Fatiando a Terra: Open-source tools for geophysics

Santiago R. Soler\* and Lindsey J. Heagy

Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia

Summary

During this talk we will present the Fatiando a Terra project, a collection of open-source Python libraries designed for geophysical applications. We will describe how the project started as a simple library part of a PhD Thesis in South America, how it created a community around it that actively collaborates to its development and its current state. We will introduce the tools available in the project and show real world examples of how they can be used to solve geophysical problems. Finally we will discuss some of the current challenges and mention the upcoming features and development plans for the future.

Introduction

The relation between geophysical research and computational solutions dates back to the very origins of silicon-based computers. Scientists and industry saw in computational power a mean to solve new challenging problems within geosciences and exploration. They offer the possibility to process large amount of data, generate visualizations for interpretation and ultimately to perform inversions to build models of the subsurface. All of which are key to improve decision making processes.

Companies and researchers started developing their own software to perform these types of tasks. Although most of them built in-house tools, the appearance of open-source software for geosciences happened as early as the 70s and 80s. Projects like Seismic Unix (Stockwell, 1999) and GMT (Wessel et al., 2019) are pioneer examples.

Fast-forwarding to this millennia, the popularity that the Python language developed among every scientific field planted the seed for an ever-growing open-source geoscientific ecosystem. Python tools aimed to solve geoscientific problems proliferated, specially since the beginnings of the second decade of the century.

In this scenario Fatiando a Terra (Uieda et al., 2013; https://www.fatiando.org) was born: a project for developing open-source Python tools for geophysics. It started in 2010 in South America as a simple Python library as part of the PhD Thesis of Leonardo Uieda, but due to its open-source license and good online documentation it started being used by researchers and industry consultants from different regions of the world. As more people started to use it, some of them became contributors by writing new features, fixing bugs, adding more test or improving the documentation. The open-source nature of the project allowed it to grow under a community-driven development.

Nowadays the project consists in a set of Python libraries for geosciences, each one of them with a very specific scope of application. Its main goals are to provide open-source software tools that are easy to use and also well designed, tested and documented. During this talk we'll provide an overview of the tools available in the project and demonstrate their functionalities using examples from research and industry applications.

The Project

In its origins, Fatiando a Terra consisted in a single Python library named “fatiando” (https://legacy.fatiando.org) that used to host all the features that the project offered. After years of development this designed was proved to be flawed: its code base grew too large for a single library, making it hard to maintain and extend its functionalities. Moreover, it hosted a wide variety of functionalities: from production ready tools up to toy problems meant to be used for educational purposes.

In 2018 the project decided to redesign its code base by splitting the old “fatiando” library into several smaller libraries, each one with a very specific scope. This simplifies both the adoption of the libraries and also their development. Most users look for only a subset of the tools offered by the project, and having them divided in libraries reduces the size of the libraries they depend on. As a side effect, anyone interested in changing the code of one of our libraries now needs to familiarize with a smaller code base, making it easier for the community to contribute to the project.

By that time the geoscientific Python ecosystem had already seen a major growth. Libraries like SimPEG (Cockett et al., 2015), GemPy (de la Varga et al., 2019), pyGIMLi (Rücker et al., 2017) and ObsPy (Obspy, 2019) were already established and providing scientists and industry with a wide range of tools for research and exploration. The project decided to invest on extending this growing geoscientific ecosystem, rather than reinventing features already developed by other packages. The introduction of smaller and narrow scoped libraries allowed other projects to use on them, avoiding having a single large dependency for only a portion of its features.

The project is currently formed by five libraries: Verde, Boule, Harmonica, Pooch and Ensaio.

**Verde**

Verde hosts tools for spatial data processing, interpolation and gridding. Its core interpolation methods are inspired by machine learning, hence its interface reassembles the one of the popular Scikit-learn (Pedregosa et al., 2011). Besides, it offers analysis tools that accompany the interpolators, like trend removal, blocked or windowed operations, cross-validation, k-folding, grid projection and more coordinates manipulation utilities.

**Boule**

Boule is a very thin library that hosts classes for representing geodetic reference ellipsoids for the Earth and for celestial bodies of the solar system like the Moon, Mars, Venus and Mercury. These classes also offer methods to perform coordinate conversions between geodetic and geocentric spherical systems, and to compute the normal gravity generated by these ellipsoids on any external point through a closed-form analytic solution (Li and Götze, 2001).

**Harmonica**

Harmonica offers functions and classes for processing and modelling gravity and magnetic data. It hosts functions for forward modelling the gravity fields of point sources, rectangular prisms and also tesseroids (a.k.a spherical prisms). It can perform gravity corrections from a simple Bouguer to a full terrain correction through forward modelling digital elevation models with prisms. Regular grids can be transformed using FFT-based filters like upward derivative, upward continuation and reduction to the pole, among others. It also offers ways to perform interpolation, gridding and upward continuation through the equivalent sources technique. Lastly, but not least, it can also read data stored in popular formats like `.gdf` files provided by the ICGEM Calculation Service (Barthelmes, 2013) and `.grd` files from Oasis Montaj©.

**Pooch**

The most general purpose library in the project is Pooch, which offers an easy to use interface for downloading and caching data from the web. Originally designed for scientific applications and to be used by other software packages, Pooch can download data from the web through a large range of protocols, cache it locally at a desired location and also check the integrity of those files. This thin library is currently being used by other projects in the scientific Python stack, like SciPy (Virtanen et al., 2020), scikit-image (van der Walt et al., 2014), MetPy (May et al., 2016) and icepack (Shapero et al., 2020), among others.

**Ensaio**

Lastly we introduce Ensaio, a small library that hosts open licensed datasets that are useful for running examples and tutorials, for teaching and for probing our codes. It actually uses Pooch under the hood to download and cache those datasets locally so its codebase ends up being very slim. All the gallery examples and user guides of Harmonica and Verde use Ensaio to fetch the example datasets.

Example: processing gravity data

We can process gravity data using some of the Fatiando a Terra libraries, along with some other packages from the scientific Python stack like Numpy (Harris et al., 2020), Xarray (Hoyer & Hamman, 2017) and Pandas (McKinney, 2010; The Pandas Development Team, 2023).

For this example we are going to use an open gravity dataset over the Bushveld Igneous Complex in Southern Africa (made available by NOAA NCEI). The gravity data and a topography grid for the area are available for download through Ensaio. We will then use Pandas and Xarray to load the downloaded files into our Python script:

The downloaded gravity data consists in values of the observed gravity. The first step would be to compute the gravity disturbance by removing the normal gravity from the observed data. We can use Boule to calculate the normal gravity that the reference ellipsoid generates on every observation point (see Fig. 1(a)).

import ensaio

import pandas as pd

import xarray as xr

path\_gravity = ensaio.fetch\_southern\_africa\_gravity(

version=1

)

path\_topography = ensaio.fetch\_earth\_topography(version=1)

data = pd.read\_csv(path\_gravity)

topography = xr.load\_dataarray(

path\_topography

)

The gravity disturbance is mainly governed by the effect of the topographic masses. We can remove this effect by forward modelling the topography. We will first create a model of the topography using rectangular prisms, assigning a density contrast to each one of them:

import boule as bl

ellipsoid = bl.WGS84

normal\_gravity = ellipsoid.normal\_gravity(

data.latitude,

data.height\_geometric\_m

)

gravity\_disturbance = (

data.gravity\_mgal – normal\_gravity

)

And then compute their gravitational effect on the observation points. By removing the terrain effect from the disturbance we obtain the Bouguer gravity disturbance (see Fig. 1(b)).

import harmonica as hm

density = np.where(

topography\_geometric > 0, 2670,

1040-2670

)

topography\_model = hm.prism\_layer(

coordinates=(

topography\_geometric.easting,

topography\_geometric.northing

),

surface=topography\_geometric,

reference=0,

properties={"density": density},

)

Since the Bouguer disturbance is governed by the effect of deep anomalous sources, we are going to separate the residual field from the regional field using deep equivalent sources (see Fig. 1(c)).

coordinates = (

data.easting\_m,

data.northing\_m,

data.height\_geometric\_m

)

terrain\_effect=topography\_model.prism\_layer.gravity(

coordinates, field="g\_z"

)

gravity\_bouguer = (

gravity\_disturbance – terrain\_effect

)

Finally, we will create a regular grid of the residual field using shallower equivalent sources (see Fig. 1(d)). The observed high values correlate with the igneous intrusions present in the Bushveld Complex.

# Define deep equivalent sources

deep\_sources = hm.EquivalentSources(

damping=1000, depth=500e3

)

deep\_sources.fit(

coordinates, gravity\_bouguer

)

# Compute regional field

gravity\_regional = deep\_sources.predict(

coordinates

)

gravity\_residual = (

gravity\_bouguer – gravity\_residual

)

The full code for running this example in addition with more detailed explanation of the process can be found in https://www.fatiando.org/tutorials.

import pyproj

eq\_sources = hm.EquivalentSources(

damping=10, depth=10e3

)

eq\_sources.fit(

coordinates, gravity\_residual

)

# Define grid coordinates and grid the data

grid\_coords = vd.grid\_coordinates(

region=region,

spacing=2 / 60,

extra\_coords=2200,

)

# And a mercator projection

projection = pyproj.Proj(

proj="merc", lat\_ts=data.latitude.mean()

)

# Grid the residuals

residual\_grid = eq\_sources.grid(

coordinates=grid\_coords,

data\_names=["gravity\_residual"],

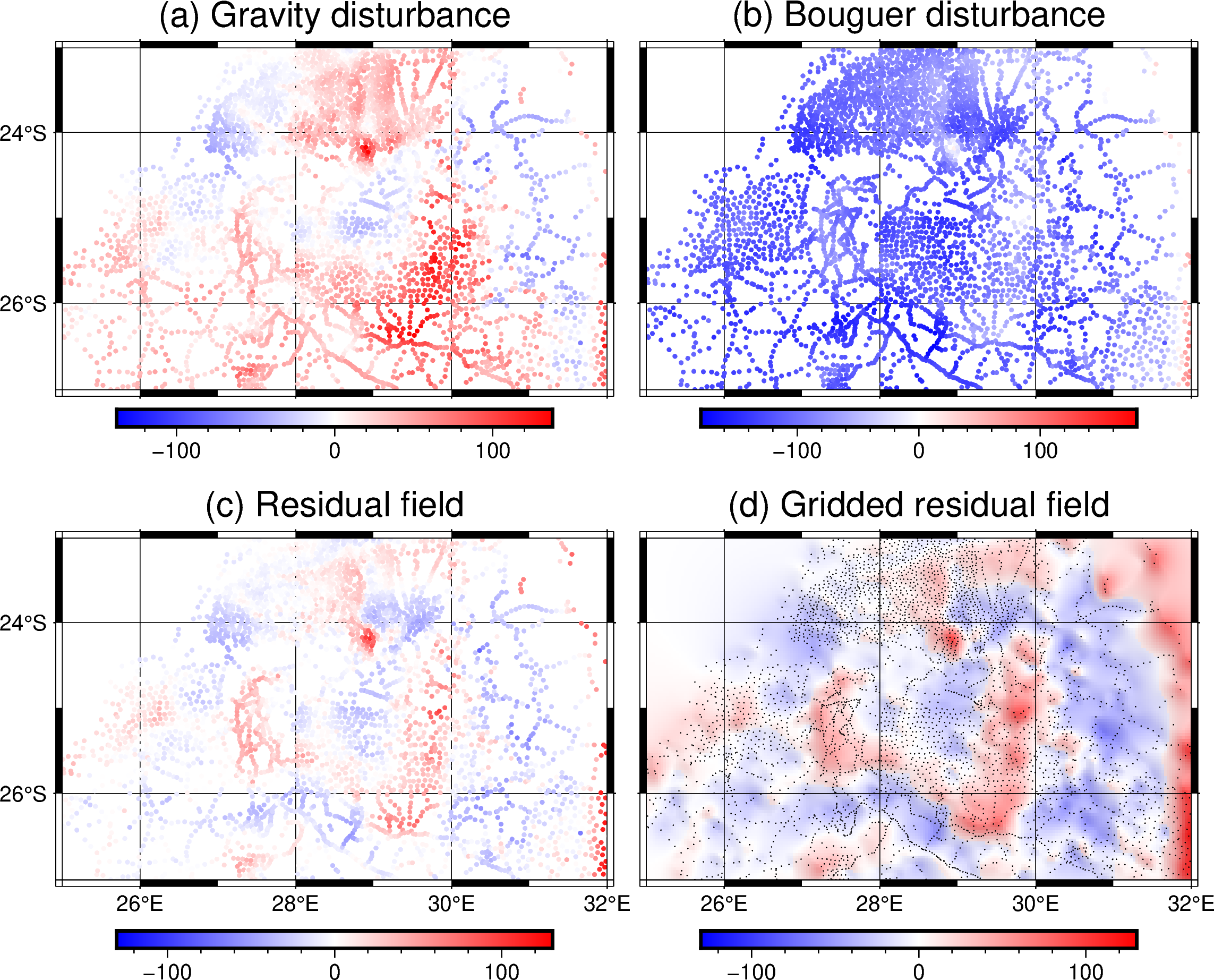
dims=("latitude", "longitude"),

projection=projection,

)

Conclusions

After more than a decade since them Fatiando a Terra project was born, the project lives thanks to a network of collaborators and users that actively participate in its development. It keeps proving that a collaborative and community-driven software development for science and industry is possible under an open-source environment. Its development turns out to be useful not only for their direct users, but also to the whole ecosystem: by providing tools that other projects can depend on and by setting an example of how to develop and maintain open-source geoscientific tools.

Figure 1.Fields obtained in different stages of the gravity data processing workflow for the Bushveld Igneous Complex: (a) gravity disturbance on the observation points obtained after removing the normal gravity from the observed gravity; (b) Bouguer gravity disturbance obtained after removing the terrain effect from the disturbance; (c) residual gravity field produced by removing the regional field using deep equivalent sources; (d) grid of the residual gravity field generated with equivalent sources.

Fatiando has been used by students as researchers as core parts of their Thesis and scientific articles, but also by the industry. This highlights the importance of nursing open-source tools that people are willing to invest time in learning and harnessing for their specific goals.

The future of the project expects exciting new features: a new magnetic forward modelling functions are being developed for Harmonica along with new types of equivalent sources and FFT-based transformations. The inversion framework for the old `fatiando` library is missing in the new tools, but we are looking forward to implement an improved one, focusing on performing types of inversions that are missing in other packages of the geophysical Python ecosystem.

Acknowledgments

We are indebted to the developers and maintainers of all the open-source software and particularly to the users and contributors of Fatiando a Terra. Without their gratitude, effort and contributions this project wouldn't exist.

References

Barthelmes, F. (2013). Definition of functionals of the geopotential and their calculation from spherical harmonic models: Theory and formulas used by the calculation service of the International Centre for Global Earth Models (ICGEM); http://icgem.gfz-potsdam.de/ICGEM/ ; revised edition. Scientific Technical Report; 09/02; ISSN 1610-0956. https://doi.org/10.2312/GFZ.B103-0902-26

Cockett, R., Kang, S., Heagy, L. J., Pidlisecky, A., & Oldenburg, D. W. (2015). SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications. Computers & Geosciences, 85, 142–154. https://doi.org/10.1016/j.cageo.2015.09.015

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled Arrays and Datasets in Python. Journal of Open Research Software, 5(1), 10. https://doi.org/10.5334/jors.148

Li, X., & Götze, H. (2001). Ellipsoid, geoid, gravity, geodesy, and geophysics. GEOPHYSICS, 66(6), Article 6. https://doi.org/10.1190/1.1487109

May, R., Arms, S., Marsh, P., Bruning, E., Leeman, J., Bruick, Z., & Camron, M. D. (2016). MetPy: A Python Package for Meteorological Data. UCAR/NCAR-Unidata. https://doi.org/10.5065/D6WW7G29

McKinney, W. (2010). Data Structures for Statistical Computing in Python. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

The ObsPy Development Team (2019). ObsPy 1.1.1 (1.1.1). Zenodo. https://doi.org/10.5281/ZENODO.1040770

The Pandas Development Team (2023). pandas-dev/pandas: Pandas (v1.5.3). Zenodo. https://doi.org/10.5281/ZENODO.3509134

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Rücker, C., Günther, T., & Wagner, F. M. (2017). pyGIMLi: An open-source library for modelling and inversion in geophysics. Computers & Geosciences, 109, 106–123. https://doi.org/10.1016/j.cageo.2017.07.011

Shapero, D., Badgeley, J., & Hoffman, A. (2020). icepack: Glacier flow modeling with the finite element method in Python (v1.0.0). Zenodo. https://doi.org/10.5281/ZENODO.1205640

Stockwell, J. W. (1999). The CWP/SU: Seismic Un∗x package. Computers & Geosciences, 25(4), 415–419. https://doi.org/10.1016/S0098-3004(98)00145-9

Uieda, L., V. C. Oliveira Jr, and V. C. F. Barbosa (2013), Modeling the Earth with Fatiando a Terra, Proceedings of the 12th Python in Science Conference, pp. 91 - 98.

de la Varga, M., Schaaf, A., & Wellmann, F. (2019). GemPy 1.0: Open-source stochastic geological modeling and inversion. Geoscientific Model Development, 12(1), 1–32. https://doi.org/10.5194/gmd-12-1-2019

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: Image processing in Python. PeerJ, 2, e453. https://doi.org/10.7717/peerj.453

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. Geochemistry, Geophysics, Geosystems, 20(11), 5556–5564. https://doi.org/10.1029/2019GC008515