Plan Overview

A Data Management Plan created using DMPonline.be

Title: (GPUCL/24/012)Antarctic Climate Causality: Unraveling links between the components of the system and their variability (ACCU)

Creator: Sebastian Berghald

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Project abstract:

Antarctica and the Southern Ocean play a key role in the Earth's climate system, influencing global heat balance and carbon uptake. Recent anomalies, including a drastic sea ice decline, anomalous snowfall, and unprecedented heat waves, challenge our understanding of the region's response to climate change. This proposal aims at an improved understanding of recent changes occurring in the Antarctic climate by investigating the interconnections between the ice sheet surface mass balance (SMB), sea ice, ocean temperature, and ice shelf melt, utilizing observations and a recently developed fully coupled ice-sheet–ocean–sea-ice–atmosphere–land model, PARASO, covering Antarctica and the Southern Ocean. The interactions between climate components cover both direct links and indirect links due to varying large-scale atmospheric processes. The methodology integrates observations and simulations, employing causal frameworks to identify relationships. The work packages involve identifying links between atmospheric circulation, sea ice, and SMB variability from observations (WP1), unravelling the role of large-scale atmospheric variability (WP2), exploring potential feedbacks related to sea ice changes (WP3), and synthesizing contributions to recent and future climate changes (WP4). By employing advanced causal inference methods, the project strives to overcome challenges in attributing causality from observational data. In a complementary way, the PARASO model allows for a comprehensive analysis of the interplay between different components that cannot be identified from observations alone. Moreover, the model will be used for future Antarctic climate projections. This proposal's outcomes will advance our understanding of recent Antarctic climate changes and their interlinkage, offering insights into the future of the coupled Antarctic climate system.

ID: 213288

Start date: 16-09-2024

End date: 16-09-2028

Last modified: 21-03-2025

(GPUCL/24/012)Antarctic Climate Causality: Unraveling links between the components of the system and their variability (ACCU)

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: N(ew data) or E(xisting data)	Indicate: D(igital) or P(hysical)	Indicate: Audiovisual Images Sound Numerical Textual Model SOftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
PARASO 1	WP2: 25 years of simulations with PARASO	N	D	M	NetCDF	>5TB	-
PARASO 2	WP3: 15 years of simulations	N	D	M	NetCDF	>5TB	
PARASO 3	WP4: 30 years of simulations	N	D	M	NetCDF	>5TB	
ERA5	relevant variables and periods	E	D	N	NetCDF, Grib	>5TB	
RACMO	relevant variables and periods	E	D	N	NetCDF	<100GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

ERA5 data is freely available at and retrieved from Copernicus CDS (https://cds.climate.copernicus.eu/). Monthly output from RACMO2.3p2 will be retrieved from Zenodo archives (https://doi.org/10.5281/zenodo.7760491). If available in time, RACMO2.4p2 will be given preference. Both ERA5 and RACMO model data (and potentially other RCM and/or observational data) will only be stored temporarily.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)?

If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

The produced data itself will have relevant metadata stored in the files themselves. They will also be accompanied by a README file including all necessary information. In the PhD thesis and associated publications further descriptions will be provided and/or references included (e.g. regarding the model setup). All publications will be accompanied by electronic notebooks (Jupyter), which provide code and explanations to reproduce the results.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

All data generated and intended for long-term storage will be given persistent identifiers and referenced in publications.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)
- ManGO

1) Relevant notebooks with code will be saved in the RCS (working group) Gitlab
2) Processed data intended to be publicly available will be made available through KU Leuven Research Data Repository
3) Processed data useful for follow-up research in the RCS group will be stored on large volume storage
4) Raw PARASO (model output) data will be stored on Storage4Climate VSC Tier-1 Data Service iRODS for 7 years and then

removed.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data intended for long-term storage (i.e. related to publications and/or the PhD thesis) will be made publicly available if feasible. For other data, appropriate measures will be taken where necessary.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The costs for large volume storage are expected to be around 100EUR per year and will be carried by RCS working group (Nicole van Lipzig).

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- Certain data cannot be kept for 10 years (explain below)

Data from other sources (like ERA5) which are easily available in other repositories won't be saved. Also temporary data which are not relevant for publications and/or the PhD thesis won't be retained. Raw PARASO model output will be removed after 7 years. This is to avoid generating unnecessary demand for long-term storage.

Where will these data be archived (stored and curated for the long-term)?

- Large Volume Storage (longterm for large volumes)

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

See above.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Other (specify below)

Relevant processed data and codes will be made publicly available on Gitlab (codes) and KU Leuven RDR (data).

If access is restricted, please specify who will be able to access the data and under what conditions.

Question not answered.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other (specify below)

KU Leuven RDR (data), Gitlab (codes)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- MIT licence (code)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

KU Leuven RDR

What are the expected costs for data sharing? How will these costs be covered?

see above.

Responsibilities

Who will manage data documentation and metadata during the research project?

Sebastian Berghald

Who will manage data storage and backup during the research project?

Sebastian Berghald

Who will manage data preservation and sharing?

Sebastian Berghald, together with Profs. Nicole Van Lipzig, Stef Lhermitte and Hugues Goosse to ensure long-term availability also after the end of the PhD project.

Who will update and implement this DMP?

Sebastian Berghald