Open-Source Tools to Enable Geophysical Data Processing and Inversion

Santiago Soler (University of British Columbia)
Lindsey Heagy (University of British Columbia)
Craig Miller (GNS Science)
Leonardo Uieda (University of Liverpool)

July 28 — July 30, 2023

1 Overview

One of the main goals of geosciences is to obtain information about the structure of Earth's subsurface. Geophysics relies on measurements of physical quantities such as seismic waves, and gravitational and electromagnetic fields to characterize the inner structure of the Earth through its physical properties. The process through which geoscientists build this understanding of Earth's interior can be summarized as: (i) acquiring the data, (ii) processing it, and (iii) producing a model of the subsurface, through inversion, that is consistent with the observed data. This workflow requires mathematical and computational methods that include numerical simulations of partial differential equations, solving large-scale optimization problems, and methods for assimilating a variety of data types across a range of scales. Software tools that provide computational solutions to these problems are a major resource for geoscientists. They will surely play a fundamental role in enabling the research needed to address upcoming geoscientific problems related to energy transitions and climate change remediation.

In the last decade, there has been a transformative move within the scientific community towards reliance on and contribution to open-source software. Within the field of geophysics, the Fatiando a Terra [12, 13, 10, 5, 4] and SimPEG [2, 6] projects were early pioneers that have driven the development of Python-based open-source tools for data processing, numerical simulations, inversion, and visualization. These two projects share a common goal of providing open-source Python tools to solve geophysical problems, building foundational frameworks that simplify the implementation of state-of-the-art methodologies, and facilitating their usage in applied and basic research. Their ongoing and future development benefits not only their own communities but also the broader geoscientific community.

During this 2-day BIRS Workshop members of the communities of users, developers and contributors of Fatiando a Terra and SimPEG were invited to gather and discuss the future development roadmaps for both projects, strengthen the collaboration networks between the two projects, identify issues and points of improvement, and build solutions for them.

A total of 17 people were able to travel to Banff for the Workshop, representing nine different nationalities; while 18 people were able to join virtually for most of the sessions.

2 Invited talks

Some participants were invited to give brief talks during the time span of the Workshop, covering topics from the history of the two projects up to how researchers and professionals use these and other open-source tools for carrying out their research and work in industry and government agencies.

Some personal reflections on a decade of SimPEG

Lindsey Heagy, one of the founders of SimPEG, gave her own personal reflections on how the project got started by a small group of PhD students, motivated by a desire to work together and an interest in creating shared tools that implement all the pieces necessary to simulate and invert data for a range of geophysical applications. She also shared some insights on what it takes to create a movement, build a community and inspire other people to join.

The Fatiando a Terra Project

During this talk, Leonardo Uieda introduced the Fatiando a Terra project to all participants, providing details about its origins in Brazil, created by a group of PhD students, how it evolved in time along with a growing open-source geoscientific ecosystem, and set future goals for the project. One of these goals was to nurture the collaboration with SimPEG by jointly developing and maintaining tools that can be used by the two projects.

Complete Open-source Software Processing Workflow: From Field Data to Publication

Craig Miller showcased the workflow he and his coworkers developed in GNS Science (New Zealand) for processing, inverting, and interpreting gravity and magnetic data using open-source tools like gsolve, Fatiando a Terra and SimPEG. Craig also listed the advantages of relying on open-source software for generating this type of workflows: the value of the communities around open-source projects, the level of customization of these tools, the transparency of their algorithms, the time-saving aspect of not having to build tools from scratch, and the capacity to combine these different tools in the same framework.

Using open-source tools to bridge the gap between Geology and Geophysics

During her talk, Andrea Balza Morales presented her advances on her latest research projects, which involve methodologies for bringing together open-source tools for geology and geophysics, to produce geologically plausible models that also fit the geophysical data. Her work is oriented towards geothermal energy applications, with the goal of improving the efficiency and safety of these types of operations. In these projects, she used GemPy [3], an open-source package for geological implicit modelling, to create geologic models based on the known geology of the study area, and physical properties obtained through drilling. She then used pyGIMLi [9], an open-source Python package for geophysical modelling and inversion, to carry out deterministic inversions that adjust the position of some pilot points, that ultimately modify the initial geological model to adjust to the observed geophysical data. She also made extensive use of SimPEG for running forward simulations, and Fatiando a Terra for processing and interpolation of gravity data.

Building 3D quasi-geology models and predicting mineral resources using joint inversion and open-source code

Jiajia Sun presented the results of one of his latest projects. Using SimPEG to run a set of L_p norm joint inversions of gravity and magnetic data, he was able to generate quasi-geology models by applying a probabilistic geologic differentiation technique. From the entire set of recovered models by different L_p norm joint inversions, the probabilistic geologic differentiation classifies each cell of the 3D mesh as belonging to one of a few different rock units, based on its recovered physical property values. The probabilistic differentiation assesses the uncertainty of the classification by considering a whole set of recovered models. Finally, he

showed the results of applying this technique to a real dataset over North Decorah, Iowa, US, showing its usefulness in localizing units of interest for mineral prospectivity.

Results from a survey on open-source software in the industry

At the end of the Workshop, Sean Walker joined virtually to present the results of a survey he ran to collect information about the usage of open-source tools in the industry. The survey questions brought insight into which open-source tools professionals in the industry are currently using, how relevant are they in their workflows, what value they see in them, and why they aren't using them in certain applications. The survey results showed that members of industry see a lot of value in open-source tools. Most of the survey participants rely on them for their daily tasks. The barriers to extending their usage are related to the pre-existence of workflows that make use of proprietary software, the lack of programming proficiency of their teams, and the lack of graphical user interfaces (GUIs) for some of these tools.

3 Group Discussions

In-person participants were split into groups of four people each, while virtual participants formed their own group. During the Discussions Sessions, each group was provided with a set of guiding questions to foster the discussion around three main topics:

- improving the communication within our communities,
- future developments of our projects, and
- increasing the diversity and sustainability of the projects.

Each group was encouraged to capture all outputs of these discussions and report them back to the other participants after each Discussion Session was finished. Based on the ideas developed during these sessions, the Workshop Organizers compiled a selection of actionable tasks. On the last day of the Workshop, participants chose one of those actionable items to start working on them, forming new sets of groups. In-person and virtual participants were encouraged to work and collaborate together within the groups.

In the following sections, we'll summarize the main discussion points for each topic, and list the progress made during the Workshop to develop solutions and take action after the identified issues and new ideas.

3.1 How can we improve the communication within our communities?

Participants identified avenues for improvement in the various communication channels of the two communities and came up with different ways to solve them. These strategies include improving the documentation of the software libraries, creating more user-friendly resources like comprehensive tutorials and examples, and exploring various outreach methods like newsletters, YouTube videos, and workshops. These initiatives aim to foster better communication, streamline collaboration, and ensure valuable information reaches community members, effectively preventing information fragmentation.

Amongst the other ideas that were discussed were:

- creating pages for frequently asked questions (FAQs) in the documentation of the software packages;
- establishing cross-links between the websites of the two projects pointing to relevant information;
- expanding the "Getting Started" guides;
- creating a comprehensive guide for new developers;
- scheduling periodic joint meetings between the communities of Fatiando a Terra and SimPEG;
- establishing office-hours for helping users and contributors to solve issues and ask questions;
- collecting all scientific articles that make use of the two projects and upload a list of them to the projects' websites;

- and extending the references to literature in the functions and classes of the packages, especially to provide details about the mathematical and physical backgrounds of their implementations.

3.2 What future research can be enabled by SimPEG and Fatiando?

Participants discussed the types of software tools that might be needed to perform current and further research, and which features are missing in the current packages of both projects.

One of the main discussion points was expanding the capabilities of Fatiando tools to process potential fields data from raw observations, such as drift and base station corrections, tide correction, and data levelling amongst others. Participants also discussed:

- optimizing terrain corrections using rectangular prisms for high-resolution elevation models;
- implementing geometry inversions for undulating layers like Mohorovicic discontinuity [11], the basement of sedimentary basins, and bathymetry under ice sheets;
- and FFT-based inversions like the Parker-Oldenburg method [8, 7].

Regarding SimPEG, participants proposed: improving and extending the magnetotellurics inversions; improve the ability to interoperate with other open-source packages that implement other forward models (e.g. ground-penetrating radar, empymod [14]); enabling the use of p-norms for the data misfit terms; and further developing advanced inversion methods that include geological and petrophysical knowledge (e.g. PGI [1]).

Other ideas were discussed, like speeding up the gravity and magnetic simulations in SimPEG using Choclo (one of the Fatiando a Terra libraries); developing functionalities for procedural model generation; exploring stochastic inversions; and smoothing the interoperability between the two projects.

3.3 How can we increase the diversity of our communities?

All groups started the discussion by acknowledging that there are many different types of diversity: cultural, social, racial, gender, sexual orientation, disability, by technical proficiency, amongst others. Considering this diversity spectrum, participants discussed strategies to increase the involvement of people from underrepresented groups, why this is a major goal for our communities, and proposed to write a Diversity Statement to establish the communities' scope and commitment to increase the diversity within them.

Participants came up with actionable tasks for lowering barriers that some underrepresented groups find while approaching these two projects. Amongst them we can find: translating tutorials and essential documentation pages to other languages to lower the language barrier for non-English speakers; implementing mechanisms of inclusion and accessibility that other organizations usually make use of, like accessible colormaps, alt-text for figures, and more accessible documentation websites; optimize the software tools so they can be used on low-end computers, allowing users without access to high-performance computational resources to use them.

Groups also explored the opportunity of reaching existing programs and organizations geared to underrepresented groups and undeveloped regions of the world for collaborations, like Geoscience Without Borders, SEG Women Networking Committee, GeoLatinas, Earth Science Women's Network, Women in Mining, the Equator project, sub-committees in professional associations (e.g. American Geophysical Union, European Geophysical Union). They also proposed extending our participation in other open-source software communities, like the Software Sustainability Institute, pyOpenSci, and Open Science Labs.

They explored the possibility of creating mentorship programs to train new developers and contributors from underrepresented groups, pairing mentee and mentors that have common backgrounds; hosting local workshops and events in the native language of the participants; organizing workshops, meet-ups and sessions at major geoscientific conferences to engage with a broader audience.

3.4 How to make these projects more sustainable?

All groups agreed that in order to promote long-term sustainability for the two projects, two key aspects must be addressed: (i) identifying avenues to increase representation from a wider variety of industry and

academic institutions, and (ii) new generation of developers and contributors should be trained. Aligned with the diversity discussions, people from underrepresented groups should be encouraged to make contributions to both projects and trained to become developers and maintainers. One actionable task that came up during these discussions is to improve the guides to contribute in the documentation pages of the two projects: adding more clear and updated instructions on how to set up a development environment; providing details and examples on code style and design; giving insight on how to write and execute tests; build the documentation; and ultimately, requesting contributions to be included in the main branch of the source code.

4 Outcome of the Meeting

The discussion that participants held during the workshop helped to identify points of improvement for the future of the SimPEG and Fatiando a Terra projects. The main topics were around future developments for the software packages within the two projects; how to promote more mutual collaboration and interoperability between the two of them; and how to lower barriers to increase diversity within our communities and train a new generation of developers and maintainers of these open-source tools. Since the participants were mostly graduate students and early-career researchers, most of the discussions were about how can we start contributing back and make our communities more open to people from underrepresented groups.

From all the discussions, some actionable items were selected and participants formed new groups to start working on them. Among them we can find: the creation of a comprehensive list of scientific publications that make use of SimPEG and Fatiando a Terra, as a way to promote the work of authors that rely on these tools and to provide an overview of the impact these two projects have in the geoscientific community; improving the guidelines for contributing to the two projects, by extending the documentation pages that help developers to setup the development environment and contribute with new code; move SimPEG tutorials to their own website built using JupyterBooks.

The discussions also led to some more long-term goals for the two projects, like developing a new joint tutorial that covers processing some geophysical data up to running an inversion using both Fatiando a SimPEG; promoting the collaboration between the two projects, for example by speeding up SimPEG's potential fields simulations with Fatiando's tools; scheduling joint periodic virtual meetings between the two projects to promote more interactions between the two communities; define joint and better roadmaps for the two projects; and keep improving the guides for new developers to join.

This workshop proved to be an excellent opportunity for the communities of Fatiando a Terra and SimPEG to get together and create new bonds that will trigger future scientific and technical collaborations. It provided graduate students and early-career researchers the room to share their perspectives and show their motivations and expectations for the future of these two projects. Covering from new developments that will help their current and future research and industry applications, to considering ways to build stronger and more diverse communities with better communication channels, their discussions were fruitful and enlightening.

References

[1] Thibaut Astic, Lindsey J Heagy, and Douglas W Oldenburg. Petrophysically and geologically guided multi-physics inversion using a dynamic gaussian mixture model. *Geophysical Journal International*, 224(1):40–68, aug 2020.

[2] Rowan Cockett, Seogi Kang, Lindsey J. Heagy, Adam Pidlisecky, and Douglas W. Oldenburg. SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications. *Computers & Geosciences*, 85:142–154, dec 2015.

[3] Miguel de la Varga, Alexander Schaaf, and Florian Wellmann. GemPy 1.0: open-source stochastic geological modeling and inversion. *Geoscientific Model Development*, 12(1):1–32, jan 2019.

[4] Fatiando a Terra Project, Chris Dinneen, Mariana Gomez, Lu Li, Agustina Pesce, Santiago Rubén Soler, and Leonardo Uieda. Boule v0.4.1: Reference ellipsoids for geodesy and geophysics, October 2022.

[5] Fatiando a Terra Project, Federico D. Esteban, Lu Li, Vanderlei C Oliveira Jr, Agustina Pesce, Nicholas Shea, Santiago R. Soler, Matthew Tankersley, and Leonardo Uieda. Harmonica: Forward modeling, inversion, and processing gravity and magnetic data, August 2022.

[6] Lindsey J. Heagy, Rowan Cockett, Seogi Kang, Gudni K. Rosenkjaer, and Douglas W. Oldenburg. A framework for simulation and inversion in electromagnetics. *Computers & Geosciences*, 107:1–19, oct 2017.

[7] Douglas W. Oldenburg. The inversion and interpretation of gravity anomalies. *GEOPHYSICS*, 39(4):526–536, August 1974.

[8] R. L. Parker. The rapid calculation of potential anomalies. *Geophysical Journal International*, 31(4):447–455, March 1973.

[9] C. Rücker, T. Günther, and F. M. Wagner. pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers and Geosciences*, 109:106–123, 2017.

[10] Leonardo Uieda. Verde: Processing and gridding spatial data using Green's functions. *Journal of Open Source Software*, 3(29):957, 2018.

[11] Leonardo Uieda and Valéria C.F. Barbosa. Fast nonlinear gravity inversion in spherical coordinates with application to the South American Moho. *Geophysical Journal International*, 208(1):162–176, oct 2016.

[12] Leonardo Uieda, Vanderlei C. Oliveira Jr, and Valéria C. F. Barbosa. Modeling the earth with Fatiando a Terra. In Stéfan van der Walt, Jarrod Millman, and Katy Huff, editors, *Proceedings of the 12th Python in Science Conference*, pages 96 – 103, 2013.

[13] Leonardo Uieda, Santiago Soler, Rémi Rampin, Hugo van Kemenade, Matthew Turk, Daniel Shapero, Anderson Banihirwe, and John Leeman. Pooch: A friend to fetch your data files. *Journal of Open Source Software*, 5(45):1943, January 2020.

[14] Dieter Werthmüller. An open-source full 3D electromagnetic modeler for 1D VTI media in Python: empymod. *Geophysics*, 82(6):WB9–WB19, 2017.