https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

Open Access
Earth System
Science
Data
Discussions

Reconstruction of daily snowfall accumulation at 5.5km resolution over Dronning Maud Land, Antarctica, from 1850 to 2014 using an analog-based downscaling technique

Nicolas Ghilain[1], Stéphane Vannitsem[1], Quentin Dalaiden[2], Hugues Goosse[2], Lesley De Cruz[1], and Wenguang Wei[3]

[1]Royal Meteorological Institute of Belgium, Brussels, Belgium
[2]Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[3]Key Laboratory of Regional Climate–Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Correspondence: Stéphane Vannitsem (stephane.vannitsem@meteo.be)

precip rate

Abstract. The surface mass balance (SMB) over the Antarctic Ice Sheet displays large temporal and spatial variations. Due to the complex Antarctic topography, modelling the climate at high resolution is crucial to accurately represent the dynamics of SMB. While ice core records provide a means to infer the SMB over centuries, the view is very spatially constrained. General circulation models (GCMs) estimate its spatial distribution over centuries, but with a resolution that is too coarse to capture the large variations due to local orographic effects. We have therefore explored a methodology to statistically downscale snowfall accumulation, the primary driver of SMB, from climate model historical simulations (1850-present day) over the coastal region of Dronning Maud Land. An analog method is set up over a period of 30 years with the ERA-Interim and ERA5 reanalyses (1979-2010 AD) and associated with snowfall daily accumulation forecasts from the Regional Atmospheric Climate Model (RACMO2.3) at 5.5 km spatial resolution over Dronning Maud in East Antarctica. The same method is then applied to the period from 1850 to present day using an ensemble of ten members from the CESM2 model. This method ~~enables to~~ derives a spatial distribution of the accumulation of snowfall, the principal driver of the SMB variability over the region. A new dataset of daily and yearly snowfall accumulation based on this methodology is presented in this paper (MASS2ANT dataset, http://doi.org:10.5281/zenodo.4287517, Ghilain et al. (2021)), along with comparisons with ice core data and available spatial reconstructions. It offers a more detailed spatio-temporal view of the changes over the past 150 years compared to other available datasets, allowing a possible connection with the ice core records, and provides information that may be useful in identifying the large-scale patterns associated to the local precipitation conditions and their changes over the past century.

1 Introduction

In the context of the global climate warming, polar ice sheets have increasingly gained attention, due to the threat of a massive sea level rise at the global scale (Garbe et al. , 2020). While the Greenland Ice sheet is eroding at an increasing speed both from the base and the surface (Lenaerts et al. , 2019), the Antarctic Ice Sheet is sometimes viewed as subject to a mitigation

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

wording seems to tie iceshelf melt to sea level rise – which is not the case

mechanism to the observed melting of the ice shelf through an increased coastal precipitation due to a higher atmospheric humidity (Shepherd et al. , 2018; Krinner et al. , 2014; Agosta et al. , 2013; Medley and Thomas , 2019). This effect has been documented regionally and has been evidenced to be large over the Antarctic Peninsula (Thomas et al. , 2008) and over some coastal areas (Frezzoti et al. , 2013), but is less clear for the rest of the Antarctic Ice Sheet (Monaghan et al. , 2006; van den Berg et al. , 2005; Lenaerts et al. , 2012). However, the study of other coastal regions of Antarctica has revealed methodological limitations in estimating the surface mass balance, especially because the tools and observations available have a too low spatial resolution in regions where snowfall is orography induced and spatial variations are high (Eisen et al. , 2008). Dronning Maud Land (DML, 20°W – 45°E, map on Figure 2) is one of the sectors for which such uncertainty exists. Like any other coastal sector of Antarctica, the gradient of snowfall accumulation follows the large scale topography, with maxima over the coast and minima in the interior (Rotschky et al. , 2007), but it is highly variable at the scale of a few kilometers because of the ice rises and rumples that punctuate the coast (Lenaerts et al. , 2014). The available reconstitutions of surface mass balance distribution over the last 200 years are deduced from the interpolation of the quality controlled SMB estimations from ice cores drilled at several locations (Favier et al. , 2013), assuming an averaged (Rotschky et al. , 2007) or a time-dependent spatial distribution (Medley and Thomas , 2019).

again are they discussing the whole icesheet or just iceshelves

again is the study over ice shelves?

General circulation models provide another source of information for the SMB over the last millenium, but, as the local topography is not well resolved by these models, the estimation of snowfall accumulation over the coast is inaccurate (Tetzner et al. , 2019). ~~On another side,~~ Alternatively regional climate models (RCMs) have been adapted over the polar regions (van Wessem et al. , 2016; Agosta et al. , 2019) allowing the modelling of more detailed processes determining SMB at a scale of a few kilometers (Agosta et al. , 2019; Mottram et al. , 2020), but with the limitation of the period considered, generally not before the satellite era (past 40 years). The time frame of 40 years, though of relevance for climate analysis, seems too short to determine trends in snowfall accumulation potentially related to climate warming due to the uncertainty ~~on~~ in the process at play, the large interannual variability in the region and the unknown history of the local SMB.

Indications of a recent increase of SMB in Dronning Maud Land have been found at some locations (Lenaerts et al. , 2013; Schlosser et al. , 2016; Medley et al. , 2018; Philippe et al. , 2016), but a stationary or decreasing trend has been found elsewhere (Thomas et al. , 2017; Vega et al. , 2016; Altnau et al. , 2015; Schlosser et al. , 2014). Most of the coastal ice core drilling sites are situated on top of an ice rise, a choice that may be due to the technical difficulties posed by the lateral flow of the ~~icefields~~ ice shelves and the low volume of snow accumulated, ~~both~~ preventing ~~to~~ accurate~~ly~~ date ing of the strata (Matsuoka et al. , 2015; Goel et al. , 2020). ~~But~~ Moreover, the drilling sites may not be representative of larger areas (Cavitte et al. , 2020; Kausch et al. , 2020). Extrapolating the findings from ice core sites to large areas may thus be difficult.

I don't think that is correct

Statistical methods can be set up to emulate the precipitation field relating the spatial pattern of precipitation given by a RCM to large scale dynamics of the global climate model. We have explored such a methodology to statistically downscale the dominant SMB component over the area , the accumulated precipitation, from the climate model historical simulations (1850-present day). A method based on the search for analogs has been set up over a period of 30 years with the ERA-Interim reanalysis (1979-2010 AD) and associated with snowfall from the Regional Atmospheric Climate Model (RACMO2.3) at 5.5 km spatial resolution over Dronning Maud Land in East Antarctica. The method is then applied to the period from 1850 to present

why refer to SMB when you calculate precip. Precip alone is interesting and ~~so~~ arguing you are improving SMB weakens your case day using an ensemble of 10 simulations from the CESM2 model. This method ~~enables to~~ derive 10 time-dependent spatial distribution of the accumulated snowfall over Dronning Maud Land at 5.5 km resolution from CESM2 ensemble members. In addition, we provide the time series of the 10 first principal components of each of the 4 variables used that can be used to characterize the regional synoptic situations. We present first the method and the data used, then the structure of the database, and finally the validation with ice core records and a discussion of the uncertainties.

s s

2 Method and input data

Statistical downscaling follows a general scheme composed of three to four steps, involving several sources of information. In our case, at least three sources are required: 1) a GCM with coarse resolution but spanning a long time period, 2) a global reanalysis with coarse resolution spanning a shorter (recent) time frame, but of high quality, and finally 3) a regionally optimized RCM, with a finer resolution but spanning a short recent period. The objective is to evaluate the relationship (called here Perfect Prog - PP -) between the reanalysis and the RCM fields and then to apply it to the GCM in order to downscale its results to the scale of the RCM. Applying the PP directly to the GCM turned out to be unsuccessful due to discrepancies between the GCM climatology and the reanalyses. Therefore, three essential steps (Figure 1) are envisaged: the setting of a PP between the reanalysis and the RCM, the definition of a correction scheme of the GCM with the reanalysis over a common time frame to produce an "emulated" GCM, and finally the use of the PP on the "emulated" GCM for the whole period of the GCM integration.

but you are computing precip??

2.1 Input data

Four datasets are used in the development of the downscaled GCM: the reanalyses ERA-Interim and ERA5, the GCM runs composed of are the 10 -members- CESM2 simulations, and the optimized reanalysis-driven RCM integration RACMO2.3.

The reanalyses ERA-Interim (Dee et al. , 2011) reanalysis has been largely validated all over the Earth surface, and also over Antarctica, where the surface mass balance deduced from it was found to be consistent with satellites observations of CloudSat (Palerme et al. , 2017), and exhibits a significant correlation with the observed interannual variability (Wang et al. , 2016). Because of its relatively coarse spatial resolution (0.75°in latitude-longitude) which cannot resolve the atmospheric circulation over sharp local orographic changes, it has been however observed that the local scale snowfall over DML coast, which can be heavy due to orographic uplift, is not properly estimated by ERA-Interim (Palerme et al. , 2017). Heavy snowfall days from ERA-Interim correspond to less than 20% of the days identified by an automatic weather station near Kohnen (Welker et al. , 2014). By contrast, large-scale fields over the continent and Southern Ocean have been assumed to be more accurate. Therefore, ERA-Interim fields of relative humidity at 850 hPa, geopotential height at 500 hPa, mean sea level pressure, sea ice extent, sea surface temperature and air temperature at 850 hPa were extracted from ECMWF archives. Snowfall was extracted the same way, for comparison. More recently, the ERA5 reanalysis (Hersbach et al. , 2020) has been released with a higher spatio-temporal resolution in comparison with ERA-Interim (0.25°and 1 hour). The same fields have been extracted from the Copernicus Climate Change Service Data Store (C3S CDS). The comparison of ERA5 snow accumulation with meteorological

I believe CloudSat measured ~~ppe~~ precip – not SMB

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

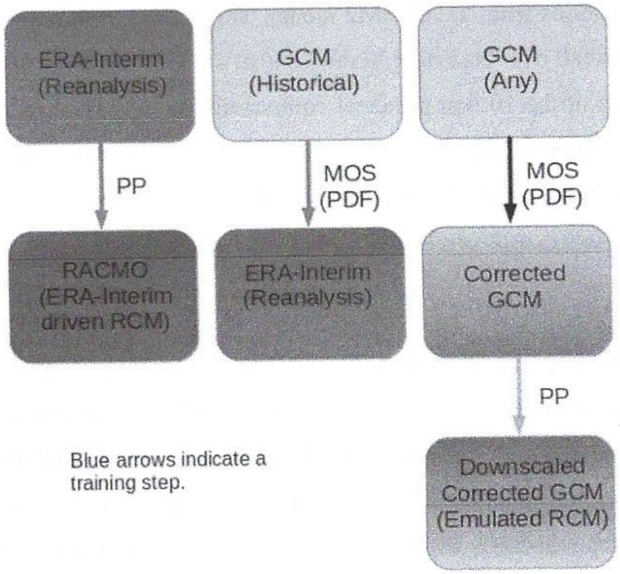


Figure 1. The statistical downscaling can be generally decomposed into 3 successive steps, including 1) the association of the reanalysis with the regional climate model by a Perfect Prog (PP, Maraun and Widmann (2018)), 2) a correction of the historical GCM to make it unbiased compared to the reanalysis using Model Output Statistics, and 3) the successive correction of the GCM followed by the application of the Perfect Prog., giving way to an emulated RCM estimation using the corrected GCM.

observations over the Antarctic Peninsula indicates an increased ability to identify strong precipitation events, which indicates a strong consistency between the synoptic weather patterns and the observed precipitation (Tetzner et al. , 2019).

The Regional Climate Model The Regional Atmospheric Climate MOdel version 2.3 (RACMO2.3) was forced at its boundaries by the ERA Interim reanalysis, including an upper-air relaxation. This provides a simulation of the atmospheric variables and precipitations over the Antarctic Ice Sheet from 1979 to 2016 (Lenaerts et al. , 2013; van den Berg and Medley , 2016; van Wessem et al. , 2016). A simulation at 5.5 km resolution over specific regions of Antarctica, including Dronning Maud Land, allows to better emphasize the orographic effects of precipitation and the potential wind redistribution of snow. The latter configuration at higher resolution allows studying in more detail a region with complex surface topography and SMB records (Lenaerts et al. , 2018, 2014). RACMO2.3 has been recognized to have the best fit to recent AIS SMB observations compared to other atmospheric and reanalysis models (Wang et al. , 2016; Rignot et al. , 2019).

The General Circulation Model The Community Earth System Model version 2 (Danabasoglu et al. , 2020) provides a model framework that allows the reconstruction of the evolution of atmospheric variables at 1 to 2 degree resolution over the globe. A set of ten historical CESM2 runs using forcing from CMIP6 have been made available for the industrial period (1850 - 2014, Historical CESM2 CAM6). A different initialization has been generated for each member. The CESM2 CAM6 provides new historical runs issued from the latest development of the CESM model, and is shown to be an improvement over CESM1 (Danabasoglu et al. , 2020).

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

between the two models??

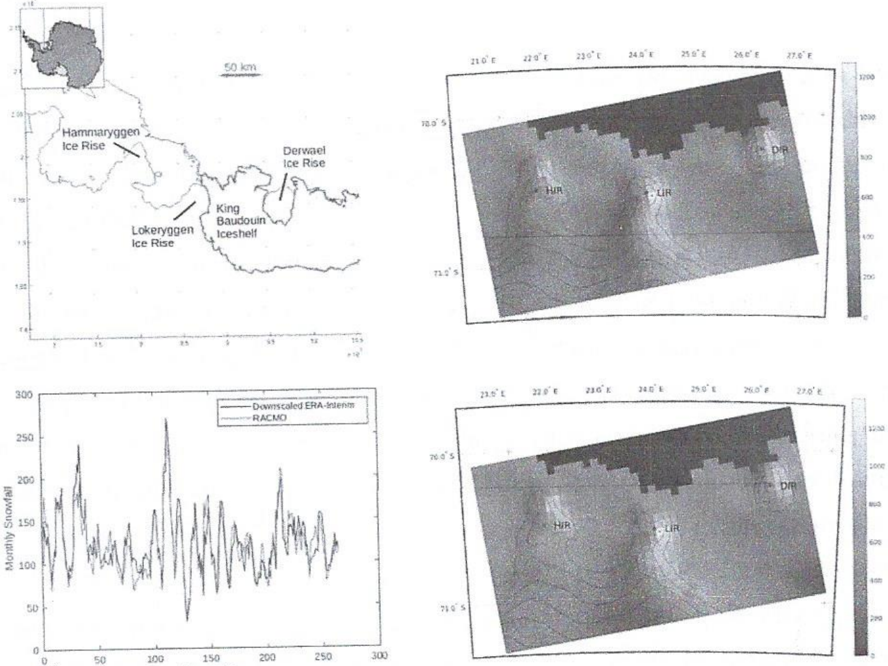
2.2 The Method

The comparison of the meteorological and snowfall fields estimated from the high resolution (5.5 km) RCM RACMO2.3 and from the low resolution (regridded to 1°latitude-longitude) reanalysis ERA-Interim has revealed large local differences in the snowfall amounts and spatial distribution over the Antarctic coast in Dronning Maud Land. ~~Two methods have first been unsuccessfully tried. First,~~ the analysis of back-trajectories (HYSPLIT model on NCEP/NCAR reanalysis (Stein et al. , 2015)) of strong snowfall events has not lead to a strong relation between the location of the origin of those air masses and the intensity of associated high precipitation amounts. This result contrasts with the success obtained with ERA-Interim and ERA5 in identifying the atmospheric rivers responsible ~~of~~ for such events (Gorodetskaya et al. , 2020). ~~Secondly,~~ another approach based on the Random Forest technique was tested, but failed probably due to unclear and non-systematic relations between large scale atmospheric patterns and snowfall intensity. In addition, since causal links can be model-dependent (Vannitsem et al. , 2019), a method that does not over-exploit those causal links is required to guarantee the transferability to other models in presence of large model uncertainties.

This is very vague

A downscaling method for snowfall based on the search for analogs has then been set up (Maraun and Widmann , 2018). The search for analogs by passes the problem of non-systematic relations, as long as similar synoptic situations provide on average similar precipitation amounts. The choice of a daily time step is in accordance with the duration of high precipitation events over Antarctica's coast, which account for about 60% of the annual snowfall (Schlosser et al. , 2010; Reijmer and van den Broeke , 2003; Turner et al. , 2019), last between 12 hours to 7 days long, and therefore can most of the time be associated to synoptic atmospheric situations (Reijmer and van den Broeke , 2003; Schlosser et al. , 2010; Welker et al. , 2014; Gorodetskaya et al. , 2020). The meteorological fields identified to be the most explanatory (and possibly replicable by climate models) for precipitation rates over the Antarctic continent have been decomposed in Empirical Orthogonal Functions (EOFs). The principal components weights (PCs) of the dominant EOFs (the first ten) are used for the selection of the analogs (Sneyers and Gossens , 1988; Hannachi , 2004). To select the large-scale meteorological fields of interest for the downscaling method, we have built ~~for~~ a set of 50 points from the RACMO2.3 domain an analog database from the association between the principal components weights (PCs) of different fields from ERA-Interim and the RACMO2.3 precipitation over a 11 years period (1979-1990). These points correspond to the highest annual snowfall accumulation over the area. The fields tested were: geopotential height at 700 and 500 hPa, air temperature at 500 and 850 hPa, relative humidity at 700 and 500 hPa, surface atmospheric pressure, sea-ice cover, and total precipitation (liquid and solid). One to four fields were used in the different tests. The statistical evaluation of the performance was based on correlation and root-mean square difference at both daily and annual scales after a bias correction (quantile mapping). The analysis over the validation period (1991-2000) indicates that the optimal choice is reached with four fields: 1) geopotential height at 500 hPa, 2) surface pressure, 3) relative humidity at 700 hPa and 4) total precipitation. The same evaluation methodology was used to select 1) the optimum number of closest analogs used for the estimation, namely 20 analogs, 2) the way to select the best estimation from the distribution of analogs: the mean of the ensemble seems the most appropriate, 3) the optimum minimum geographical spanning of the reanalysis fields: latitude 40°S-90°S, longitude 10°W-60°E, and 4) the length of the training period: 3 is a minimum, but we use 11 years without a

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021



(upper left)

(inset)

Figure 2. The map shows the Princess Ragnhild Coast, which is part of Dronning Maud Land (red box), and the whole Antarctic continent. Verification of the performance of the downscaling scheme trained on 11 years (1979-1990) of ERA-Interim data and RACMO2.3 daily snowfall and applied on ERA-Interim for a 10 years period (1991-2000). (Lower left) The monthly time series comparison over a location of maximum of annual snowfall accumulation shows a high degree of accuracy, while the spatial comparison of the accumulated snowfall (in mm) over 1996 on Princess Ragnhild Coast in presence of ice rises (eg Derwael Ice Rise – DIR -, Lokeryggen Ice Rise – LIR and Hammaryggen Ice Rise - HIR) illustrates the high degree of fidelity of the analog method (top right) in reproducing the RCM (bottom right) accumulation patterns, especially in the West-East difference of accumulation around ice rises (Kausch et al. , 2020).

large statistical gain. As expected, the spatial and temporal variations of RACMO2.3 have been preserved (Figure 2 shows an example of time series comparison and a comparison over Princess Ragnhild Coast characterized by the presence of ice rises near the coast). The root mean square difference on a daily basis over the validation period ranges from 10% in large accumulation areas near the coast to 15% in the inner regions. The advantage of the analog method is that it allows one to identify the major types of weathers systems delivering precipitation and their occurrence over time in the form of PCs. ~~This could~~ be of ~~interest for the analysis of a~~ frequency change in ~~the precipitation over long time~~ periods.

The downscaling method was then repeated over the ensemble of the ten climate runs from the CESM2 model over the period 1850-2014. Before the downscaling could be applied, we first needed to make the PCs compatible with the reanalyses. A simple bias correction based on the linear regression of EOFs (Feudale and Tompkins , 2011; Yu et al, 2018), followed by a quantile mapping, transforms the CESM2 original PCs into "Reanalysis-like" PCs. The operation is done after the verification

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

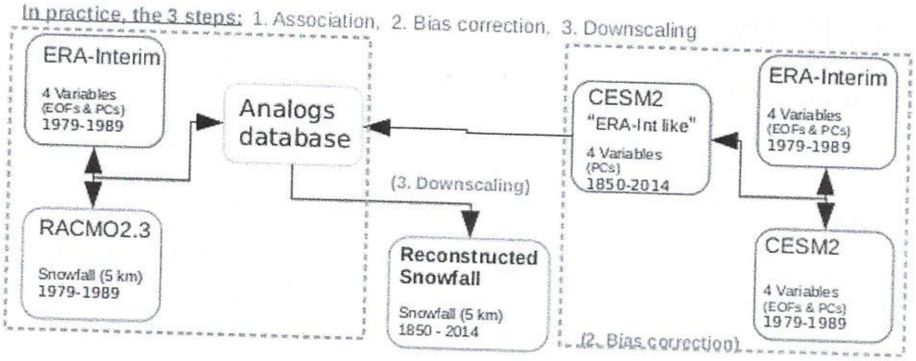


Figure 3. The analogs method follows 3 steps: 1) the building of the analogs database using the association of the reanalyses PCs to the snowfall accumulation from RACMO2.3, 2) the transformation of CESM2 into "reanalysis like" fields, 3) the search in the analogs database. The second step ensures the compatibility of the principal components between the database and the GCM.

I got lost on what the definition of the members is

are they CESM ? members

of the similarities of the EOFs among the members. The principal components can then be compared to the analog database in search for the closest events. As during the training process, quantile mapping defined over a 10 years period with the reanalysis is applied to the complete time series to obtain the final downscaled estimations (Figure 3). No significantly high differences are expected from any one member versus another, as the initializations have been carefully controlled and the spectral analysis of the time series of PCs give very similar results (Danabasoglu et al. , 2020).

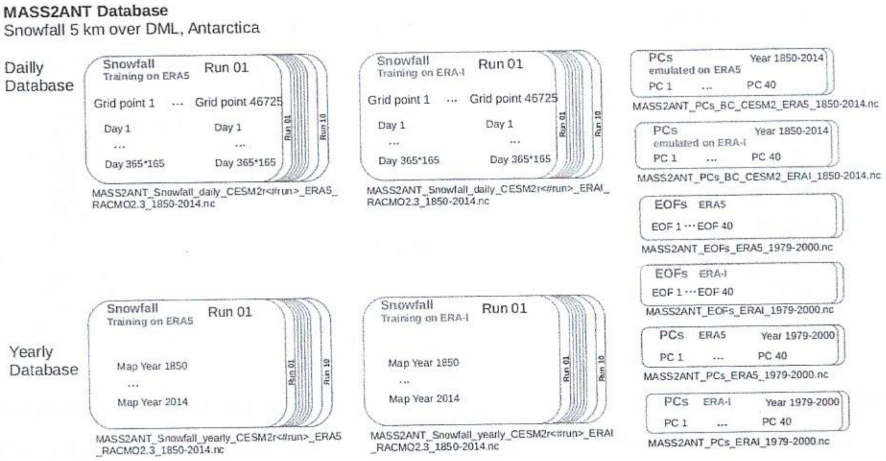
3 The dataset

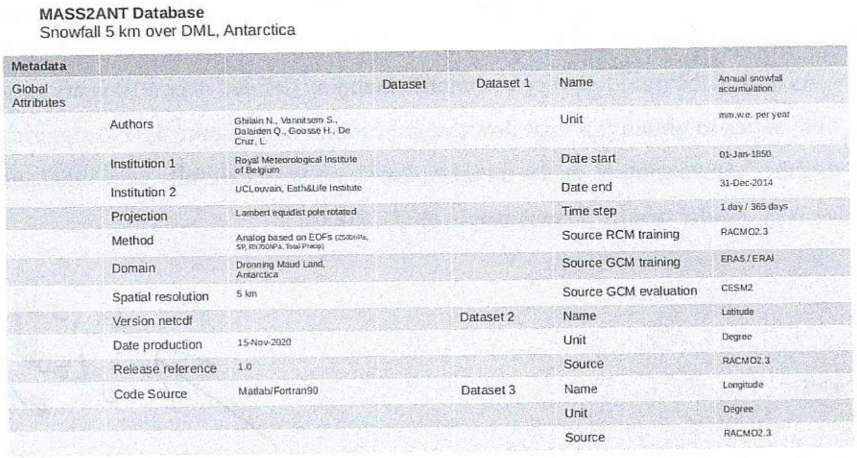
3.1 Structure of the dataset

this section should go to the end after illustrating that the dataset is interesting

The dataset is composed of 46 files (netcdf4 format) in total, which accounts for about 500GB, when uncompressed. The yearly accumulated snowfall maps are provided in polar stereographic projection at 5.5 km resolution, conforming to the RACMO2.3p5.5 run set-up: 1 file per downscaled CESM2 run, for each re-analysis used for training (20 files). The daily dataset is provided in the form of time series (20 files): 10 files (netcdf4 format), each containing the time series of daily snowfall accumulation over each grid point of the RACMO2.3 domain on the ice sheet (DML coastal region) for the period 1850-2014, 365 days per year. Each file corresponds to a downscaling of one of the 10 runs from CESM2, based on the training on ERA-Interim and RACMO2.3. Latitude, longitude, corresponding line and column in the RACMO2.3 domain are stored for each grid point. Extraction of time series for a specific location is therefore straightforward, and, thanks to the column and line, it is possible to recompose daily maps. Due to the size of the daily files, only a set of 2 files are stored along with the annual data files on Zenodo, and the whole set is available on request. Another set of 10 files with the same structure is stored with the results from the application of the same method trained on ERA5. In addition the time series of the 40 principal components used are stored in separate files, one for ERA-Interim trained, and one for ERA5 trained simulations (Figure 4), all provided with embedded metadata. PCs time series for the 40 EOFs for all the CESM2 members are stored in 2 additional files. The

no idea what this is




MASS2ANT Database
Snowfall 5 km over DML, Antarctica

Metadata						
Global Attributes			Dataset	Dataset 1	Name	Annual snowfall accumulation
	Authors	Ghilain N., Vannitsem S., Dalaiden Q., Goosse H., De Cruz, L.			Unit	mm.w.e. per year
	Institution 1	Royal Meteorological Institute of Belgium			Date start	01-Jan-1850
	Institution 2	UCLouvain, Eath&Life Institute			Date end	31-Dec-2014
	Projection	Lambert equidist pole rotated			Time step	1 day / 365 days
	Method	Analog based on EOFs (Z500hPa, SP, Rh700hPa, Total Precip)			Source RCM training	RACMO2.3
	Domain	Dronning Maud Land, Antarctica			Source GCM training	ERA5 / ERAI
	Spatial resolution	5 km			Source GCM evaluation	CESM2
	Version netcdf			Dataset 2	Name	Latitude
	Date production	15-Nov-2020			Unit	Degree
	Release reference	1.0			Source	RACMO2.3
	Code Source	Matlab/Fortran90		Dataset 3	Name	Longitude
					Unit	Degree
					Source	RACMO2.3

Figure 4. The MASS2ANT database (NetCDF files) includes 2 files for yearly accumulated snowfall (10 members each), and one per reanalysis used for training. The daily estimations are stored per year and per reanalysis used for training. In addition, PCs time series for the 40 EOFs for all the CESM2 members are stored in 2 separate files, as well as the 40 EOFs of the emulated CESM2. Embedded metadata include units, version geographical coordinates, and temporal stamps.

EOFs maps from ERA-Interim and ERA5 are stored in 2 separate files. At last, PCs time series from ERA-Interim and ERA5 fields are stored in 2 files.

3.2 Examples of maps (annual, daily)

As expected, daily fields differ among the different members of the downscaled CESM2 results. When a high precipitation event over DML can be present on a specific day in member 7, it may not be the case for all the other members (Figure 5).

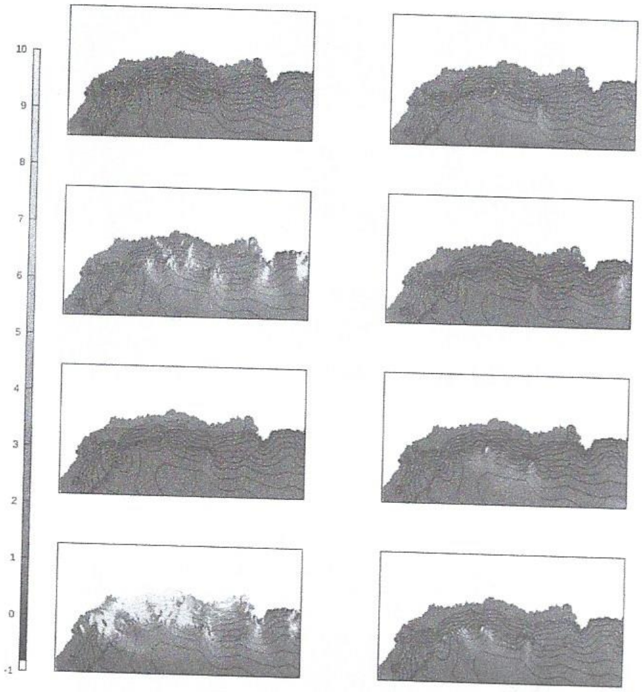


Figure 5. Recomposed maps of downscaled daily snowfall (in mm) over DML on the 1st day of 1850 for members 1 to 8. A saturation at 10 mm per day has been defined for visualization.

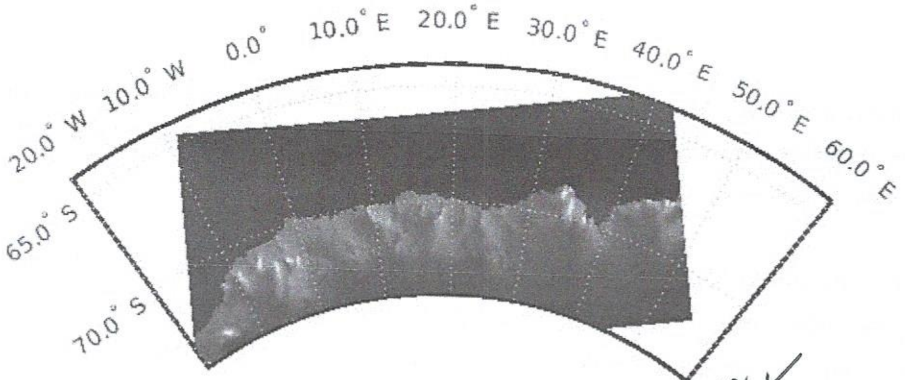


Figure 6. Map of downscaled snowfall accumulated over a year (CESM2 member 1).

which year

seems like you need to show the annual precip for all members

While a large disparity is observed on daily maps, the effect is much smaller on the annual ~~accumulation~~ precip (Figure 6). We illustrate this by a daily map (members 1 to 8) and an annual accumulation over the entire domain of the downscaled CESM2 member 1, trained with ERA5 (Figure 5).

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021

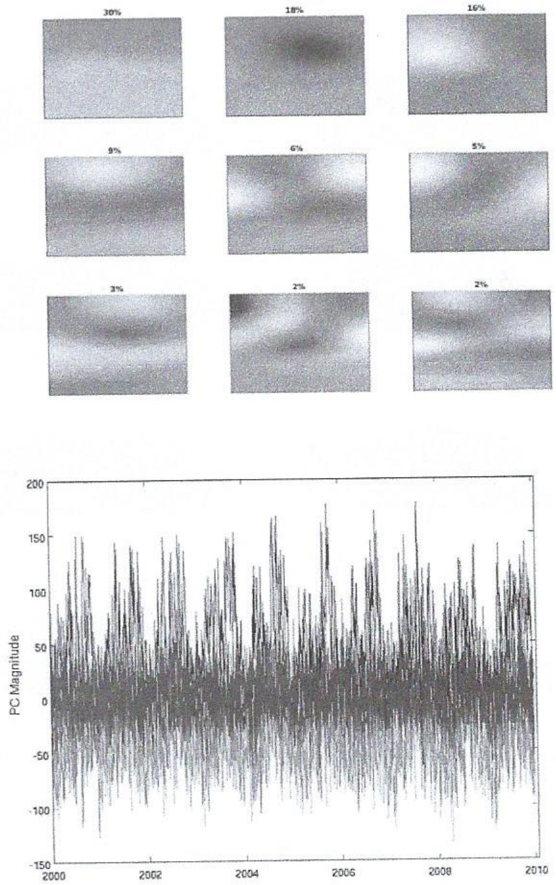


Figure 7. The configurations of the 9 first EOFs for surface atmospheric pressure with their relative occurrence are ordered by importance in the decomposition of the variability of the field (represented here unitless, covering the total area 40°S-90°S, 10°W-60°E). The time series of the associated PCs for one year show the different amplitudes.

For each date, the PCs are provided, and associated to the EOFs for each meteorological field. An analysis of the association between snowfall regimes with the synoptic patterns is then possible. In Figure 7, we show an illustration of the first 9 EOFs for Surface atmospheric pressure over the GCM domain considered (40°S-90°S, 10°W-60°E), as well as the time series of the associated PCs, showing their various amplitudes.

3.3 Comparison with ice core measurements

The annual variation of the surface mass balance is estimated from ice cores retrieved from a set of locations over the coastal DML. The length of the records differs between the different sites and can cover more than 150 years. The annual precip ~~accumulation~~ computed from the downscaled daily snowfall is compared to SMB from the different ice core records from Thomas et al.

Table 1. Ice cores drilled over Dronning Maud Land used for comparison, extracted from Thomas et al. (2017)

Ice Core	Latitude	Longitude	Start date	End date	Reference
S100	-70.2439	4.8	1850	1999	Karczmarska et al. (2004)
S20	-70.2472	4.8183	1956	1996	Isaksson et al. (1999)
H72	-69.2	41.08	1850	1999	Nishio et al. (2002)
B04	-70.6167	-8.3667	1892	1981	Schlosser (1999)
DIR	-70.25	26.34	1850	2011	Philippe et al. (2016)
E91	-73.6	-12.4333	1932	1991	Isaksson et al. (1996)
DML16 FB9813	-75.1669	5.0006	1850	1997	Oerter et al. (2000)

(2017), after a conversion of ice core estimates to water equivalent height (Figure 1, Figure 8). A good overall match between 185 the downscaled snowfall and the ice cores is found, as well as a large bias reduction compared to the use of CESM2 without downscaling.

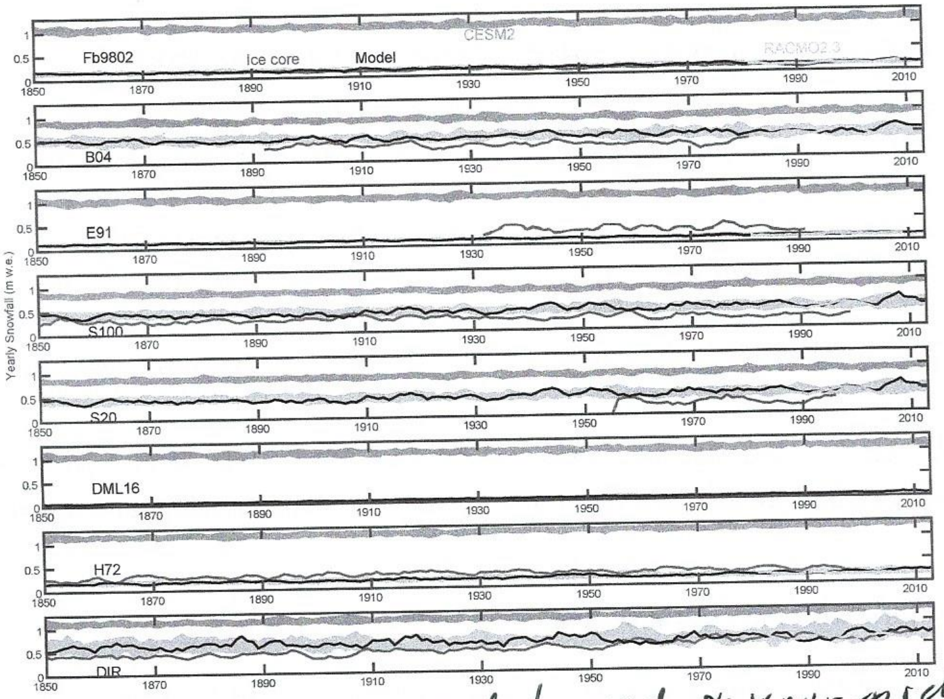
3.4 Comparison with other sources and other studies

Almost all the available datasets of spatial estimates of snowfall accumulation or SMB over Dronning Maud Land are 1) at high resolution (20-30 km (Lenaerts et al. , 2018; Agosta et al. , 2019)) but spanning the last 30 years, 2) spanning a longer period 190 but at low resolution (0.75°, Medley and Thomas (2019)), or 3) at high resolution but averaged in time (Rotschky et al. , 2007). Over other regions, downscaling using a physical orographic model has been used to obtain maps at high resolution from a RCM over a longer period of time (Agosta et al. , 2013). In this study, we compare our dataset to two other sources. The first reconstruction covers Western Dronning Maud Land (Rotschky et al. , 2007), which offers a static view of the accumulation patterns based on the spatial interpolation of ice core data (data available from the PANGAEA website http://doi.pangaea.de/ 195 10.1594/PANGAEA.472297). A visual comparison with the CESM2 member 1 downscaled (trained with ERA5) shows an enhanced variability along sharp elevation changes (Figure 9). The second reconstruction over the entire domain is based on ice core datasets and climate patterns from GCM over the last 150 years on a yearly basis (Medley and Thomas , 2019), of which we only consider the reconstruction using ERA-Interim (data available at https://earth.gsfc.nasa.gov/cryo/data/antarctic-accumulation-reconstructions). The downscaled CESM2 member 1 averaged over the total period seems in agreement with the 200 averaged reconstruction based on the ice core records, especially to reconstruct the gradient from the coast to inner land, but depicts much more details related to topography (Figure 10).

4 Uncertainties

The uncertainties in the daily snowfall estimates compiled into this dataset can arise from 1) the choice and accuracy of the model sources (regional climate model, re-analysis, "historical" climate model), and 2) the accuracy of the downscaling

https://doi.org/10.5194/essd-2021-12
Preprint. Discussion started: 27 January 2021



is the ice core data red & your model black

Figure 8. The time series of yearly accumulated snowfall extracted from the database (CESM2 member 1, training with ERA-Interim, and all members in shaded gray) is compared to the SMB estimated from ice cores available from (Thomas et al. , 2017) (Ice core data available on the webpage of PAGES2k/Antarctica2k and in https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/00940). RACMO2.3 time series has been superimposed (cyan) and the 10 CESM2 simulations (shaded blue), showing the large bias reduction thanks to the use of RACMO2.3 for the training. A 4-year moving average has been applied on the time series for the visualization.

There seems to be about a 25% variation between the model & the ice core data

method (choice of parameters, choice of predictive variables). In this section, we report a (non-exhaustive) quantification of the uncertainty level related to both the choice of model sources and of the parameters of the method.

4.1 Sampling in choice of analogs

A simple way to assess the effect of the sampling of the analogs on the estimation uncertainty is to use the bootstrap techniques. We have created a distribution of 100 ensembles of analogs, with the 20 analogs randomly drawn with repetition at the ice core sites (Figure 11). The downscaled CESM2 member 1 with 100 different analogs draws represented at the ice core sites. The uncertainty is between 5 and 10 % of the mean value.

also the variations in the ice core and model data do not seem well correlated. That should be quantified

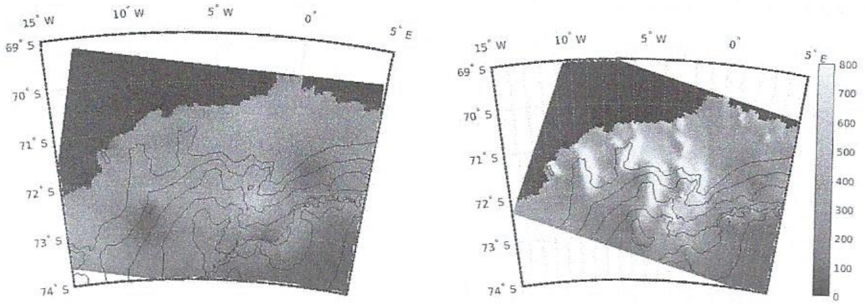


Figure 9. The visual comparison of the average annual snowfall accumulation over Western Dronning Maud Land between the database from (Rotschky et al. , 2007) and the downscaled CESM2 member 1 (trained with ERA5) shows more pronounced asymmetrical patterns near sharp surface elevation changes. The difference in magnitude should be taken with caution, as only the snowfall component is provided by the MASS2ANT database, while (Rotschky et al. , 2007) is the integration of the total surface mass balance. / which figure is which

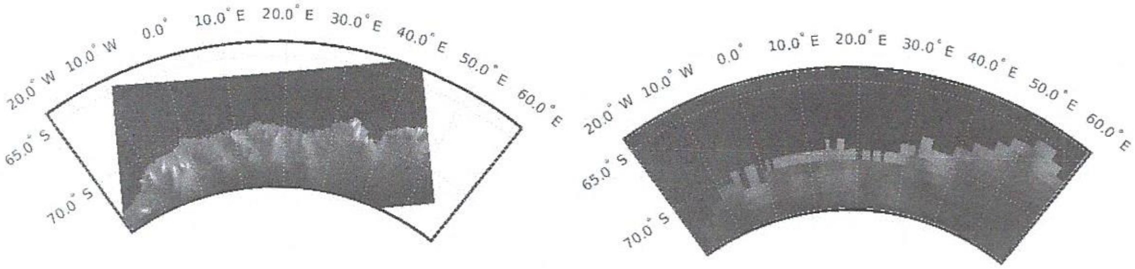


Figure 10. The coast-to-inland gradient of the downscaled CESM2 member 1 averaged over the total period seems in agreement with the averaged reconstruction based on the ice core records of (Medley and Thomas , 2019), but depicts much more details that could be useful in detailed analysis of ice core records representativity (same color scale).

4.2 Choice of reanalysis used for training

The choice of reanalysis used for training the analogs database has an effect on the downscaling obtained from CESM2. The choice of reanalysis influences the result, at equivalent training scores (Figure 12). This is one of the reasons for extending the database to ERA5 as training reanalysis instead of ERA-Interim. More independent reanalyses could be used to further assess the uncertainty. The difference between CESM2 downscaling using ERA-Interim or ERA5 in the training reveals an uncertainty of 25% on ~~in~~ average at the ice core sites.

4.3 Choice of "historical" GCM runs

CESM2 simulations consist of 10 members, corresponding to different initializations (Danabasoglu et al. , 2020). The inherent model uncertainty is therefore impacting the results and is quantified here (Figure 13). The scatter of the 10 members of

Table 2. Summary of the uncertainty levels associated to the choices of input model or sampling method. The uncertainties (mean uncertainties) are estimated on yearly accumulation of snowfall, relative to the basic methodology followed to derive the database, for at least one time series over the domain.

	Snowfall Uncertainty (% mean yearly accumulation)
Sampling in choice of analogs (1)	5 to 10 %
Choice of reanalysis for training (2)	25%
Choice of "historical" GCM runs (3)	17-22 % (std) to max 30-38%

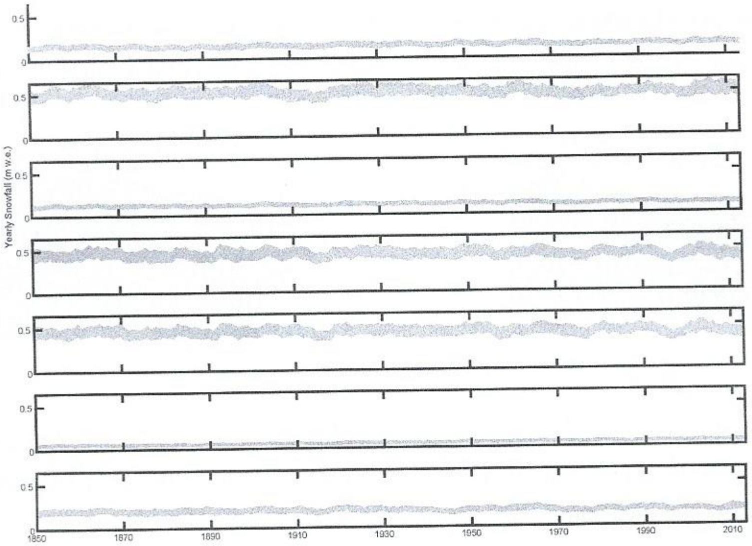


Figure 11. The downscaled CESM2 member 1 with 100 different analogs draws represented at the ice core sites. The uncertainty is between 5 and 10 % of the mean value.

CESM2 downscaling using ERA-Interim in the training reveals an uncertainty of 20% (standard deviation) and up to 40% at some ice core sites, maybe a direct consequence of the CESM2 members not being temporally correlated.

4.4 Choice of RCM model

At last, the choice of the RCM critically drives the dynamical link between large-scale patterns and local snowfall, and the average and maximum amplitude of snowfall accumulation. An assessment of the uncertainty caused by the choice of RCM could be done using another RCM, with equally proven quality. For example, the Modele Atmospherique Regional (MAR) is

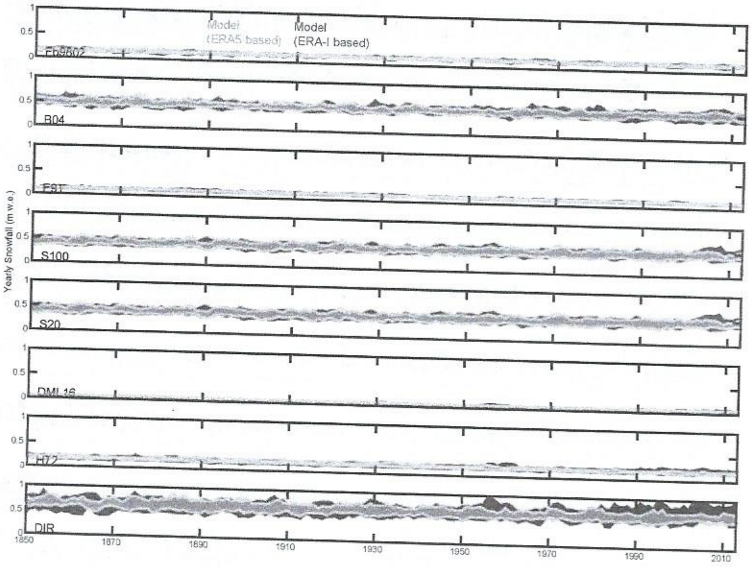
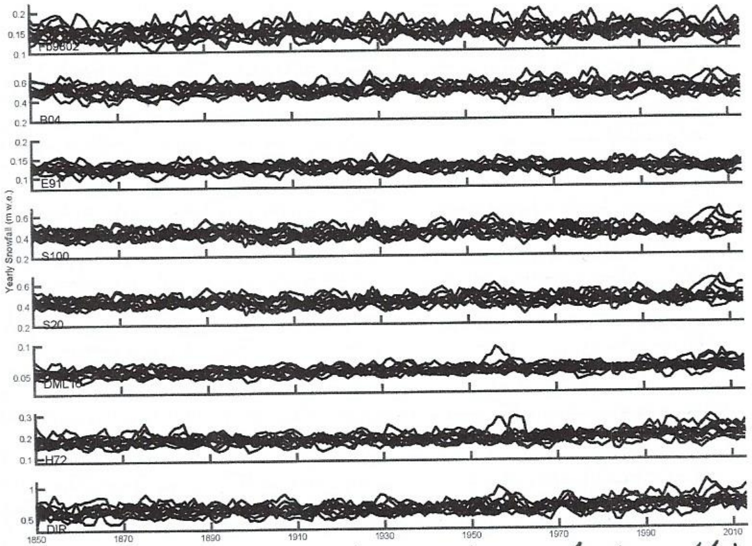


Figure 12. The difference CESM2 downscaling using ERA-Interim or ERA5 in the training reveals an uncertainty of 25% on average at the ice core sites.

another recognized RCM used for polar climates at high resolution. The newest simulations over Antarctica at 30 km resolution (Agosta et al. , 2019) have shown similarities with RACMO2 at 27 km resolution, but revealed localized differences over the ice sheet related to a more realistic sublimation of falling snowfall in comparison with RACMO2 (Gallée et al. , 2013). A simulation at the same resolution with another RCM optimized for polar regions, like MAR, could be of great interest to better frame the uncertainty linked to RCM physics. The uncertainty linked to the choice of RCM has not been evaluated in this study and could be envisaged to extend the database if new RCM simulations at such resolution are available in the future.

5 Conclusions

We propose here a reconstruction of snowfall evolution over Dronning Maud Land, Antarctica, at 5.5 km resolution using an analog-based downscaling technique. This technique has allowed us to exploit the detailed spatio-temporal estimation of snowfall from 30-year RACMO2.3 simulations in combination with synoptic patterns from recent reanalyses to statistically downscale the historical runs from CMIP6 (CESM2 model). The resulting database stores the ensembles of daily accumulated snowfall from 1850 to 2014, the pertinent information for synoptic patterns analysis (the principal components weights and empirical orthogonal functions from four large-scale meteorological fields), and the annual evolution of accumulated snowfall over Dronning Maud Land. The database can be used to analyze the detailed contribution of snowfall to the surface mass



It would be interesting to find a trend in this record but the uncertainty seems to preclude that. ~~W~~ Without

Figure 13. The scatter of the 10 members of CESM2 downscaling using ERA-Interim in the training reveals an uncertainty of 20% (standard deviation) and up to 40% at some ice core sites.

some demonstration of the product utility, it is

balance over the region, its evolution and its association to synoptic weather conditions. The method can be easily replicated with new RCM and GCM simulations.

hard to know how applicable the product is to glaciological analysis

6 Data availability

The files of the dataset (the annual snowfall, the PCs and the EOFs) are available on Zenodo platform (http://doi.org:10.5281/zenodo.4287517). However, due to size limitations, only 2 daily snowfall files out of 20 have been stored there, the whole set is available on request to the contact author.

Author contributions. Project proposal and funding: SV & HG; Scientific supervision: SV; Design of the method: LDC, HG, SV & NG; Climatological Analysis NG, WW & SV; Implementation, tests, method tuning, verifications, database creation, validation: NG & SV; Analysis of the results: all authors; Preparation data material: QD & NG. Manuscript drafting: NG; Manuscript revisions: all authors.

Competing interests. The auhors declare that they have no conflict of interest.