

# GeoBO: Python package for Multi-Objective Bayesian Optimisation and Joint Inversion in Geosciences

Sebastian Haan

14 November 2020

## Summary

A critical decision process in geophysical data collection is how to efficiently combine a variety of sensor types and where to collect new observations, in particular if measurements are very costly or resources are limited. This may include drill-core placements or the layout of a geophysical survey. Bayesian Optimisation (BO) solves this decision making problem by finding global optimal solutions given multiple, often distinct measurements and model uncertainties. One of the major challenges for applying BO in geoscience still lies in constructing a 3D model from sparse, typically 2D sensor data that can be computationally efficiently evaluated and that transforms the combined data and uncertainties from multiple sensor measurements coherently into a posterior function approximating the true model; this is also known as an inversion problem.

**GeoBO** is a Python package that solves both, Multi-Objective Optimisation and Joint Inversion, within one probabilistic framework: from prior selection and input, data fusion and inversion, to the final sensor optimisation and real world model output. Fig.1 shows a graphical overview model of **GeoBO**; the two main process steps are summarized in the following.

First, **GeoBO** jointly solves multi-linear forward models of 2D survey data (e.g. magnetics and gravity) to 3D-geophysical properties using Gaussian Process kernels (see, e.g., Rasmussen and Williams 2006; Melkumyan and Ramos 2009), without the requirement of prior geological knowledge. In addition, sparse drillhole measurements can be taken into account. One of the key features is to consider correlations between geophysical properties by solving simultaneously multiple forward models (Melkumyan and Ramos 2011; Reid et al. 2013). Thus, this approach provides a better constrained combined solution of multiple, distinct sensor types, rather than taking their individual solutions. Another practical advantage of this probabilistic approach is that predictions are described by posterior distributions for each location (voxel or cube cell), which quantify the uncertainty in the predictions and their credible intervals. Most other work in the field of joint inversion only report point estimates of the quantities of interest (see, e.g., Zeyen and Pous 1993; Gallardo-Delgado, Pérez-Flores, and Gómez-Treviño 2003; Moorkamp et al. 2011).

While this inversion method is limited to linear forward models, the reconstructed posterior distribution of the geophysical properties can be further refined using site-specific knowledge or more complex inversion methods for sampling non-linear forward models, such as **hIPPYlib** (Villa, Petra, and Ghattas 2018), **Obsidian** (Scalzo et al. 2019), or **GemPy** (Varga, Schaaf, and Wellmann 2019).

In a second step, the BO in **GeoBO** allows the user to define objectives in an acquisition function (see

for BO methods, e.g., Mockus 1975; Brochu, Cora, and De Freitas 2010; Sebastian Haan 2020) which typically has to balance between a) exploration, i.e. querying points that maximise the information gain and minimize the uncertainty of a geophysical site model; b) exploitation, i.e. querying points that maximise the reward (e.g. concentrating search in the vicinity locations with high value such as minerals); and c) minimize the number of samples given a cost function for any new measurement (e.g. costly drill-cores). The reconstructed 3D model output of the joint inversion model is then used to query for the next most promising point based on the acquisition function, which guides the search for a user-defined optimum.

GeoBO allows applications to specify priors as additional options, such as a) the typical correlation lengthscale of the geological structure via GP hyperparameters, b) the gain/cost parameters in the BO acquisition function, and c) the correlation coefficients between different geophysical properties. This package includes forward models for gravity and magnetics surveys (Yaoguo and Oldenburg 1998), drillcores, and test-sets of geological models and data.

While it has been tested on solving the problem of allocating iteratively new expensive drill-core samples (Sebastian Haan 2020), the same method can be applied to a wide range of optimisation and sensor fusion problems.

## Installation, Dependencies and Usage

GeoBO can be installed using `setuptools` or `pip` and requires the following python packages: `numpy`, `matplotlib`, `scikit_image`, `scipy`, `rasterio`, `pandas`, `pyvista`, `skimage`, `PyYAML` (for details see `requirements.txt`). The installation can be tested by running the script with included synthetic data and default settings.

To use GeoBO, change the main settings such as filenames and parameters in `settings.yaml` (see description there), and run

```
cd geobo/
python main.py settings.yaml
```

Core features of GeoBO:

- Joint probabilistic inversion tool by solving simultaneously multi-linear forward models (e.g. gravity, magnetics) using cross-variances between geophysical properties (cross-variance terms can be specified by use in `settings.yaml`)
- Computation of complete posterior distribution for all geophysical properties (described by their mean and variance value at each location/voxel)
- Generation of ranked proposal list for new most promising drillcores (optional: non-vertical drillcores) based on global optimisation of acquisition function (see `settings.yaml`)
- Templates for acquisition function to use in Bayesian Optimisation (see ‘futility’ functions in `run_geobo.py`)
- Flexible parameter settings (`settings.yaml`) for exploration-exploitation trade-off and inclusion of local 3D cost function in acquisition function (see function ‘`create_costcube`’ in `run_geobo.py`)

Other features are:

- Package includes geological survey/drillcore sample as well as synthetic data and functions for synthetic data generation

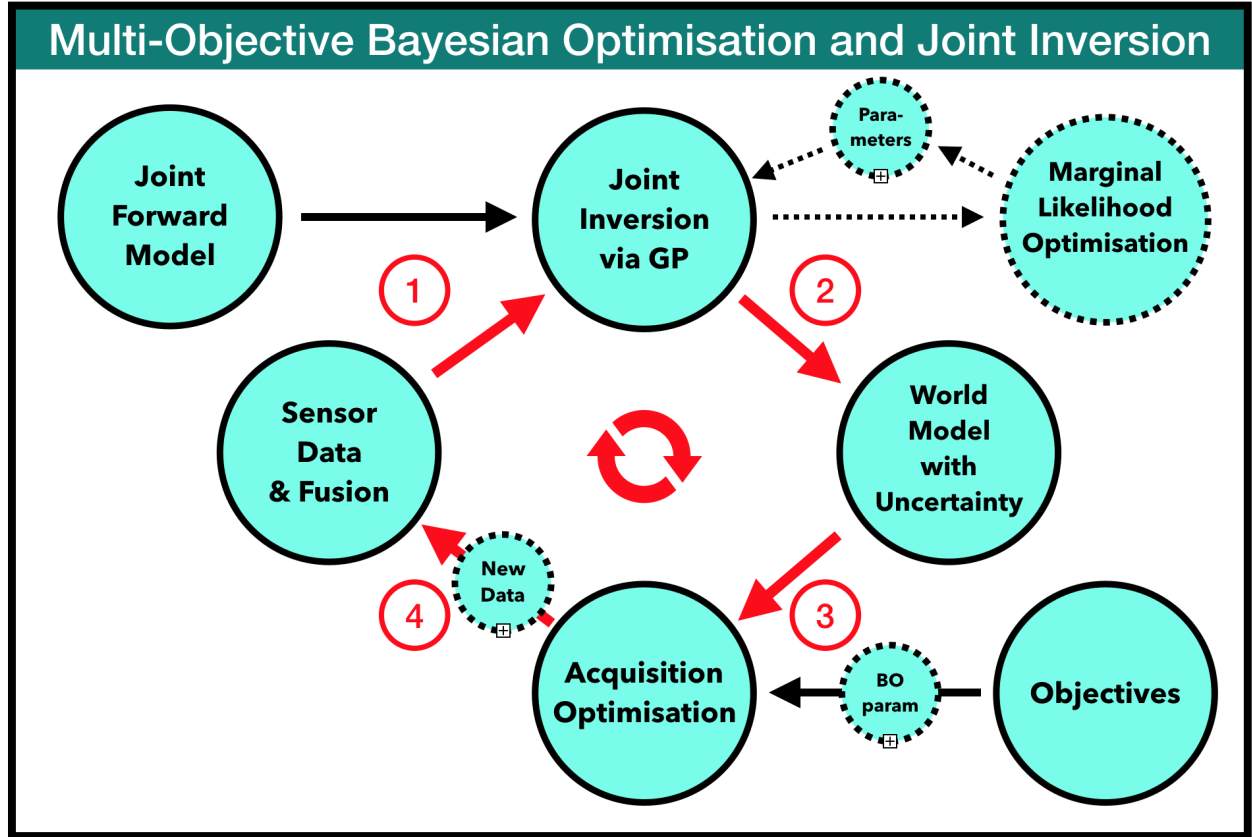


Figure 1: Overview of the probabilistic framework for multi-objective optimisation and joint inversion; the process stages are: 1) Joint inversion via Gaussian Processes based on multiple sensor data fusion and forward model, 2) Posterior model generation of 3D multi-geophysical properties. 3) Maximisation of acquisition function to allocate optimal new sensor location and type. 4) New acquired data is combined with existing data; process repeats until maximum number of iterations is achieved. For more details see Sebastian Haan (2020)

- Generation of 2D/3D visualisation plots of reconstructed cubes and survey data (see plot settings in settings.yaml)
- 3D Cube export in VTK format (for subsequent analysis, e.g., with ParaView)
- Options to include any pre-existing drillcore data (see settings.yaml for feature names as well as format of sample drilldata)
- Included linear forward models: density-to-gravity and magnetic susceptibility-to-magnetic field; custom linear forward models can be added (sensormodel.py)
- Library of Gaussian Process (GP) kernels including sparse GP kernels (kernels.py)
- Flexible settings for any cube geometry and resolution (cube length/width/height and voxel resolutions)
- Optional marginal GP likelihood for optimization of GP hyperparameters and inversion process

## Results and Output

First, the joint inversion generates the reconstructed properties and their uncertainties as 3D cubes, such as density and magnetic susceptibility. Other 3D physical properties can be obtained by adding custom forward models (see section *Custom Linear Forward Models* in README). Secondly, Bayesian Optimisation uses these probabilistic inversion results to query new potential measurements based on a specific set of objectives as defined in an acquisition function, which guides the search for a user-defined optimum. Finally a list of all new measurement proposals ranked from maximum to minimum improvement is generated including a map of the most promising new measurement locations.

## Acknowledgements

Sebastian Haan would like to acknowledge Dietmar Muller, Fabio Ramos, and Ben Mather for their very valuable contributions to this project. This research was supported by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

## References

- Brochu, Eric, Vlad M Cora, and Nando De Freitas. 2010. “A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning.” *arXiv Preprint arXiv:1012.2599*.
- Gallardo-Delgado, Luis A, Marco Antonio Pérez-Flores, and Enrique Gómez-Treviño. 2003. “A versatile algorithm for joint 3D inversion of gravity and magnetic data.” *Geophysics* 68 (3): 949–59. <https://doi.org/10.1190/1.1581067>.
- Melkumyan, Arman, and Fabio Ramos. 2009. “A Sparse Covariance Function for Exact Gaussian Process Inference in Large Datasets.” In *IJCAI*, 9:1936–42.
- Melkuyman, Armon, and Fabio Ramos. 2011. “Multi-Kernel Gaussian Processes.” In *IJCAI*, 22:1408. 1.
- Mockus, J. 1975. “On Bayesian methods of optimization.” *Towards Global Optimization*, 166–81.
- Moorkamp, Max, Björn Heincke, Marion Jegen, Alan W Roberts, and Richard W Hobbs. 2011. “A framework for 3-D joint inversion of MT, gravity and seismic refraction data.” *Geophysical Journal International* 184 (1): 477–93. <https://doi.org/10.1111/j.1365-246X.2010.04856.x>.

- Rasmussen, Carl Edward, and Christopher KI Williams. 2006. *Gaussian Process for Machine Learning*. MIT press.
- Reid, Alistair, O’Callaghan Simon Timothy, Edwin V Bonilla, Lachlan McCalman, Tim Rawling, and Fabio Ramos. 2013. “Bayesian Joint Inversions for the Exploration of Earth Resources.” In *IJCAI*, 2877–84.
- Scalzo, Richard, David Kohn, Hugo Olierook, Gregory Houseman, Rohitash Chandra, Mark Girolami, and Sally Cripps. 2019. “Efficiency and Robustness in Monte Carlo Sampling for 3-d Geophysical Inversions with Obsidian V0. 1.2: Setting up for Success.” *Geoscientific Model Development* 12 (7): 2941–60. <https://doi.org/10.5194/gmd-12-2941-2019>.
- Sebastian Haan, Dietmar Muller, Fabio Ramos. 2020. “Multiobjective Bayesian Optimisation and Joint Inversion for Active Sensor Fusion.” *Geophysics*, 86(1), Pp.1-78. <https://doi.org/10.1190/geo2019-0460.1>.
- Varga, Miguel de la, Alexander Schaaf, and Florian Wellmann. 2019. “GemPy 1.0: Open-Source Stochastic Geological Modeling and Inversion.” *Geoscientific Model Development*. <https://doi.org/10.5194/gmd-12-1-2019>.
- Villa, U., N. Petra, and O. Ghattas. 2018. “hIPPYlib: an Extensible Software Framework for Large-scale Deterministic and Bayesian Inverse Problems.” *Journal of Open Source Software* 3 (30). <https://doi.org/10.21105/joss.00940>.
- Yaoguo, Li, and Douglas W Oldenburg. 1998. “3D-Inversion of Gravity Data.” *Geophysics* 63 (1): 109–19.
- Zeyen, Hermann, and Jaume Pous. 1993. “3-D Joint Inversion of Magnetic and Gravimetric data with a priori Information.” *Geophysical Journal International* 112 (2): 244–56. <https://doi.org/10.1111/j.1365-246x.1993.tb01452.x>.