Bayesian data fusion for geothermal exploration

Ramos, F.T. [1], Bonilla, E. V. [2,3], McCalman, L. [2], O'Callaghan [2], S., Reid, A. [2], Uther, W. [2], Sambridge, M. [3], and Rawling, T. [4]
[1]University of Sydney, [2]NICTA, [3]Australian National University, [4]University of Melbourne
fabio.ramos@sydney.edu.au

This research is supported by the Australian Centre for Renewable Energy (ACRE) through a measure application.

Keywords: joint inversion, data fusion, Bayesian methods, exploration, risk minimisation

Project description

The process of finding the best locations for drilling in geothermal exploration requires the collection of vast amounts of information. Gravity, magnetism, seismic reflection, radiometric, magnetotellurics and drilling data are commonly used to infer specific properties of rocks such as temperature conductivity, permeability, porosity, density and fracture levels. The availability of the data does not necessarily ensure their best usage. From a statistical perspective, it is desirable to determine the degree of dependence between each of the data sources to fuse them in a single estimation framework for all the rock properties. This ensures the correct quantification of the predictive uncertainty, which leads to risk minimisation in drilling and better decision-making. This work aims to organise the information available in Australia for geothermal activities and develop a novel statistical framework for data fusion and joint estimation of rock properties relevant to the industry.

Methodology

The approach is divided into two main components; 1) the collection and organisation of the data in a data portal; and 2) statistical methods for join inversion and characterisation of geothermal targets. More details on these developments are presented below.

The Portal

A unique property of our machine learning techniques is that data from different sensors can be fused simultaneously and automatically. To facilitate this, we developed a portal system that could provide all available data from any sensors given a particular region in a single, machine-readable format suitable for the inference software.

Traditionally, datasets are stored in a variety of text-based and binary formats depending on the type of sensor (ArcGIS, ERS, SEG-Y etc.), in different projections and with idiosyncratic labelling systems. We saw the need for an open data format, which was flexible enough to allow different kinds of sensor or simulation data to be stored together and represented in a self-describing way. The format should be universal enough that a variety of open-source tools for reading it already existed. The HDF5 data format (www.hdfgroup.org) meets all these requirements and was chosen as the portal's output.

Communication with the portal occurs via a simple HTTP-based interface. Software can easily implement control of the portal directly, and we have also written a web front-end that allows the user to select, visualise, and download available datasets in a single HDF5 file, ready for input into our data fusion software. We also anticipate being able to serve and visualise results from our algorithms using this system.

Bayesian data fusion

The second phase of the project consists of developing the data fusion engine using statistical machine learning, in particular, nonparametric Bayesian methods (see e.g. Rasmussen and Williams, 2006). Nonparametric Bayesian methods are flexible and principled approaches to modelling uncertainty in real-world problems, allowing the incorporation of information from various sensor modalities and different types of prior knowledge. Moreover, these techniques have the advantage of automatically determining statistical dependencies between input measurements (e.g. observations of the gravitational and magnetic fields) to jointly estimate other parameters involved in the underlying physical process (e.g. density and magnetic susceptibility).

Our first step towards the development of a unified non-parametric Bayesian framework for data fusion in geothermal exploration is the construction of probabilistic models for geophysical inversion problems. Our goal here is twofold. Firstly, we want to develop more accurate joint inversion methods that allow us to reason about physical properties of the earth by exploiting commonalities between different sensor modalities (e.g. gravity, magnetics, seismic) and by using prior information (e.g. from geological studies of a specific region). Secondly, we aim at *quantifying the uncertainty* of the different variables in the process. For example, rather than placing hard constraints on parameters based on previous geological studies, we can introduce this information probabilistically in our models. More importantly, the output of our methods is, in general, a full distribution over the parameter of interest.

Our methods can avoid over-fitting (i.e. poor generalization, see e.g. Mackay, 1992) and enable the use of kernels that capture nonlinear relationships, providing a simple but effective platform for multi-task learning and sensor data fusion in spatial stochastic processes.

These methods have opened major opportunities to overcome the challenges by extending the applicability of machine learning techniques to large-scale exploration problems, integrating data acquisition and data acquisition in a principled and unifying statistical model.

Geophysical Inversion with Gaussian Processes

Given a set of measurements such as gravity, magnetics and seismic data, and the location where they were obtained, a Gaussian process (GP) prior is placed over the space of functions mapping spatially dependent geophysical quantities (such as density) to measurements. Our goal is to reason about these quantities probabilistically. Hence we perform inversion by conditioning the distribution of the corresponding property (or parameter) on the observations yielding a posterior distribution that balances prior knowledge and data fit. The result is a probability distribution over the inverted quantity (for example, density) for any point in space. The method produces continuous outputs with an associated uncertainty, which can be used, for example, to reason about optimal locations where to acquire more measurements or to figure out the locations where to carry out new explorations (drilling) so that a predefined risk that depends on the desirable properties is minimized.

Results to date

Portal

The portal is the pipeline that feeds geophysical and geological datasets into our data fusion algorithms. As such, the first stage of the pipeline is to convert incoming files into a unified internal representation suitable for further processing. We currently use HDF5 for this internal representation, though as the volume of data stored increases a distributed file system might be chosen. The file importers also add metadata attributes such as the bounding box, sensor type, creation date, and description to a database, used by the backend to facilitate queries and requests.

The portal backend connects to this database and uses it to serve requests for a collection of datasets in a particular region. The communication protocol is an HTTP-based REST interface, a stateless protocol using URL requests easily implemented in a variety of languages. When a request is received by the backend, it opens the associated HDF5 files of all requested datasets, crops them to the region of interest, then re-writes a new HDF5 file that contains all this data in a single file. That output file is then sent to the user as a download.

Even though we anticipate our software directly interfacing with the portal system, we have developed a web interface to facilitate members of the geothermal community exploring the available data. The interface has standard functionality such as an interactive map that can be panned and zoomed with the mouse. Geo-referenced visualisations of the datasets and their bonding boxes are automatically overlaid when selected. Regions of interest can be chosen on the map either by clicking and dragging with the mouse or entering co-ordinates directly. Check-boxes for each available dataset allow the user to determine the contents of the output file, which is collated on the server when the user hits 'download'. See Figure 1 for a screen-shot of the portal web interface.

The portal is currently populated with representative datasets from Geoscience Australia, state governments, and geothermal companies. Data acquisition and conversion into the portal system is ongoing, but datasets with existing support include all public gravity station data, surface geology and fault lines from all states, and high-resolution gridded magnetics, topographics and radiation. We have already acquired and are currently in the process of converting additional magnetics, 2- and 3-D seismic lines, magnetotellurics, depth-to-basement models, various drill-hole measurements and micro-seismic. We anticipate eventually collecting all available data relevant to the geothermal exploration problem in Australia.

Inversion of gravity and magnetics with drill hole data

In this section we present preliminary results of our nonparametric Bayesian inversion method on a synthetic experiment regarding gravity data when additional magnetic data are presented along with drill hole observations. The goal is to reason about the density of the underlying anomaly responsible for the observed data.

Figure 2 shows the results for a synthetic example of a density anomaly, the corresponding gravity observations and additional magnetic data and drill hole observations. These synthetic data have been generated with a linear forward model as in Li and Oldenburg (1998). Incorporating drill hole data in our model is straightforward as these can be considered as actual observations of the parameter of interest (i.e. density). Similarly, as our approach addresses probabilistic dependencies between the outputs (in this case gravity and magnetics) through the covariance (kernel) function, our technique provides a flexible way of performing joint inversions, where the relationships between the different modalities can be learnt from data.

We see that the constrained nonparametric Bayesian inversion, using drill hole observations and magnetic data, not only improves the quality of the mean of the density distribution but it also reduces the uncertainty around these predictions.

References

Li Y., and Oldenburg D.W., 1998. 3D inversion of gravity data. Geophysics, 63(1):109-119.

Mackay, D. J. C., 1992. Bayesian interpolation. Neural Computation, 4(3):415-447.

Rasmussen, C. E., and Williams C. K. I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts.

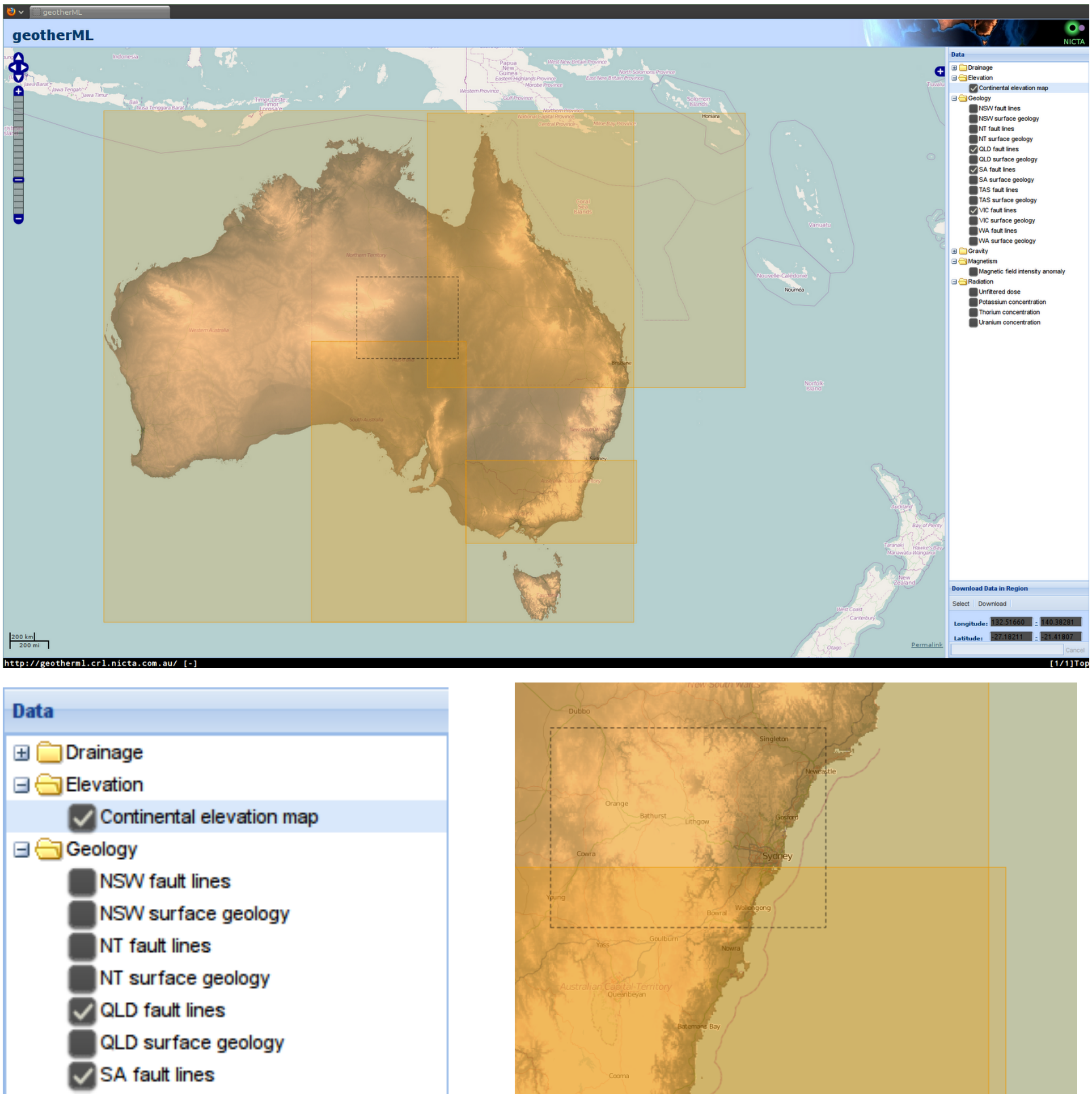

Figure 1: *Screenshots from the portal web interface. (Top) A figure depicting the portal web interface. A list of datasets is on the right, with check-boxes for selecting those in which the user is interested. The bottom right contains tools for selecting data in a region with the mouse, or entering lat/long coordinates manually. The central pane contains a zoomable visualisation including partially transparent shaded bounding boxes of the checked datasets, and a dotted line indicating the selected region. (Bottom left) A close-up of the list of datasets with checkboxes. (Bottom right) A close-up of a selected region on the map, with two overlapping bounding boxes and a region around Sydney currently selected.*

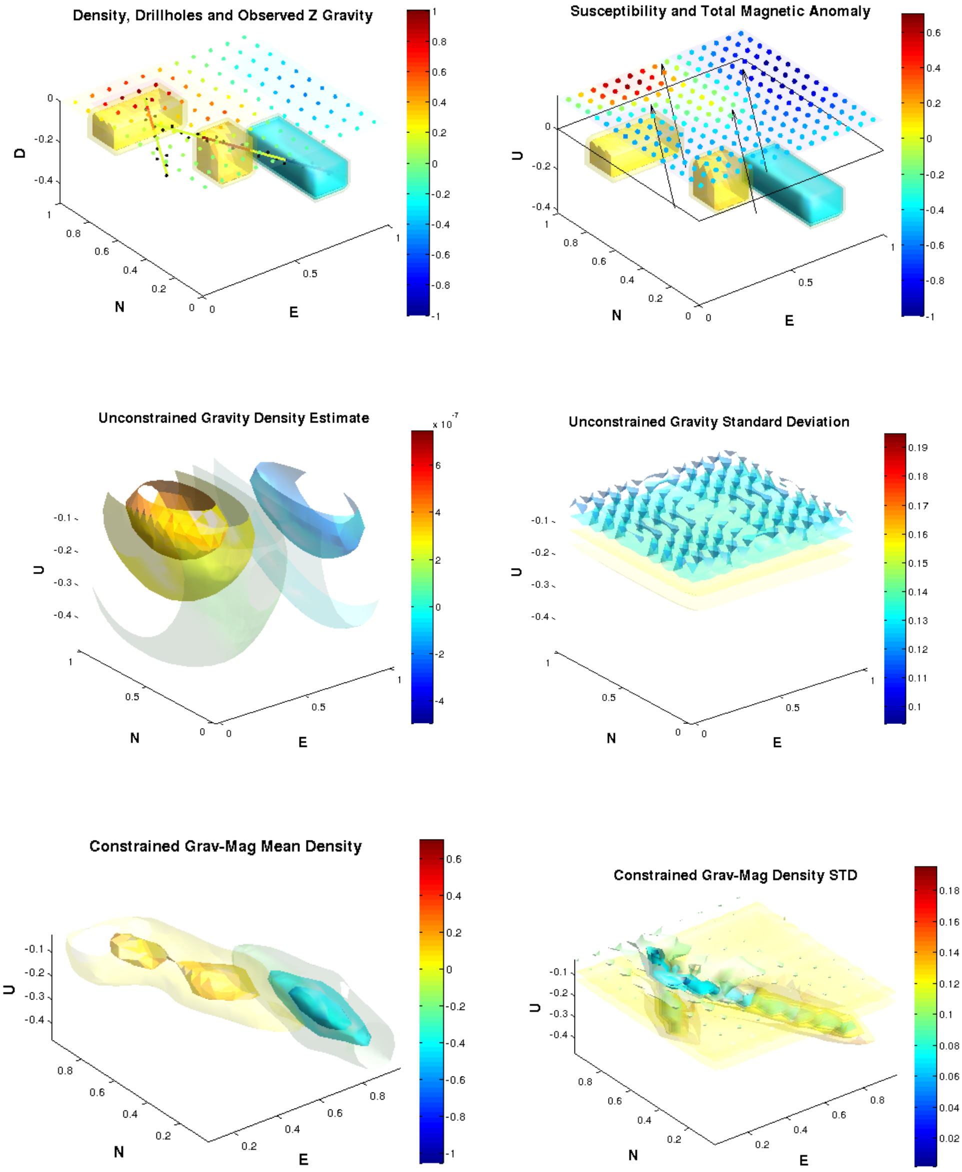

Figure 2: *Density inversion scenario using a Matérn covariance function. (Top left) The density anomaly along with the gravity observations and drill hole observations. (Top right) Forward simulated Total Magnetic Anomaly observations for the same body. (Middle) Results of the unconstrained nonparametric Bayesian inversion with only gravity data showing the mean of the density distribution (left) and its standard deviation (right). (Bottom) Results of the nonparametric Bayesian inversion using gravity data, magnetic data and drill hole data showing the mean of the density distribution (left) and its standard deviation (right).*