AGCMs are not dominated by the first-order physical link between temperature and water vapour saturation pressure, but by errors in the representation of other processes such as atmospheric circulation and cloud microphysics.

Temperature analysis

- Reviewer comment: *The statistical analysis of near-surface temperature needs to be expanded. Only the RMSE is currently shown, but that fails to represent the mean bias ... Moreover, there is no analysis of statistical significance whatsoever ...*
- Reply: The mean bias has no sense in this analysis because of the risk or error compensation. We compute the error statistics using monthly means over the entire mean annual cycle. Therefore, one could, for example, obtain a zero annual mean bias in spite of a strong positive summer bias, compensated for by an equally strong, but opposite, winter bias. The RMSE (or, alternatively, the mean absolute bias) is therefore much more meaningful. To allow for a visual assessment of the reality of the improvement, we added error bars indicating the regional mean inter-model standard deviation of the simulated temperature errors, which shows reduced spread in CMIP6 compared to CMIP5 and, depending on the regions, substantial improvement of the mean root mean square errors. As this short analysis of temperature errors is not the focus of the paper and only very briefly mentioned in the main text, we do not think that a more detailed analysis of this aspect is warranted.

Map of the regions

- Reviewer comment: *One smaller issue is that Figure 1 should mask out regions south of 82° South. What are the white areas.*
- Reply: The map of the studied regions (Figure 1) has been modified to mask out regions south of 82° South where there is no data from CloudSat.

Brief communication : Evaluating Antarctic precipitation in ERA5 and CMIP6 against CloudSat observations

Marie-Laure Roussel¹, Florentin Lemonnier¹, Christophe Genthon¹, and Gerhard Krinner²

¹Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, Sorbonne Université / CNRS / École Normale Supérieure – PSL Research University / École Polytechnique – IPP, Paris, France

²Institut des Géosciences de l'Environnement, CNRS, Univ. Grenoble Alpes, 38000 Grenoble, France

Correspondence: (Marie-Laure Roussel (marie-laure.roussel@lmd.polytechnique.fr))

Abstract. CMIP5, CMIP6 and ERA5 ~~antaretic~~ Antarctic precipitations are evaluated against CloudSat data. At continental and regional scales, ERA5 and ~~CMIP models median~~ the median CMIP models are biased high, with insignificant improvement from CMIP5 to CMIP6 ~~despite near-surface temperature improvement~~. However, ~~less models yield outlying overestimation~~ there are less positive outliers in CMIP6. AMIP configurations perform better than ~~historical~~ the coupled ones and, surprisingly, relative errors in areas of complex topography are higher (up to 50%) in the 5 higher resolution models. The seasonal cycle is well reproduced by the median of the CMIP models, but not by ERA5. ~~There is limited progress~~ Progress from CMIP5 to CMIP6 ~~and being limited, there is~~ still room for improvement.

1 Introduction

Antarctica is the largest freshwater reservoir on Earth. Because of its sea-level equivalent of 57.9 ± 0.9 m (Morlighem et al., 2019), even minor changes of the ice sheet mass balance can have important consequences for global sea level. Apart from a small contribution from ice deposition, precipitation is by far the dominant positive term in the ice sheet mass balance, ~~at equilibrium~~. At equilibrium it is compensated for by meltwater drainage and ice discharge (e.g., Lenaerts et al., 2019) (e.g., Favier et al., 2019). Precipitation is the main source of interannual mass balance variability of the ice sheet (e.g., Boening et al., 2012) and is projected to increase in a warmer future (e.g., Krinner et al., 2008; Frieler et al., 2015) (e.g., Frieler et al., 2015). Therefore, an evaluation of the most recent CMIP6 (World Climate Research Programme (WCRP) Coupled Model Intercomparison Project phase 6) coordinated climate model simulations (Eyring et al., 2016) is timely.

Over the last decades, numerous technical developments have led to an increased number of meteorological measurements. In this study, precipitation over almost the entire Antarctic continent is analysed at a climatological time scale using a large-scale snowfall data-set that is entirely-independent from climate models, ~~and thus provides the opportunity for allowing~~ objective evaluation. The reference for snowfall rate used here is the map produced by Palerme et al. (2014) based on the CloudSat satellite radar, which provided the first 4-year surface snowfall climatology for Antarctica. It has recently been followed by its complete three-dimensional version (Lemonnier et al., 2019b). We use these satellite observations to assess the Antarctic precipitation rates simulated by the CMIP6 (~~World Climate Research Programme (WCRP) Coupled Model Intereomparison Project phase 6~~) models (Eyring et al., 2016) in diverse models in various setups, at the continental and regional spatial scales,

25 and at the annual and seasonal time scales. We further assess progress with respect to the preceding CMIP phase 5 (Taylor
et al., 2012). ERA5 reanalyses are also used and evaluated in this comparison, because outputs are often used as a reference,
particularly in less monitored areas, and because of its foreseeable use as driver for regional climate models, the continental
and climatological precipitation rates of which are strongly determined by the driving global model (e.g., Di Luca et al.,
2012). [Using new reanalyses and output of the most recent CMIP exercise, this work provides a brief update of the analysis by](#)
30 [Palerme et al. \(2017\), which focused on CMIP5 and ERA-Interim.](#)

2 Data and methods

2.1 Data

2.1.1 Snowfall : CloudSat radar

The instrument on the CloudSat satellite platform is a RADAR operating at 94GHz and looking at nadir. The Cloud Profiling
35 Radar (CPR) measures the back-scattered signal of hydrometeors. Based on micro-physical parameters (Wood et al., 2015)
and the diffusion properties of the ice particles, the snowfall rate can be computed. Constrained by the satellite orbit, this
measurement can be performed up to 82°S. Many sources of error are related to this measurement: the various assumptions
as well as the low frequency of passage of the satellite on the Antarctic induce uncertainties. (Lemonnier et al., 2019a) study
allowed to improved confidence in the CPR snowfall retrieval over peripheral areas by a comparison with *in-situ* measure-
40 ments (within maximum 25% error). In this work, we use data from the 2007-2010 Antarctic three-dimensional climatology
(Lemonnier et al., 2019b) yielding the vertical distribution of the snowfall rate with a resolution of 1° latitude and 2° longitude
- optimizing the agreement with *in-situ* observations (Souverijns et al., 2018; Palerme et al., 2014). Recently the need to take
into consideration the effect of soil echoes has been highlighted (Palerme et al., 2019), because it affects the measurement of
CPR especially in the areas of complex topography, such as mountains and fjords. Some abnormal values are ignored in this
45 dataset, but not highly impacting averages. Here we consider the radar information at the level of 1200 meters above ground
level to assess the surface snowfall rate.

2.1.2 ~~Surface air temperature: SCAR Reader data~~

~~Changes in the quality of the representation of observed precipitation rates are briefly assessed in the light of temperature biases
with respect to SCAR Reader AWS and manned station data (Turner et al., 2004). For each station and model, we identified the
50 nearest grid point and used a spatial regression (based on the neighboring grid points) of surface temperature against surface
altitude in order to correct for altitude differences between the model and the observations. SCAR Reader data were used only
when at least 10 years of observations were available, and the model output was averaged over the number of years of available
observations, centered around the mean year of these observations between 1979 and 2005 (in order to evaluate progress from
CMIP5 to CMIP6).~~

55 2.1.2 CMIP5 and CMIP6 global climate models

The Coupled Model Intercomparison Project (CMIP, Taylor et al., 2012; Eyring et al., 2016) is coordinated by the World Climate Research Programme (WCRP). Its main objective is to improve modeling and future predictions combining the natural variability of the climate system and its response to modification of the radiative forcing in coordinated experiments (see <https://es-doc.org/cmip6-experiments/>). The available model outputs taken into account in this study are listed in table A1 of the Appendix A. CMIP, which started in 1995, is currently in its 6th phase.

Here we evaluate CMIP5 and CMIP6 model output from the *amip* and *historical* experiments. ~~*amip* is the configuration of the atmospheric circulation used when~~ In the *amip* configuration, an atmospheric circulation model uses observed sea surface temperatures (SST) and sea ice (from 1979 to 2014) ~~are selected as forcing~~ as prescribed boundary conditions. The so-called *historical* simulations are coupled ocean-atmosphere experiments. In both setups, observed time-varying atmospheric composition (anthropogenic, natural and volcanic influences), solar forcing, land use etc. based on observations are prescribed. In addition, *highresSST-present*, defined in the framework of HighResMIP (Haarsma et al., 2016), is a configuration available in the CMIP6 archive similar to *amip* with forced SST, but with a higher horizontal resolution. The experiment is designed to allow evaluating the sensitivity of climate model output to spatial resolution, and to help understanding the origins of model biases. The historical CMIP6 model outputs, driven by observed boundary conditions, end in 2014, while the observational period ended in 2005 in the earlier CMIP5 exercise. We therefore preferentially restrain the CMIP5 output to before 2005, complementing them by output from the RCP8.5 scenario run until 2014 where appropriate (see figure C1), because the realized CO₂ emissions between 2006 and 2014 closely follow those of that high emission scenario (Hayhoe et al., 2017). The start of our analysis period is 1979, corresponding to the beginning of the satellite period. We use all available CMIP5 and CMIP6 ~~model~~ models, although it is well known (e.g., Masson and Knutti, 2011) that models managed by the same group or sharing a common development history yield very similar output, potentially biasing multi-model means. We preferentially use median model output, which is less sensitive to such effects, and quantify inter-model dispersion by the 25 and 75% percentiles, which are insensitive to outliers. Furthermore, although the *highresSST-present* multi-model ensemble of opportunity contains several versions of most models at low and high resolution, we do not restrain our choice to the high-resolution model versions; nevertheless, on average, the *highresSST-present* ensemble of opportunity used here has, on average, a substantially higher resolution than the *amip* and *historical* CMIP6 ensembles.

2.1.3 ERA5 reanalyses

ERA5 (Copernicus Climate Change Service (C3S), 2017) is the latest global reanalysis of the atmosphere made by the European Centre for Medium-Range Weather Forecasts (ECMWF) based on historical observation data since 1979 with the Integrated Forecasting System (IFS) model and its data assimilation system. Outputs from these reanalyses have high spatial horizontal and vertical resolutions (30 kilometers, 137 vertical levels). In this work, the monthly averages of the ERA5 reanalyses are used for the 40 years from 1979 to 2018.

2.2 Methods

For precipitation, we consider the entire Antarctic ice sheet, including ice shelves, where CloudSat satellite observations are available (i.e. north of 82°S). In order to evaluate the performances of the models ~~to reproduce~~ in reproducing the various precipitation regimes of Antarctica, we examine both regional and seasonal averages. We consider the four standard meteorological seasons that are December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). These are studied separately on the plateau (all areas above 2250 meters) and ~~or several~~ several peripheral and intermediate regions (defined by latitude and longitude, and an altitude ~~lower than~~ below 2250 meters), as there are some seasonal signature differences mostly due to the sea-ice and the circumpolar current variations during the year with significant impact on precipitation patterns on the ice sheet margin ~~(?)~~ (Palerme et al., 2017). Six regions have been selected based on latitude, longitude and altitude to distinguish main geographical patterns: Plateau, East Antarctic Coast, the Peninsula, the Filchner-Ronne and Ross Ice Shelves, and the remaining part of the West Antarctic Ice Sheet. These are shown in figure Figure 1 and described in Appendix B.

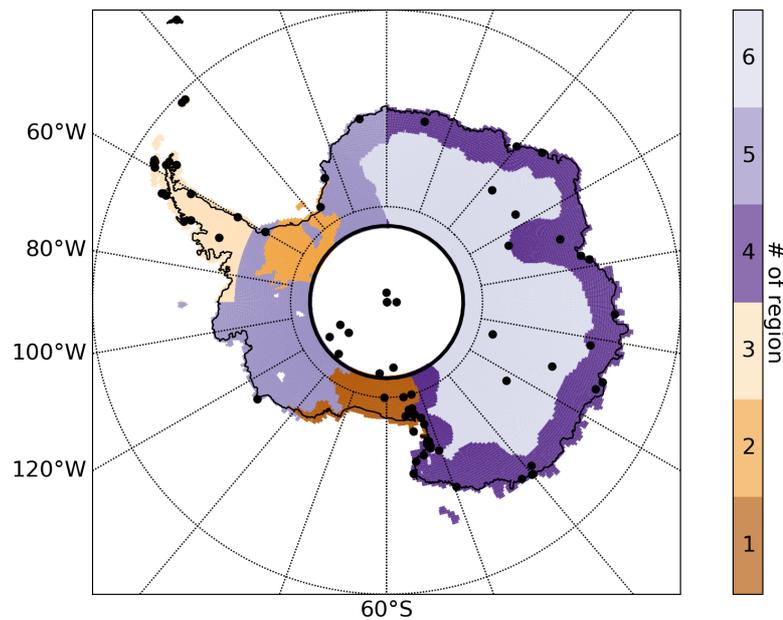


Figure 1. Map of the studied regions on the ERA5 grid. Numbers refer to the regions defined in table B1. Black dots indicate SCAR Reader temperature measurement stations (AWS and manned stations). The black line indicates the 82°S latitude circle.

To test the sensitivity of our conclusions concerning ERA and the CMIP outputs to the relatively short ~~Four-year~~ 4-year CloudSat period, we compare the CloudSat 4-year time series with multiple time periods of the same length extracted from the 40-year climatology of ERA5 and with the average of the 2007-2010 CloudSat period. We made 20 draws of 4 random

years to process the samples for the evaluation against the 2007-2010 CloudSat period. This number of 20 samples has been chosen because there is no significant difference in the results with more samples. As we will show below (see section 2.3.1), our conclusions are not very sensitive to these choices.

105 Furthermore, as historical CMIP5 outputs are only available for years up to 2005, a direct comparison from 2007 to 2010 is not possible between CMIP5 and CloudSat. Annual mean snowfall (averaged over the whole Antarctic continent north of 82°S) starting in 1979 is available until 2005 for CMIP5, until 2014 for CMIP6, and until 2018 for ERA5. Over this period, there is a slight positive mean precipitation trend in the CMIP ensembles (strongest, about 2% per decade, in the CMIP5 and CMIP6 *historical* simulations), but the variations induced by this trend over the model periods are substantially weaker than
110 the absolute differences between the model means and the CloudSat observational average. Therefore, and because our results are not particularly sensitive to the choice of model years, CMIP output is averaged over the entire respective simulation period for comparison with CloudSat.

[A more detailed statistical analysis using the Welch t-test, presented in Appendix D, demonstrates that the greatest confidence is attached to the ice-shelves, the Peninsula and the East coast \(Ross, Filchner, Peninsula and Low East regions\) when comparing the snowfall means from CloudSat to ERA5 and the CMIP \(both 5 & 6\) datasets. However, some uncertainties remain in areas of complex topography, due to sublimation of snow below the first level of CloudSat, which is likely to influence the total snowfall amount taken into account here. In the interior of the Antarctic continent, the comparison has to be treated with caution as the snowfall means from CloudSat, ERA5 and the CMIP datasets are significantly different. This may be mainly due to the CPR of CloudSat that underestimates snowfall means, as a major part of it comes from microphysical processes occurring below the first CloudSat level in this region. Therefore, the comparison is focused on the differences between CMIP5 and CMIP6, while the CloudSat results are kept for information purposes only as the single source of observation over these areas. In addition, the test points out that there is no major reduction of the reliability of the comparison between CloudSat and the CMIP experiments when the whole temporal coverage is considered \(instead of a 4-years time-series\). Conversely, there is a more significant influence of the selected years of the ERA5 dataset, which is more sensitive to the inter-annual variability.](#)

115
120

125 2.3 Results

2.3.1 Continent-wide climatological snowfall rates

Figure 2 displays the annual precipitation for the entire continent ("All") and the defined regions for CloudSat, ERA5 (both the 2007-2010 period and the average of 20 draws of four random years with associated standard deviation) and the various CMIP5 and CMIP6 ensembles. For all CMIP experiments, the ensemble median of the continental mean precipitation is above
130 the 2007-2010 CloudSat average of 186 mm water equivalent per year. Following [Palerme et al. \(2017\)](#) who compared CMIP5 models to CloudSat snowfall measurements, we have identified CMIP5 and CMIP6 models that have continent-wide mean snowfall rates within 20% of the CloudSat average value of 186 mm water equivalent per year, that is, between 150 and 223 mm per year. Not a single CMIP5 and CMIP6 model falls below this lower bound. Conversely, a substantial fraction of CMIP models, both in CMIP5 and CMIP6, exceeds the upper bound of 223 mm per year. As a result, only 58% of the CMIP6

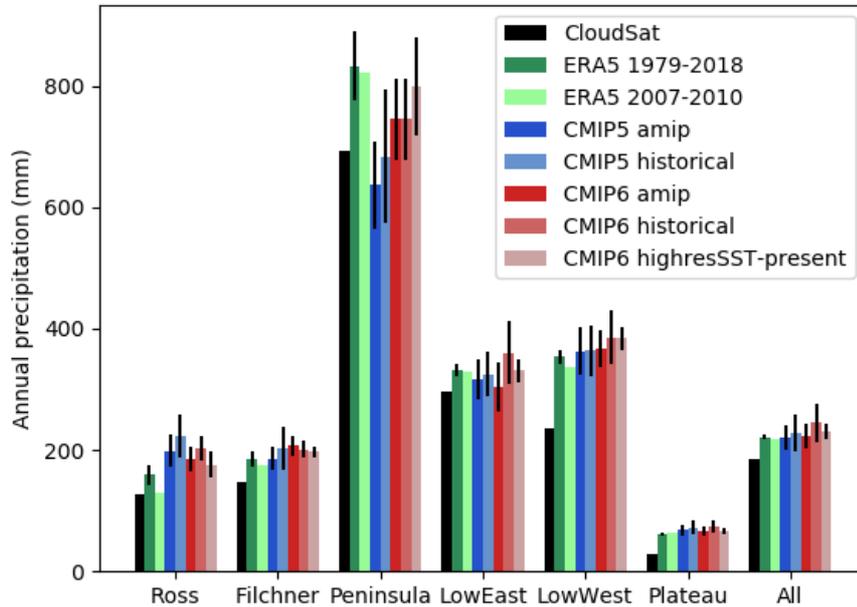


Figure 2. Average snowfall rate per region and for the entire continent north of 82°S for CloudSat (blueblack, 2007-2010), ERA5 (computed on 20 draws of 4 random years between 1979 and 2018, with standard deviation, and the 2007-2010 average), and the various CMIP ensembles. For CMIP, the ensemble medians and the 25th and 75th percentiles are indicated.

135 *amip* models fall within the $\pm 20\%$ range around the CloudSat value, and this number decreases to 38% for CMIP6 *highresSST-present*, the other ensembles lying between these extreme values. The atmosphere-only *amip* runs less frequently exceed the $\pm 20\%$ bound (56% and 58% within the 20% range for CMIP5 and CMIP6, respectively) than the coupled *historical* runs (43% and 48% within the 20% range for CMIP5 and CMIP6, respectively). We must note that the median model precipitation rate shows no improvement from CMIP5 to CMIP6; if anything, compared to CMIP5, there is even a degradation in the CMIP6
 140 median *historical* simulation with respect to CloudSat.

There is therefore a systematic high bias, exacerbated a higher spatial resolution, and no substantial improvement is obvious on the continental scale from CMIP5 to CMIP6; prescribed observed oceanic boundary conditions (SST and sea ice) in the *amip* runs lead, unsurprisingly, to more realistic simulated precipitation rates than in the corresponding coupled runs.

From CMIP5 to CMIP6, one can note, positively on the positive side, that the number of models with extreme positive precipitation biases is reduced. In the CMIP5 *historical* ensemble, for example, 4 models exceed (in one case very substantially)
 145 the maximum of the CMIP6 ensemble at 353 mm, which is almost twice the observed 2007-2010 rate.

Interestingly, ERA5 similarly exhibits a positive mean precipitation bias of about 30 mm per year, and is therefore not better, at least compared to the CloudSat climatology, than the CMIP5 and CMIP6 median models.

2.3.2 Regional averages

150 Figure 2 shows that ERA5 and the CMIP6-highresSST models, which have higher horizontal resolutions that should enable a better spatial representation of the small scale processes, particularly those induced by topography, do not exhibit reduced errors in the Peninsula region and in West Antarctica (regions named LowWest, Filchner and Ross). Relative errors with respect to the CloudSat measurement can exceed 50% in these regions, compared to the lower regions of East Antarctica (LowEast) where it is as low as a few percents.

155 Conspicuously, all CMIP ensembles and ERA5 exhibit positive biases with respect to CloudSat in all regions. The strongest relative biases are located in the Plateau region, that is, above 2250 m, where the CloudSat mean is about 29 mm of water equivalent per year, while the ERA mean for the same period is 65 mm per year, and the CMIP ensembles have even stronger biases. In most regions, the *amip* simulations exhibit lower biases than the coupled *historical* simulations in the CMIP5 and CMIP6 ensembles, as already seen for the continental mean values.

160 There is no clear overall improvement in the performance of the CMIP6 ensemble over the CMIP5 ensemble. There is degradation in some regions (for example the Peninsula) and improvement in others, such as the Plateau region, where the improvement in the *amip* configuration is modest (see also Figure 3), but important because of the large spatial extent of the East Antarctic Plateau, and on the Ross Ice Shelf. In ~~dry~~ these plateau and ice shelf areas, the *highresSST-present* runs ~~eonistently performs~~ consistently perform better than the other CMIP6 runs. This is contrary to expectations that higher spatial resolution, ~~allowing by leading to~~ a better representation of topographical effects, would ~~principally in principle~~ allow better representing precipitation rates in regions with steep topography, that is, mostly coastal areas.

2.3.3 Seasonal averages

Figure 3 displays the observed and simulated seasonal variations of precipitation separately for the high (>2250 m) and low (<2250 m) regions of the continent. The CMIP ensembles capture the weak annual cycle in the plateau regions, characterized
170 by a maximum in DJF and a minimum in SON, but, as reported above, they overestimate the average precipitation rate substantially. ERA5 does not capture this seasonality and simulated maximum precipitation rates in MAM and JJA. In the lower reaches of the continent, the CMIP ensembles and ERA5 do capture the observed seasonality, with maximum precipitation rates typically in MAM. This is very probably linked to the availability of oceanic moisture, driven by sea ice around the continent and the delayed annual temperature cycle in the Southern Ocean, and to the seasonality of meridional atmospheric
175 circulation (Gentson and Krinner, 1998).

3 Discussion and conclusion

The CloudSat precipitation climatology provides the possibility to evaluate climate models and reanalyses against model-independent satellite-derived data. By comparing ERA5 reanalysis output from multiple random 4-year periods against output

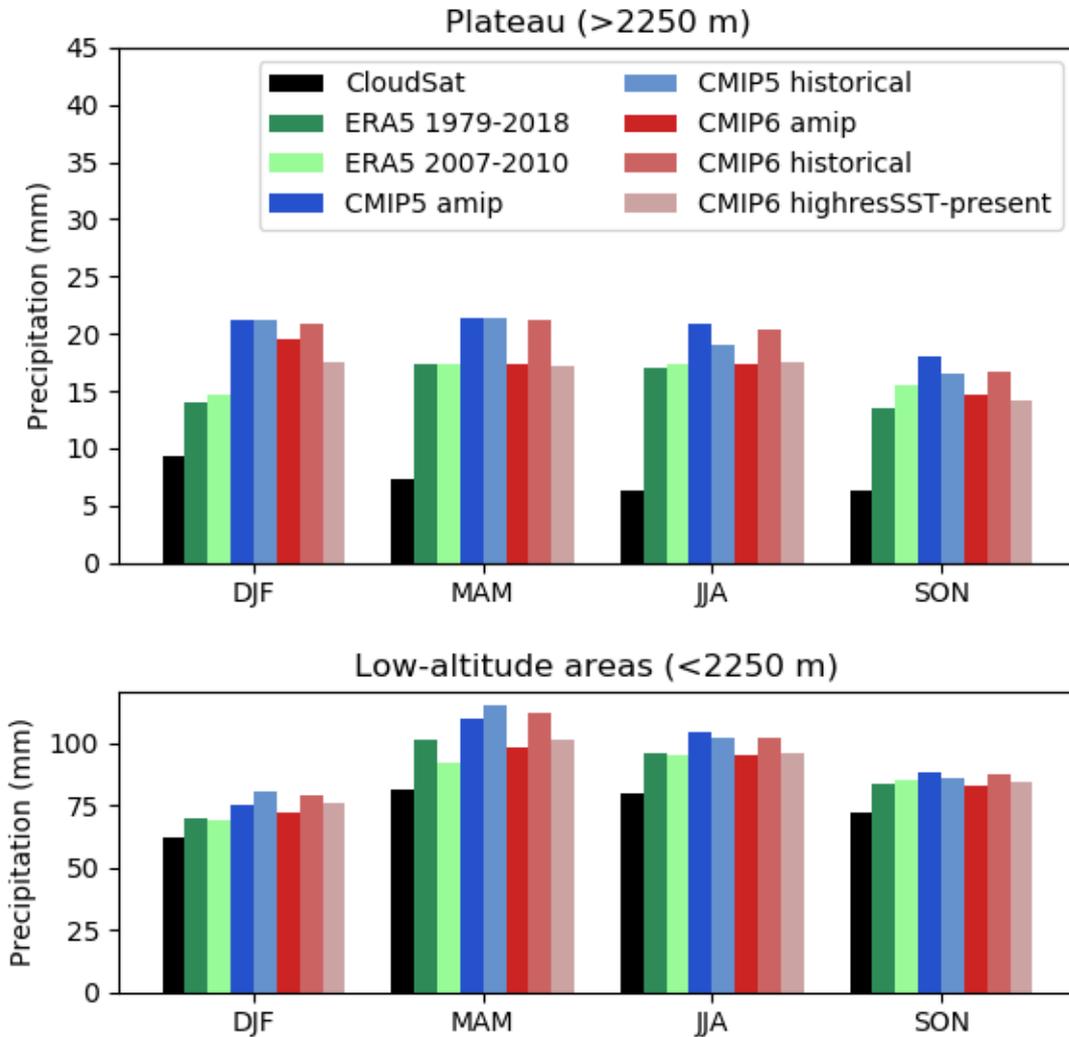


Figure 3. Seasonal averages of the observed mean 2007-2010 CloudSat, mean ERA 5 (random and 2007-2010 averages), and ensemble median long-term average CMIP5 and CMIP6 snowfall rates per season for the plateau areas (top) and the low-lying reaches of the continent (bottom).

for the 4-year observational period (2007-2010) and the satellite-derived data, we have shown that on regional scales, a 4-year period is long enough to draw robust conclusions about misfits between the models and the satellite data set.

The main results of this short study are that: 1) All CMIP model ensemble medians and ERA5 overestimate the continental mean precipitation rates; 2) The positive biases are particularly strong in the plateau regions; 3) There is no measurable improvement, in terms of continental and regional mean precipitation rates and their seasonality, from CMIP5 to CMIP6; 4) The seasonal cycle of precipitation, both on the plateau and lower (coastal) regions, is reasonably well captured by the median CMIP models; 5) Median precipitation rates tend to be better reproduced in the atmosphere-only *amip* configurations than in the coupled *historical* setups; 6) Positive precipitation biases in particular in the Peninsula region are exacerbated at higher resolution in the ~~higtresSST-present~~highresSST-present ensemble; 7) The CMIP6 ensemble suffers less than CMIP5 from outliers with very strong positive precipitation biases.

~~It is interesting to~~ We note that although there is no progress in the representation of large-scale mean precipitation and of its seasonality from CMIP5 to CMIP6, there is a concomitant ~~measurable~~ slight progress in the representation of surface air temperature. Regional-scale multi-model median root-mean square errors are reduced by typically 5 to 10% between these successive CMIP generations (see [Figure E1](#) in the annex). ~~Conversely, this indicates that an improved simulation of near-surface temperature does not necessarily lead to an improved representation of precipitation-generating processes~~ This indicates that in spite of a clear physical link between temperature and precipitation changes on long time scales (e.g., Frieler et al., 2015), precipitation errors in current-generation AGCMs are not dominated by the first-order physical link between temperature and water vapour saturation pressure, but by errors in the representation of other processes such as atmospheric circulation and cloud microphysics.

Author contributions. This research was designed by all authors. MLR and GK carried out the data analysis. MLR wrote the initial draft and all coauthors contributed to the writing.

200 *Competing interests.* The authors declare no competing interests.

Acknowledgements. This paper is a contribution to the [CALVA and APRES3](#) ~~project~~ projects supported by the French polar institute IPEV ~~and the French national research fund ANR~~. Support by the French space agency CNES for program EECLAT is also acknowledged. MLR benefits a PhD grant by Ecole Polytechnique. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5 and CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

References

- Boening, C., Lebsack, M., Landerer, F., and Stephens, G.: Snowfall-driven mass change on the East Antarctic ice sheet, *Geophysical Research Letters*, 39, 2012.
- 210 Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, dataset available from <https://cds.climate.copernicus.eu/cdsapp>, 2017.
- Di Luca, A., de Elfa, R., and Laprise, R.: Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations, *Climate Dynamics*, 38, 1229–1247, 2012.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development (Online)*, 9, 2016.
- 215 Favier, V., Krinner, G., Amory, C., Gallée, H., Beaumet, J., and Agosta, C.: Antarctica-regional climate and surface mass budget, *Current Climate Change Reports*, 3, 303–315, 2017.
- Frieler, K., Clark, P. U., He, F., Buizert, C., Reese, R., Ligtenberg, S. R., Van Den Broeke, M. R., Winkelmann, R., and Levermann, A.: Consistent evidence of increasing Antarctic accumulation with warming, *Nature Climate Change*, 5, 348, 2015.
- 220 Genthon, C. and Krinner, G.: Convergence and Disposal of Energy and Moisture on the Antarctic Polar Cap from ECMWF Reanalyses and Forecasts, *Journal of Climate*, 11, 1703–1716, [https://doi.org/10.1175/1520-0442\(1998\)011<1703:CADOEA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<1703:CADOEA>2.0.CO;2), 1998.
- Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geoscientific Model Development*, 9, 4185–4208, <https://doi.org/10.5194/gmd-9-4185-2016>, <https://www.geosci-model-dev.net/9/4185/2016/>, 2016.
- 225 Hayhoe, K., Edmonds, J., Kopp, R., LeGrande, A., Sanderson, B., Wehner, M., and Wuebbles, D.: Climate models, scenarios, and projections, in: *ClimateScience Special Report: A Sustained Assessment Activity of the U.S. Global Change Research Program*, edited by Wuebbles, D., Fahey, D., Hibbard, K., Dokken, D.J. and Stewart, B., and Maycock, T., 589, pp. 186–227, U.S. Global Change Research Program, Washington, DC, USA, <https://digitalcommons.unl.edu/usdeptcommercepub/589>, 2017.
- 230 Krinner, G., Guicherd, B., Ox, K., Genthon, C., and Magand, O.: Influence of oceanic boundary conditions in simulations of Antarctic climate and surface mass balance change during the coming century, *Journal of Climate*, 21, 938–962, 2008.
- Lemonnier, F., Madeleine, J.-B., Claud, C., Genthon, C., Duran-Alarcon, C., Palerme, C., Berne, A., Souverijns, N., Van Lipzig, N., Gorodetskaya, I., et al.: Evaluation of CloudSat snowfall rate profiles by a comparison with in situ micro-rain radar observations in East Antarctica, *Cryosphere*, 13, 943–954, 2019a.
- 235 Lemonnier, F., Madeleine, J.-B., Claud, C., Palerme, C., Genthon, C., L'Ecuyer, T., and Wood, N. B.: CloudSat-inferred vertical structure of precipitation over the Antarctic continent, <https://doi.org/10.1029/2019JD031399>, 2019b.
- Lenaerts, J. T. M., Medley, B., van den Broeke, M. R., and Wouters, B.: Observing and Modeling Ice Sheet Surface Mass Balance, *Reviews of Geophysics*, 57, 376–420, <https://doi.org/10.1029/2018RG000622>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018RG000622>, 2019.
- 240 Masson, D. and Knutti, R.: Climate model genealogy, *Geophysical Research Letters*, 38, 2011.
- Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., et al.: Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet, *Nature Geoscience*, pp. 1–6, 2019.

- Palerme, C., Kay, J., Genthon, C., L'Ecuyer, T., Wood, N., and Claud, C.: How much snow falls on the Antarctic ice sheet?, *The Cryosphere*, 8, 1577–1587, 2014.
- Palerme, C., Genthon, C., Claud, C., Kay, J. E., Wood, N. B., and L'Ecuyer, T.: Evaluation of current and projected Antarctic precipitation in CMIP5 models, *Climate dynamics*, 48, 225–239, 2017.
- Palerme, C., Claud, C., Wood, N. B., L'Ecuyer, T., and Genthon, C.: How does ground clutter affect CloudSat snowfall retrievals over ice sheets?, *IEEE Geoscience and Remote Sensing Letters*, 16, 342–346, 2019.
- 250 Souverijns, N., Gossart, A., Lhermitte, S., Gorodetskaya, I. V., Grazioli, J., Berne, A., Duran-Alarcon, C., Boudevillain, B., Genthon, C., Scarchilli, C., et al.: Evaluation of the CloudSat surface snowfall product over Antarctica using ground-based precipitation radars, *Cryosphere*, 12, 3775–3789, 2018.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, 2012.
- 255 Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., and Iagovkina, S.: The SCAR READER project: toward a high-quality database of mean Antarctic meteorological observations, *Journal of Climate*, 17, 2890–2898, 2004.
- Wood, N. B., L'Ecuyer, T. S., Heymsfield, A. J., and Stephens, G. L.: Microphysical Constraints on Millimeter-Wavelength Scattering Properties of Snow Particles, *Journal of Applied Meteorology and Climatology*, 54, 909–931, <https://doi.org/10.1175/JAMC-D-14-0137.1>, 260 <https://doi.org/10.1175/JAMC-D-14-0137.1>, 2015.

Appendix A: CMIP5 & CMIP6 version models

CMIP5	
<i>amip</i>	ACCESS1-0, ACCESS1-3, bcc-csm1-1-m, bcc-csm1-1, BNU-ESM CCSM4, CESM1-CAM5, CMCC-CM, CNRM-CM5, CSIRO-Mk3-6-0, EC-EARTH, FGOALS-g2, FGOALS-s2, GFDL-CM3, GISS-E2-R inmcm4, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC-ESM MIROC5, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, NorESM1-M
<i>historical</i>	ACCESS1-0, ACCESS1-3, bcc-csm1-1-m, bcc-csm1-1, BNU-ESM, CanCM4, CanESM2, CCSM4, CESM1-BGC, CESM1-CAM5, CESM1-FASTCHEM, CESM1-WACCM, CMCC-CESM, CMCC-CMS, CNRM-CM5-2, CNRM-CM5, CSIRO-Mk3-6-0, CSIRO-Mk3L-1-2, EC-EARTH, FGOALS-g2, FGOALS-s2, FIO-ESM, GFDL-CM2p1, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H-CC, GISS-E2-H, GISS-E2-R-CC, GISS-E2-R, HadCM3, HadGEM2-AO, HadGEM2-CC, HadGEM2-ES, inmcm4, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC-ESM-CHEM MIROC-ESM, MIROC4h, MIROC5, MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P, MRI-CGCM3, MRI-ESM1, NorESM1-M, NorESM1-ME
CMIP6	
<i>amip</i>	BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0, CanESM5, CESM2-WACCM CESM2, CNRM-CM6-1, CNRM-ESM2-1, E3SM-1-0, EC-Earth3-Veg, EC-Earth3, FGOALS-f3-L, GFDL-CM4, GFDL-ESM4, GISS-E2-1-G, HadGEM3-GC31-LL, INM-CM5-0, IPSL-CM6A-LR, MIROC6, MRI-ESM2-0, NESM3, NorCPM1, NorESM2-LM, SAM0-UNICON, UKESM1-0-LL
<i>historical</i>	BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0, CanESM5, CESM2-WACCM, CESM2, CNRM-CM6-1-HR, CNRM-CM6-1, CNRM-ESM2-1, E3SM-1-0, EC-Earth3-Veg, EC-Earth3, FGOALS-g3, GFDL-CM4, GFDL-ESM4, IPSL-CM6A-LR, GISS-E2-1-G, GISS-E2-1-H, HadGEM3-GC31-LL, INM-CM4-8, MIROC-ES2L, MIROC6, MRI-ESM2-0, NESM3, NorCPM1, NorESM2-LM, SAM0-UNICON, UKESM1-0-LL
<i>highresSST</i>	CMCC-CM2-HR4, CMCC-CM2-VHR4, CNRM-CM6-1-HR, CNRM-CM6-1, ECMWF-IFS-HR, ECMWF-IFS-LR, FGOALS-f3-H, FGOALS-f3-L, GFDL-CM4C192, HadGEM3-GC31-HM, HadGEM3-GC31-LM, HadGEM3-GC31-MM, INM-CM5-H, IPSL-CM6A-ATM-HR, IPSL-CM6A-LR, MPI-ESM1-2-HR, MPI-ESM1-2-XR, MRI-AGCM3-2-H, MRI-AGCM3-2-S, NICAM16-7S, NICAM16-8S

Table A1. CMIP5 and CMIP6 models considered in this study

Appendix B: Geographical delimitations for the regional analysis

Regions	1 : Ross	2 : Filchner	3 : Peninsula	4 : LowEast	5 : LowWest	6 : Plateau
Latitude	-99°,-75°	-99°,-76°	-74°,-59°	-99°,-59°	-99°,-59°	-99°,-59°
Longitude	150°,240°	270°,340°	270°,320°	0°,180°	180°,360°	0°,360°
Altitude	< 300m	< 300m	-	< 2250m	< 2250m	> 2250m
Number of stations	15	1	19	39	30	9

Table B1. Selection criteria applied to define the studied regions.

Appendix C: CMIP5, CMIP6, ERA5 and CloudSat time series of mean annual surface precipitation rates

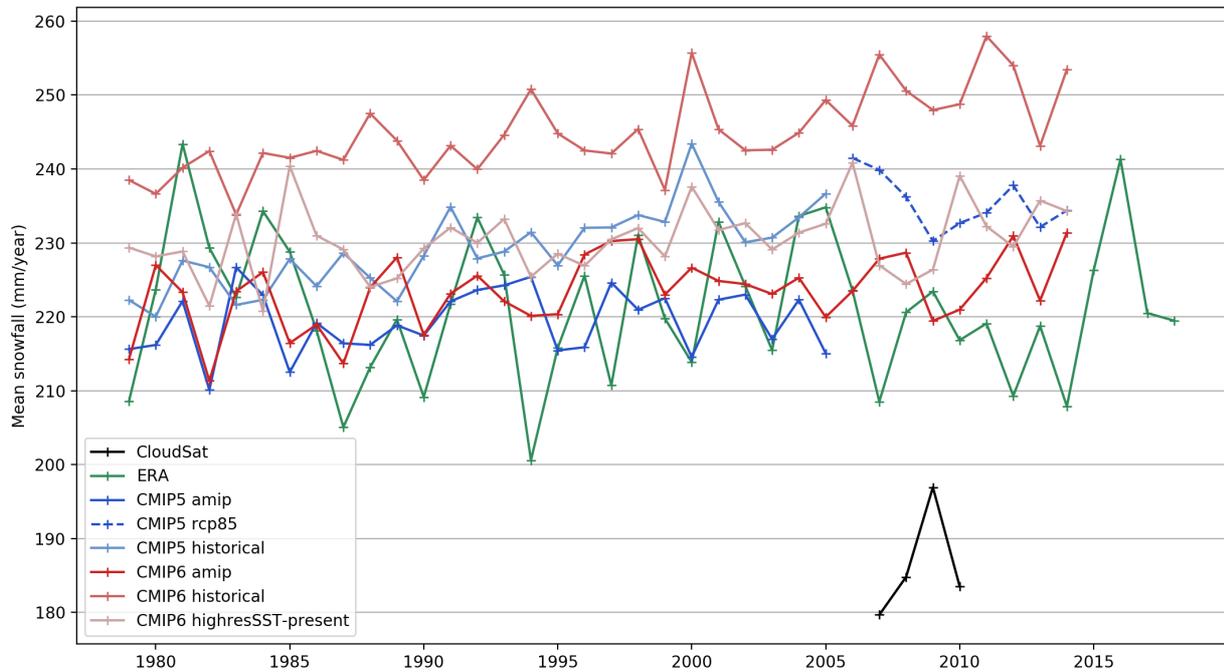


Figure C1. 1979 to 2018 ~~time-series~~ ~~time-series~~ of mean annual snowfall rate over the entire Antarctic continent. The ~~average~~ ~~median~~ values for all the models of each CMIP (5 & 6) ~~version~~ ~~are plotted~~ and the ~~standard deviations~~ ~~are represented by the colored bands~~. ~~Blue and green lines are respectively~~ ~~mean values for~~ CloudSat and ERA5 ~~annual mean values~~ ~~are plotted~~.

Appendix D: Statistical significance of the 4-year CloudSat snowfall climatology

265 Snowfall means of the CloudSat dataset have been compared to ERA5 ones and to the median of each of the CMIP (5 & 6)
experiments with a Welch's t-test. This statistical tool is used to assess the difference between the mean of two independent
groups that have normal distributions with unequal variances. Those assumptions have been verified by all the datasets at
the continent scale and for the whole year. The hypothesis tested (null-hypothesis) is that the means of the two groups are
equal. We performed the two-tailed test and analysed the p-value (rounded down to one hundredth) - in comparison with the
270 0.05 threshold - for each region (defined in the Table B1) and each season. Three methods have been carried out to chose the
time-serie of the datasets used in this test, illustrated by the Figure D1 :

- whole temporal coverage : the entire time-serie available (respectively 1979-2018 for ERA5, 1979-2005 for CMIP5 and
1979-2015 for CMIP6)
- correct years : the exact years of the CloudSat climatology (only for ERA5 and the three CMIP6 experiments)
- 275 - random years : draws of 4 non-consecutive years (tested from 1 to 10000 draws and limited to 20 draws in the main
work)

p-values are generally higher for the first 4 regions (Ross, Filchner, Peninsula and Low East) for any season and for both
ERA5 and CMIP (5 & 6). The choice of the time-serie has no major impact on the result of the test for each of the CMIP
experiments. On the contrary ERA5 and CloudSat are in a much better agreement when considering a 4-years time-serie. The
280 Figure D2 presents the detailed results considering the whole time coverage for each of the CMIP experiment and for the three
time-series considered for ERA5. The red color indicates when the null-hypothesis has to be rejected and the blue color when
it can not be rejected. One can note that the snowfall mean is significantly different at the continent scale for any season, as
well as on the plateau and the west coast. Higher p-values are mainly on ice-shelfes, peninsula and the east coast.

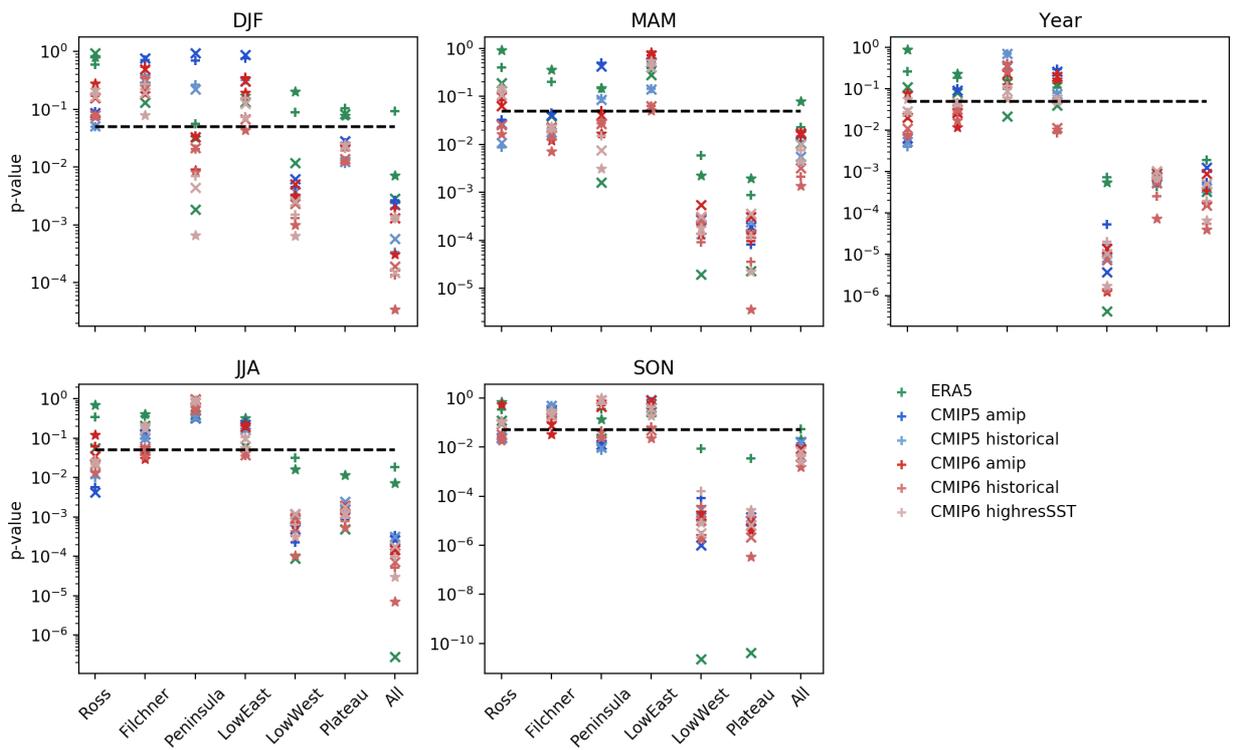


Figure D1. p-values of the Welch t-test comparing the snowfall means from CloudSat to ERA5 and each of the CMIP (5 & 6) experiments for various seasons and the whole year on each of the regions considered. The black dashed line shows the 0.05 threshold to decide whether the hypothesis is rejected or not. 'x' crosses indicate results considering the whole temporal coverage, '+' crosses the random 4-years and stars the correct 4-years of the CloudSat climatology.

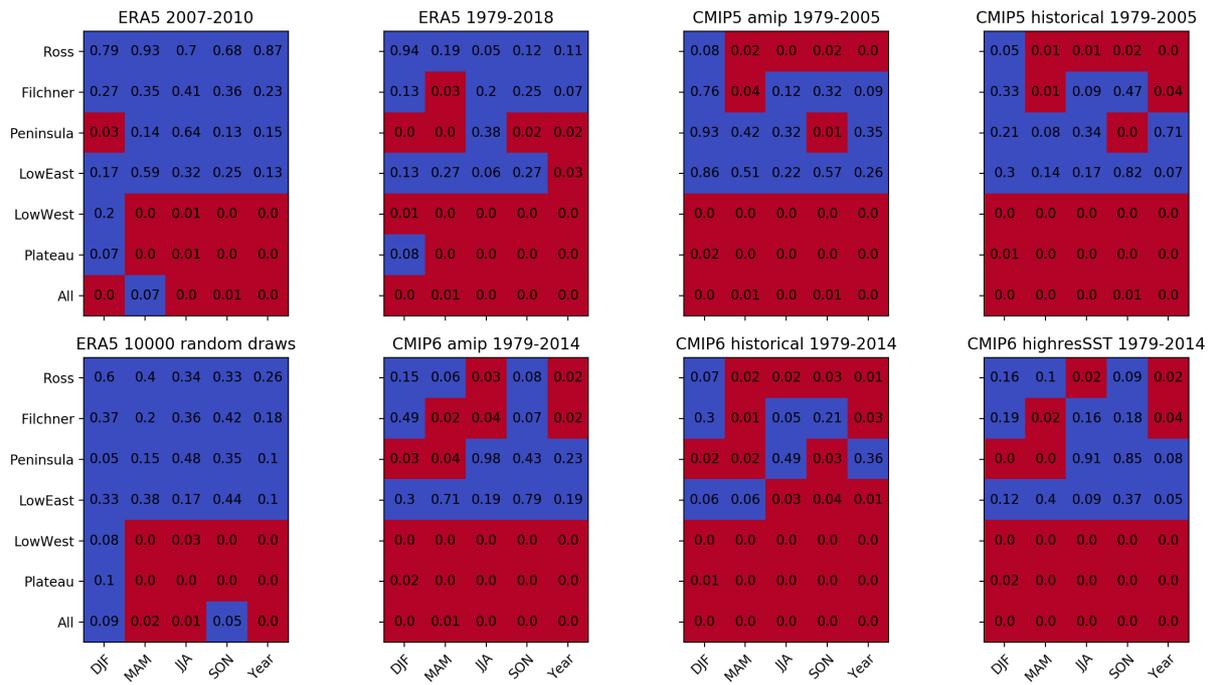


Figure D2. p-values of the Welch t-test comparing the snowfall means from CloudSat to ERA5 and each of the CMIP (5 & 6) experiments for various seasons and the whole year on each of the regions considered, considering the whole temporal coverage. Blue color indicates that the p-value is greater than the 0.05 threshold (null-hypothesis can not be rejected), red color indicates that the p-value is lower than the threshold (null-hypothesis rejected).

Appendix E: CMIP5 & CMIP6 surface air temperature comparison to SCAR Reader stations

285 Changes in the quality of the representation of observed precipitation rates are briefly assessed in the light of temperature
biases with respect to SCAR READER (REference Antarctic Dataset for Environmental Research) AWS (Automatic Weather
Station) and manned station data (Turner et al., 2004). For each station and model, we identified the nearest grid point and used
a spatial regression (based on the neighboring grid points) of surface temperature against surface altitude in order to correct
for altitude differences between the model and the observations. SCAR Reader data were used only when at least 10 years of
290 observations were available, and the model output was averaged over the number of years of available observations, centered
around the mean year of these observations between 1979 and 2005 (in order to evaluate progress from CMIP5 to CMIP6).

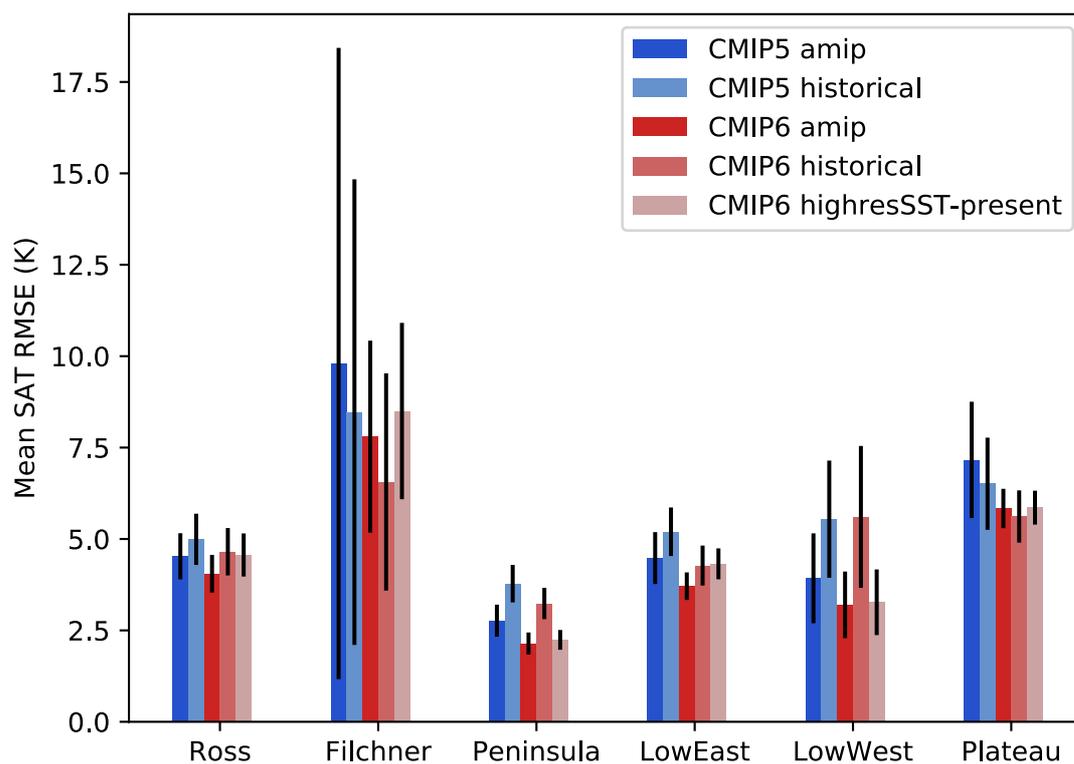


Figure E1. Multi-model median-mean of the multi-station mean root-mean square error (RMSE, in K) of simulated monthly surface air temperatures against SCAR Reader stations (AWS and manned), for the different regions. The regional mean inter-model standard deviation is shown as black error bars, indicating, for some regions, reduced inter-model spread in CMIP6 compared to CMIP5 and modest overall improvement.