Plan Overview

A Data Management Plan created using DMPonline

Title: Quantification of meltwater storage in Greenland using a statistically-optimal estimation of mass anomalies with satellite gravimetry data

Creator: Pavel Ditmar

Affiliation: Delft University of Technology

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

Project abstract:

The overall aim of the project is two-fold: (i) to develop a novel methodology for a statistically-optimal estimation of mass anomalies at the Earth’s surface from satellite gravimetry data and (ii) to apply the developed methodology to a quantification and analysis of meltwater storage in Greenland. The proposed methodology will allow mass anomalies per cell of a global grid to be directly estimated by an inversion of a set of spherical harmonic coefficients. A conducted feasibility study has demonstrated that a high-quality estimation of mass anomalies can be obtained in this way, provided that a proper regularization strategy has been identified. The developed methodology well be used to quantify the accumulation of meltwater in Greenland in the course of melt season. Our earlier study has already demonstrated that satellite gravimetry definitely can sense this process. We expect that the novel data processing methodology, in combination with a larger amount of satellite gravimetry data and their higher quality, will result in a substantial improvement of meltwater mass estimates. We intend to benefit from this in an attempt to separate the total meltwater mass that is subject to runoff into the “fast” component (primarily related to basal meltwater) and “slow” component (mostly related to meltwater storage in the firn layer). After that, we will analyse, among others, correlations between “basal” water mass and ice flow velocities. We expect that this will facilitate better understanding the complex relationships between meltwater runoff variations and ice flow velocities per drainage system (or, perhaps for largest outlet glaciers individually). In this way, we will pave the way for a more accurate forecasting the future behaviour of ice flows and resulting sea level rise in response to the on-going climate change.

ID: 117014

Last modified: 13-06-2023

Grant number / URL: ENW.GO.001.040

Quantification of meltwater storage in Greenland using a statistically-optimal estimation of mass anomalies with satellite gravimetry data

General Information

Name applicant and project number

Pavel Ditmar, NWO-project number: ENW.GO.001.040

Name of data management support staff consulted during the preparation of this plan and date of consultation.

This DMP has been reviewed by Data Steward at TU Delft: the Data Steward of the Faculty of Civil Engineering (Xinyan Fan) and of the Faculty of Aerospace Engineering (Heather Andrews).

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

The primary data to be used in the course of the project are satellite gravimetry level-2 data (spherical harmonic coefficients), which are based on K-Band Ranging (KBR) measurements provided by GRACE satellite mission, as well as KBR and Laser Ranging Interferometry (LRI) measurements provided by GRACE Follow-On (GFO) satellite mission. Together, these missions cover the time interval from Apr. 2002 until now, with an about 1-year gap in 2017-2018. Different variants of GRACE/GFO level-2 data are prepared by about 10 data processing centres/groups with a monthly temporal resolution. All these data are publicly available from, e.g., the website of ICGEM - International Centre for Global Earth Models ICGEM (http://icgem.gfz-potsdam.de). The data will be supplied by stochastic models of data noise in the form of inverse error variance-covariance matrices (i.e., normal matrices). These matrices are publicly available from a website of the Institute of Geodesy - TU Graz (https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018. In addition, we will exploit estimates of Greenland Ice Sheet (GrIS) Surface Mass Balance (SMB), as well as its individual components. These estimates will be based, in the first instance, on the Regional Atmospheric and Climate MOdel (RACMO) version 2.3p2 and will be provided by the Institute for Marine and Atmospheric research Utrecht (IMAU) at Utrecht University, which participates in the proposed project. Those estimates are produced with a spatial resolution of 1 km and a daily temporal resolution. They are provided for the entire SMB and for its 7 individual components over the total GrIS (the area is about 2,000,000 square kilometres). The storage space required by the intended input data is summarized in the following table:

Data description/ source	Format	Size per variant/ component	Number of variants/ components	Number of months	Total size
GRACE/GFO level-2 data (from ICGEM)	ASCII	0.5 MB per month	10	300	1.4 GB
Stochastic models of data noise (from TU Graz)	Binary	1.1 GB per month	1	300	330 GB
GrIS SMB estimates (from IMAU)	Binary	15.3 MB per day or ~460 MB per month	1+7 = 8	300	~1073 GB
Total					~1.4 TB

GRACE and GRACE-FO data are available on NASA's PO.DAAC ("Data hosted by the PO.DAAC is openly shared, without restriction, in accordance with NASA's Earth Science program Data and Information Policy"; https://podaac.jpl.nasa.gov/CitingPODAAC). RACMO2 dataset is published in Zenodo and it is available under a CC-BY license (https://doi.org/10.5281/zenodo.3677642).

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

In the first instance, we will produce various estimates of mass anomalies at the Earth's surface at the nodes of a global equiangular

grid. The estimates will be produced for the time interval covered by GRACE/GFO satellite missions with a monthly temporal resolution. The estimates will ultimately be stored in NetCDF format. These estimates will be used to produce time-series of meltwater storage within the GrIS and its individual parts. The estimated storage space required by the intended output data is reported in the following table:

Data description	Format	Size per month per variant	Number of variants	Number of months	Total size
Mass anomalies at the Earth's surface	NetCDF	648,000 cells x 8 bytes = 5 MB (0.5 x 0.2 degree resolution)	~100	300	~145 GB
Time-series of meltwater storage over the GrIS and its individual parts	.txt				Marginal (<5 GB)
Total					~150 GB

1.3. How much data storage will your project require in total?

- >1000 GB

In line with the tables presented under 1.1 and 1.2, it is envisioned that the total volume of stored data will be of the order of 2-3 TB (about 1.4 TB for input data, about 0.15 TB to store the project results and the rest to store intermediate data).

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

During the project, proper supporting and embedded documentation will be created as the datasets are generated. Datasets will be kept in a proper directory structure with a top-level README (.txt or .md). Guidance provided by 4TU.ResearchData will be followed when preparing these README files. In particular, they will contain information about:

- what the dataset is about
- how it was collected/generated (including experimental settings, facilities, protocols, etc.)
- software used to analyze the data (including versions, dependencies, etc.)
- directory structure setup (file naming conventions, directory tree structure)
- headers and units (if not present in the files already)
- references, caveats and assumptions

2.2 Indicate which metadata will be provided to help others identify and discover the data.

Researchers will be encouraged to follow the**disciplinary metadata standards and conventions**whenever such standards exist. All data supporting publications will be made openly available through 4TU.ResearchData. 4TU.ResearchData is a trusted and certified research data repository (it has a Data Seal of Approval certification). All datasets will be accompanied by metadata to ensure that the data are findable. Importantly, **netCDF** format also has **embedded metadata**. The idea is for these metadata to follow a **standard vocabulary** (e.g., meaning the researchers will not come up with their own 'keys' and free-text values for those 'keys'). In addition, existing **metadata standards** (https://www.dcc.ac.uk/resources/subject-areas/earth-science) will be investigated and applied when relevant. Moreover, to further aid the data discoverability, keywords describing the datasets will be added. 4TU.ResearchData is also using schema.org metadata, meaning that all datasets are indexed in Google Dataset Search. Furthermore, the 4TU.ResearchData archive assigns a Digital Object Identifier (DOI) to each dataset, which makes the datasets citable and persistently available. The DOI of each dataset will be included in the text of the related article(s) (e.g, in the Data Availability section) and the DOI of each related article(s) will be included in the citation metadata of each dataset.

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

During the course of the research project, all data will be stored on local servers (Project Drive) maintained and automatically backed up by TU Delft ICT. Data can be recovered with the help of TU Delft ICT services in the event of an incident.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

The general TU Delft ICT facilities apply on the aspects of security and protection. The data are not sensitive.

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

No personal data, not applicable

4.2 How will ownership of the data and intellectual property rights to the data be managed?

TU Delft will have ownership of all scientific data generated in the project. The datasets underlying the published papers will be publicly released following NWO's policies. During the active phase of research, the lead applicant from TU Delft will oversee the access rights to data (and other outputs), as well as any requests for access from external parties. A publication of the project results will be accompanied by a release of the associated data with an appropriate licence, see with question 5.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

All data supporting publications will be made openly available through 4TU.ResearchData. 4TU.ResearchData is a trusted and certified research data repository (it has a Data Seal of Approval certification), and ensures that research data will be preserved for at least 15 years.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- No

5.3 What data will be made available for re-use?

- All data resulting from the project will be made available

The project data will be retained for at least ten years on TU Delft servers for the purposes of further validation and analysis. The datasets underlying the figures and conclusions in academic papers will be made publicly available through 4TU.ResearchData, in line with the TU Delft Research Data Framework Policy.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available as soon as article is published

All research data underpinning research papers will be made publicly available by depositing at 4TU.ResearchData at the time of the publication of the corresponding research article, or shortly afterwards (up to several months, at most, for instance when a series of papers is being written on a similar subject, based on the same run/dataset, possibly by different members of the team). The data will be preserved for at least 15 years.

5.5 In which repository will the data be archived and made available for re-use, and under which license?

The datasets underlying the published papers will be published at 4TU.ResearchData, which is a trusted and certified research data repository (Data Seal of Approval certification). All datasets will be licensed under a CC-BY licence which requires attribution/credit for the original creation, while at the same time ensures broadest possible re-use. All datasets will be accompanied by descriptive metadata (see question 2), to ensure that the datasets are Findable, Accessible, Interoperable, and Re-usable (FAIR).

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

The primary data processing software will be written in Fortran, building upon a prototype code developed in-house earlier. Auxiliary scripts (e.g., for data downloading and visualization) will be developed by the principal investigator using the programming language of his/her preference. The developed software and scripts presented in academic papers will be shared on GitHub and those GitHub repositories will be published via 4TU.ResearchData. This way, they will be publicly available to anyone for re-use under an open licence. They will be also assigned a Digital Object Identifier (DOI), to make them citable and persistently available.

Researchers will use version control with Git for developing the code. The TU Delft Gitlab will be used for remote repositories during development, while local repositories will be in the institutional laptops/workstations. Once the codes are ready to be shared, the repository will be migrated to a public Github repo for more visibility. A snapshot of the Github repository will be archived accordingly in the 4TU.ResearchData archive.

Each repository will have a top-level README (.txt or .md) which will include information about:

- what the code does
- how to compile/run the scripts
- requirements (versions, libraries, dependencies, environment setup)
- instructions or notebook exemplifying how the code works with a small benchmark dataset

Researchers will also be encouraged to pay attention to the readability of the code and (modular) structure of the code.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Data management is not separately budgeted. 4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. We do not expect to exceed this and therefore there are no additional costs of long term preservation. The management of project data will be primarily carried out, as a part of the project, by the principal investigator.