# PyGMI - a python package for geoscience modelling and interpretation

Patrick Cole [1]

1 Council for Geoscience, South Africa

Summary

Python Geoscience Modelling and Interpretation (PyGMI) is an open-source software development project, working with a GUI, programmed in Python. The main aim of PyGMI is to provide a free geoscientific tool for processing, interpretation and modelling of geoscience data, including magnetic, gravity and remote sensing data. It also allows for research into better ways to process, analyse and model data and as such is continually updated to implement these advances.

Although is was originally intended to create voxel based three dimensional potential field models, it expanded to include tools for the processing and visualisation of raster data, import and visualisation of vector data (normally collected in surveys along lines), unsupervised and supervised cluster analysis, gravity processing, remote sensing processing and quality control tools for earthquake seismology. It is therefore a toolbox of numerous techniques to accomplish this. Examples of use include modelling of this Bushveld Igneous Complex (Cole et al., 2021) and remote sensing in the Namaqualand region, South Africa (Musekiwa, 2023).

Statement of Need

Geophysical data provide geoscientists with the ability to create a picture of the subsurface prior to more expensive endeavours such a drilling or for further research using complementary data sets. The best way to present such a picture is through 3D models. The primary need for the PyGMI project was to provide open source modelling and image processing capabilities through a GUI. Geophysically, its strength is GUI based forward modelling, with limited inversion. It makes use of scientific libraries, such as SimPEG, (Cockett et al., 2015) and scikit-learn, (Pedregosa et al., 2012) and is curently a GUI based alternative to libraries such as GemPy (Varga et al., 2019), PyGimli (Rücker et al., 2017), and Harmonica (Terra Project et al., 2024).

PyGMI Functionality

The original concept of PyGMI was a desire to have an effective 3D forward modelling package for magnetic and gravity data. There are a number of basic strategies to do this. One conventional approach follows the modelling of individual profiles, which are then "joined" up into 3D sections. Another is to form bodies from arbitrary polygonal shapes. However, using such methods, the time it takes to construct a 3D model can be prohibitive. Conventional polygonal based 3D modelling programs (3DS Max, Maya, Blender) are akin to CAD packages and require a steep learning curve, which is not always possible or tolerated by many scientists. PyGMI follows a voxel based approach (which is also used commonly in geophysical inversion), allowing the model to be 'drawn' in on model slices or layers, much like using an art program. Each voxel is assigned to a lithology from a user defined list, with its associate geophysical definitions.

There are many techniques which can be considered here (Bhattacharyya, 1964; Guptasarma & Singh, 1999; Holstein, 2003; Singh & Guptasarma, 2001a, 2001b). In this case the technique by Bhattacharyya (1964) was used, being most applicable for voxel modelling (Figure 1). The simplicity of the technique makes it well suited to rectangular prism calculations. It is described and developed into a Fortran routine, named 'mbox' by Blakely (1995) (pp. 200-201).
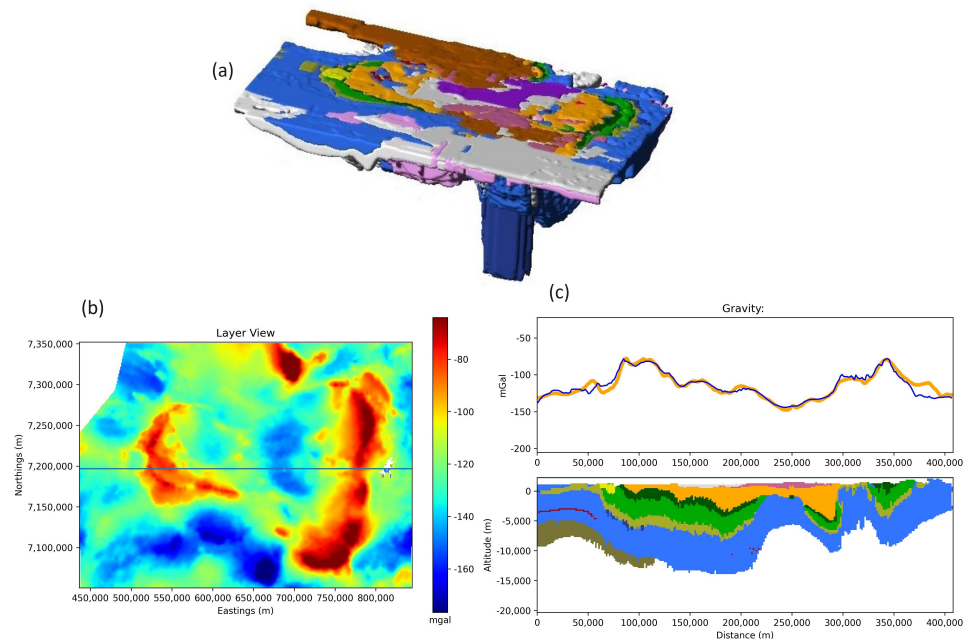
Figure 1: (a) An example of a model created with PyGMI. (b) Top view of the model, the horizontal line is the profile being modelled. (c) The side view of the profile being modelled. The lines above the model represent the observed and calculated data.

The PyGMI interface has been designed in a flow chart manner. This enables the user to see each step of the processing flow, useful in both debugging a process or teaching to new users. It also allows greater flexibility between modules.

Standard raster functions such as equations (making use of the NumExpr library, https://github.com/pydata/numexpr/), smoothing, normalisation, merging of datasets and reprojections (courtesy of GDAL/OGR contributors, 2020; Gillies & others, 2013) are included. Functions unique to potential fields such as tilt angle (Cooper & Cowan, 2006), visibility (Cooper, 2005), reduction to the pole, sun shading (Horn, 1981), IGRF calculations (Erwan et al., 2015, based on code by written by A. Zunde, USGS, S.R.C. Malin & D.R. Barraclough, Institute of Geological Sciences, United Kingdom and maintained by Stefan Maus, NOAA) have all been translated into or developed in python. The sun shading tool in particular allows for sunshade detail and light reflectance to be changed (Figure 2). These two parameters are not normally present in other software packages which usually only allow for the changing of sun dip and azimuth. Vector data, such as magnetic line data or geochemical data, can be imported courtesy of the GeoPandas library (Jordahl et al., 2020).
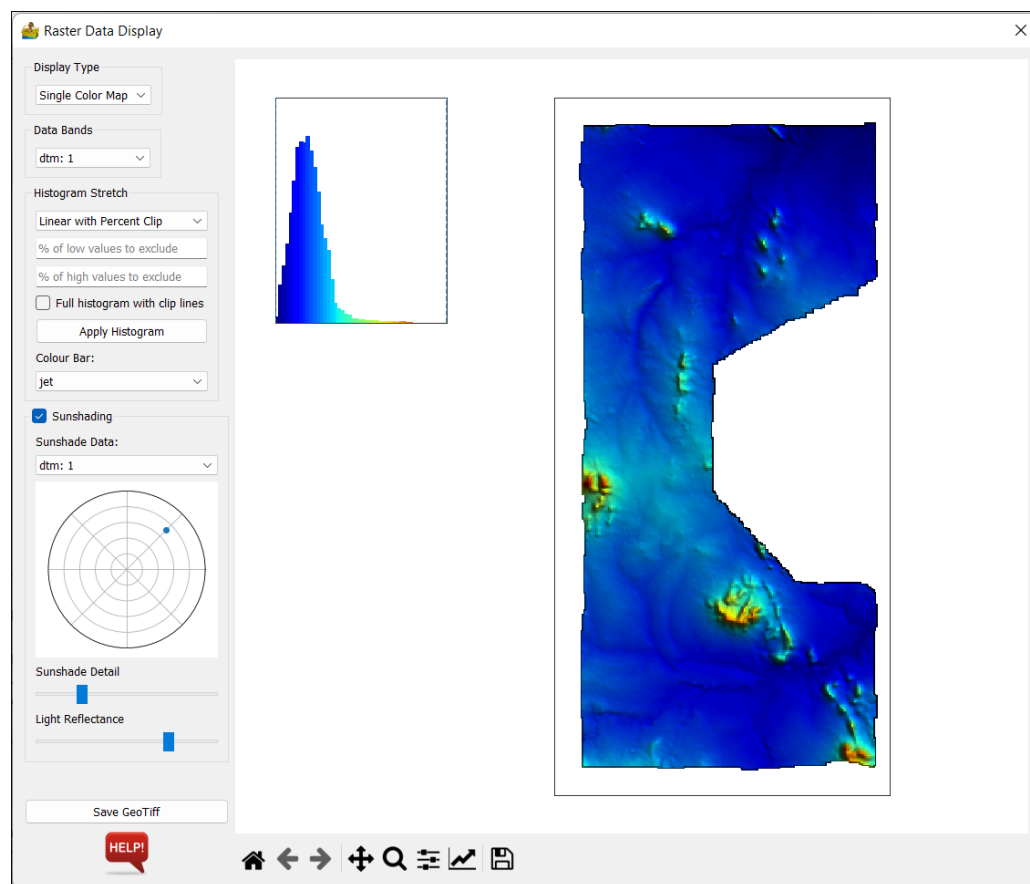
Figure 2: Sunshading being displayed on the raster interpretation module.

Gravity data processing has been developed according to the North American gravity database standards, as described by Hinze et al. (2005). It allows for data to be imported from Scintrex CG-5 gravimeters and basic processing to Bouguer anomaly to be performed.

Earthquake seismology QC tools for the open source SEISAN platform (Havskov et al., 2020) have been implemented. These include various plotting functions including RMS of time residuals, histograms of events vs time, b-values and error ellipses (Figure 3). Filtering of events can be performed via a filtering tool. A fault plane solution tool is also implemented, translated from MATLAB code written by Andy Michael, Chen Ji and Oliver Boyd and is capable of exporting results to the shapefile format.
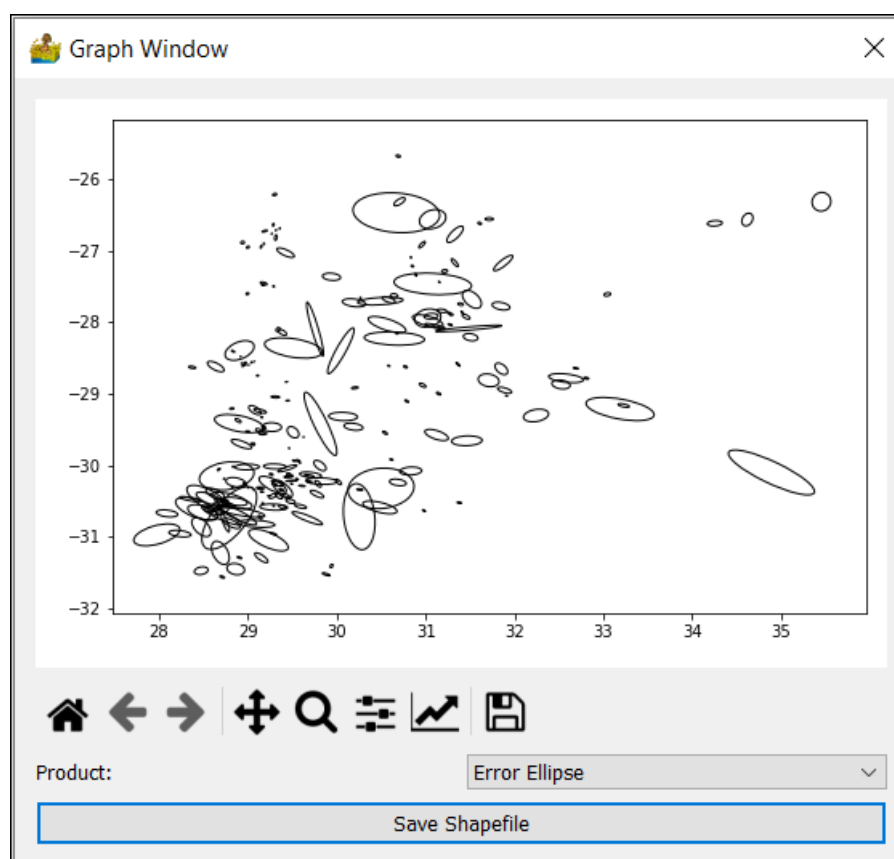
Figure 3: Error ellipses of seismic events shown on the QC tool.

Crisp and fuzzy cluster analysis routines developed by Paasche & Eberle (2009) in a joint project with the Council for Geoscience are provided, as well as supervised and unsupervised routines developed via the Sklearn library (Pedregosa et al., 2012). A graphical tool to see class representations via a scatterplot allows for relationships between scatterplots and the data to be examined interactively. Image segmentation following the method by Baatz & Schäpe (2000) has also been implemented.

Remote sensing tools are one of the major new focuses of PyGMI. Numerous major satellite formats can be read through the GDAL library via the rasterio library (GDAL/OGR contributors, 2020; Gillies & others, 2013), including Sentinel-2, ASTER and Landsat, to name a few. PyGMI's focus is to enable convenient calculation of a variety of measures, including predefined band ratios, condition indices, PCA (principal component analysis) and MNF (minimum noise fraction, Green et al., 1988). Change detection indices are also implemented as well as a viewer for change data visualisation. Finally, hyperspectral data analysis through feature detection is also possible (Haest et al., 2012).

The PyGMI project therefore aims to continue to grow and support research and development in geosciences by providing easy to use tools with a GUI frontend.

PyGMI Resources

Various resources including documentation and wiki are available:

- PyGMI Repository - https://github.com/Patrick-Cole/pygmi/
- PyGMI Documentation - https://patrick-cole.github.io/pygmi/

- PyGMI Installation Instructions - https://patrick-cole.github.io/pygmi/install.html/

Acknowledgements

References

Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XII, Beiträge Zum AGIT Symposium*, 12–23.

Bhattacharyya, B. K. (1964). Magnetic anomalies due to prism-shaped bodies with arbitrary polarization. *Geophysics*, *29*, 517–531. https://doi.org/10.1190/1.1439386

Blakely, R. J. (1995). *Potential theory in gravity and magnetic applications*. Cambridge University Press, Cambridge (UK), 441 pages. ISBN: 9780521575478

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*. https://doi.org/10.1016/j.cageo.2015.09.015

Cole, J., Finn, C. A., & Webb, S. J. (2021). Geometry of the bushveld complex from 3D potential field modelling. *Precambrian Research*, *359*, 106219. https://doi.org/10.1016/j.precamres.2021.106219

Cooper, G. R. J. (2005). Analysing potential field data using visibility. *Computers and Geosciences*, *31*, 877–881. https://doi.org/10.1016/j.cageo.2005.03.002

Cooper, G. R. J., & Cowan, D. R. (2006). Enhancing potential field data using filters based on the local phase. *Computers & Geosciences*, *32*, 1585–1591. https://doi.org/10.1016/j.cageo.2006.02.016

Erwan, T., Finlay, C., Beggan, C., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T., Boness, A., Brocco, L., Daix Canet, E., Chambodut, A., Chulliat, A., Coïsson, P., Civet, F., Du, A., Fournier, A., Fratter, I., Gillet, N., & Zvereva, T. (2015). International geomagnetic reference field: The 12th generation. *Earth, Planets and Space*, *67*, 79. https://doi.org/10.1186/s40623-015-0228-9

GDAL/OGR contributors. (2020). *GDAL/OGR geospatial data abstraction software library*. Open Source Geospatial Foundation. https://gdal.org

Gillies, S., & others. (2013). *Rasterio: Geospatial raster i/o for Python programmers*. Mapbox. https://github.com/rasterio/rasterio

Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, *26*, 65–74. https://doi.org/10.1109/36.3001

Guptasarma, D., & Singh, B. (1999). New scheme for computing the magnetic field resulting from a uniformly magnetized arbitrary polyhedron. *Geophysics*, *64*, 70–74. https://doi.org/10.1190/1.1444531

Haest, M., Cudahy, T., Laukamp, C., & Gregory, S. (2012). Quantitative mineralogy from infrared spectroscopic data. I. Validation of mineral abundance and composition scripts at

the rocklea channel iron deposit in western australia. *Economic Geology*, *107*, 209–228. https://doi.org/10.2113/econgeo.107.2.209

Havskov, J., Voss, P. H., & Ottemöller, L. (2020). Seismological Observatory Software: 30 Yr of SEISAN. *Seismological Research Letters*, *91*(3), 1846–1852. https://doi.org/10.1785/0220190313

Hinze, W. J., Aiken, C., Brozena, J., Coakley, B., Dater, D., Flanagan, G., Forsberg, R., Hildenbrand, T., Keller, G. R., Kellogg, J., Kucks, R., Li, X., Mainville, A., Morin, R., Pilkington, M., Plouff, D., Ravat, D., Roman, D., Urrutia-Fucugauchi, J., … Winester, D. (2005). New standards for reducing gravity data: The north american gravity database. *Geophysics*, *70*, J25–J32. https://doi.org/10.1190/1.1988183

Holstein, H. (2003). Gravimagnetic anomaly formulas for polyhedra of spatially linear media. *Geophysics*, *68*, 157–167. https://doi.org/10.1190/1.1543203

Horn, B. K. P. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*(1), 14–47. https://doi.org/10.1109/PROC.1981.11918

Jordahl, K., Bossche, J. V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *Geopandas/geopandas: v0.8.1* (Version v0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Musekiwa, D., C. (2023, September). Multispectral and hyperspectral remote sensing for geology mapping: Namaqualand region, south africa and namibia. *29th Colloquium of African Geology*.

Paasche, H., & Eberle, D. G. (2009). Rapid integration of large airborne geophysical data suites using a fuzzy partitioning cluster algorithm: A tool for geological mapping and mineral exploration targeting. *Exploration Geophysics*, *40*(3), 277–287. https://doi.org/10.1071/EG08028

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., Duchesnay, E., & Louppe, G. (2012). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*.

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

Singh, B., & Guptasarma, D. (2001a). Joint modelling of gravity and magnetic fields - a new computational approach. *Current Science*, *81*, 1626–1628. http://repository.ias.ac.in/28190/1/317.pdf

Singh, B., & Guptasarma, D. (2001b). New method for fast computation of gravity and magnetic anomalies from arbitrary polyhedra. *Geophysics*, *66*, 521. https://doi.org/10.1190/1.1444942

Terra Project, F. a, Castro, Y. M., Esteban, F. D., Li, L., Oliveira Jr, V. C., Pesce, A., Shea, N., Soler, S. R., Souza-Junior, G. F., Tankersley, M., Uieda, L., & Uppal, I. (2024). *Harmonica v0.7.0: Forward modeling, inversion, and processing gravity and magnetic data* (Version 0.7.0). Zenodo. https://doi.org/10.5281/zenodo.13308312

Varga, M. de la, Schaaf, A., & Wellmann, F. (2019). *GemPy 1.0: open-source stochastic geological modeling and inversion* (Version 1.0). https://doi.org/10.5194/gmd-12-1-2019