Bayesian Data Fusion for Geothermal Exploration: New Algorithms and Results

Simon T. O'Callaghan[2], Alistair Reid[2], Lachlan McCalman[2], Olena Velychko[2], Simon Carter[2,5], Lars Krieger[2,5], Michelle Salmon[2,3], Graeme Beardsmore[6], Edwin V. Bonilla[2,3], Malcolm Sambridge[3], and Tim Rawling[4], Fabio T. Ramos[1,2]

[1]University of Sydney, [2]NICTA, [3]Australian National University, [4]University of Melbourne, [5]University of Adelaide, [6]Hot Dry Rocks

fabio.ramos@nicta.com.au

This research is supported by the Australian Renewable Energy Agency (ARENA) through a measure application. NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

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

ABSTRACT

This project addresses the central question of identifying the best locations for drilling in geothermal target exploration from a probabilistic perspective. We develop algorithms to perform data fusion and joint inversions from a number of geophysical surveys to infer several rock properties such as density, magnetic susceptibility, seismic reflection, resistivity and thermal conductivity. Our methods can seamlessly integrate surveys including gravity, magnetics, seismic reflection, magnetotellurics and borehole geophysics to predict the most relevant properties for geothermal resource categorisation. As part of a probabilistic inference procedure, we also estimate uncertainties for each quantity of interest, providing principled strategies for risk minimisation in drilling projects.

This paper presents a number of updates to the project including modelling and inference using parallel tempering Monte Carlo techniques. The probabilistic approach to the inversion problem not only quantifies the risk associated with drilling, but will be able to assist in strategies for further sensing that optimally increase exploration confidence. The Bayesian formulation allows expert knowledge to be added in the form of prior distributions, making the contribution of human judgment to the final result explicit.

1. METHODOLOGY

In this paper we present a Bayesian approach to locating geothermal targets by probabilistically fusing data from multiple geophysical sensor modalities with drill-hole measurements to create a posterior distribution over the relevant geological properties.

1.1 Overview

Data from K distinct sensor modalities (such as magnetotelleurics, gravity, magnetism, seismic, and drill-holes) are considered, as components of the set $D = \{D_k\}_{k=1}^{K}$. These observations in turn relate to J spatially distributed geophysical rock properties: $g(x) = \left(g_1(x), \ldots, g_J(x)\right)$ (characterising properties including resistivity, density, susceptibility, elastic modulus and temperature).

The relationship mapping from rock properties to observations is constrained by well understood physical models, also called forward models. We denoted the model for sensor k as $f_k(g)$; each model computes an ideal (noise-free) form of the data corresponding to a particular world state, using principles such as Newton's law of gravitation, or the eikonal wave equation.

Our goal is then to reason about the distribution of rock properties given the set observations, corresponding to a geophysical inversion problem. Importantly, the inference engine does this probabilistically so as to consider different sources of uncertainty and provide support for decision-making in the process of geothermal energy exploration and characterization.

The key contributions of the developed approach are (a) considering all the modalities jointly, (b) inducing dependencies in our observations via hierarchical modeling, and (c) modeling the statistical dependencies between all the rock properties jointly. Our approach to probabilistic joint geophysical inversions therefore relies on the decomposition of the problem into the following parts: (i) a prior probability distribution over the geology in the region under study; (ii) a joint prior distribution over the rock properties; and (iii) likelihood models for the corresponding forward models given the rock properties. Figure 2 (right) shows the graphical model of our approach that will be discussed below.

1.2 Geological Prior: $P(\alpha)$, $P(c)$

Given our current knowledge of existing geothermal sites, the geological prior has been defined in two parts. The first is a prior over the spatial structure of an underlying 3D volume, while the second is a combinatorial prior over the possible rock types assigned to the parts of that structure. In particular, we have considered a layer-cake model where the prior over the structure defines different layers, which are determined by control points $\alpha \in \mathbb{R}^3$. This structural prior is denoted $P(\alpha)$.

One possible instantiation of this world model is shown below in Figure 1, although we note that the inference engine can accept any parameterised spatial function to describe the interfaces. In this case, the prior is strongly structured based domain knowledge of the target region - in this example, the model is based on the Cooper basin region. Thus the parametric model encodes knowledge that the region's geological stratigraphy consists of a basin-shaped partially metamorphosized basement overlain with multiple sedimentary layers. The basin is intruded by radiogenic granites rising from the Earth's upper mantle. These granites are of particular importance to the project because they provide a renewable heat source when insulated by the 3-5 kilometers of crust.

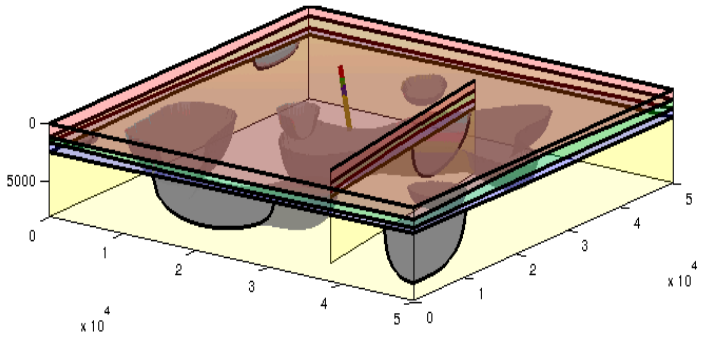


Figure 1: Geometry of an example parameterised world model. The top three layer interfaces are controlled by only 9 parameters each, while 64 parameters are dedicated to the granite intrusions. A drill-hole and a 2D slice section are also depicted within this volume.

The representation of the world as parametric interfaces allows the model to define geological structure, with segments of a particular class representing distinct rock types such as granite intrusions, or strata in a sedimentary basin. In simple terms, these regions between interfaces can be considered as a set of L layers whose corresponding rock types are unknown. Hence, to complete our geological prior, we define a distribution over rock types given a particular layer, $P(c_l)$, for $l = 1, \dots, L$.

A recent development in the project has been the progression from a prototype 2D geology representation to a 3D representation. The 3D representation offers more flexibility because, as well as allowing the use of full 3D physical models, it continues to support drills and 2D sensor models such as magnetotelleurics through the efficient extraction of 1D and 2D cuts (depicted in Figure 1).

1.3 Joint Prior over Rock Properties: $P(g)$

For each layer we define a non-spatial joint prior over rock properties. For a given geology as determined by $\boldsymbol{\alpha}$ and $\mathbf{c}$ this defines a joint prior over the corresponding spatial properties $\mathbf{g}$. The prior is drawn from a library of rock types that are determined through lab tests on extracted samples, combined with prior domain knowledge about the underlying material processes. The characterisation of multiple rock properties jointly is critical to being able to fuse multiple sensor types. A full rock property distribution spanning properties such as density, resistivity, susceptibility and seismic velocity is critical for both conducting joint inversions, and inferring unobserved properties such as fracturing. Despite its importance, the full joint distribution over properties has been largely unstudied in the literature beyond pairs of sensors. Consequently, new lab tests have been conducted as part of this project help characterise these distributions. This allows us to maintain a (non-spatial) low-dimensional distribution that, together with the geology, implicitly defines a distribution over the rock properties over the entire space. The implementation allows new rock types to be modeled by extending the library of distributions from which layer properties are drawn.

1.4 Sensor Likelihood Models: $P(D|g)$

As described above, each sensor modality D_k is constrained by its corresponding forward model $f_k(g)$. Therefore, it is sensible to assume that all sensor measurements D_k are conditionally independent given g. In other words: $P(D|g) = \prod_{k=1}^{K} P\big(D_k|f_k(g)\big)$.

Note that this is not actually a strong assumption even if our main goal is to carry out joint inversions as dependencies between observations are induced via the hierarchical structure of the model and additional statistical dependencies are exploited via the joint prior P (g).

The inference procedure uses forward models, these likelihoods map from a given world state g into the likelihood of observations D. The ideal measurements from a particular model are straightforward to compute for any sensor (linear or non-linear) provided the physical basis of the sensor is understood and can be simulated. The likelihood of D is more complicated, as it must consider how modeling and sensing errors impact on the distribution of the resulting measurement. For many sensors this can be approximated with a Gaussian distribution - this has been assumed for the models currently available.

The inference engine allows sensor likelihood models to be freely added or removed from a simulation. The library of implemented forward models currently includes gravity, magnetism, 2D seismic, a 3D thermal diffusion model and 2D magnetotelleuric models.

1.4.1 Gravity

Perturbations in the Earth's gravitational field are used to infer alterations in the local geology's density. For a specific set of geophysical spatial properties, the gravitational response on the surface can be determined exactly using Newton's gravitational law. Here we use Li's tractable approximation of the law as described in [8] for 3-D field of constant density prisms to map from a density field to a predicted gravity measurement $D_{pred}=f(g(x))$. $P(D|g)$ can then be calculated using the commonly employed Gaussian likelihood function, $P(D|g) = 2\pi^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(D-D_{\text{pred}})'\Sigma^{-1}(D-D_{\text{pred}})}$ where n is the number of gravity observations and Σ is the positive semi-definite noise matrix for the sensor.

1.4.2 Magnetism

Rocks in the Earth's crust have varying degrees of magnetic susceptibility, largely dependent on mineral iron content. While permeated by the Earth's magnetic field, the rocks have induced magnetic fields that can be detected on the Earths' Surface using precise magnetism sensors. While it is difficult to orient a sensor accurately enough to measure the three dimensional components, it is possible to measure perturbations in the total field magnitude, known as total magnetic anomaly (TMA). A susceptible volume of rock is usually modeled using infinitesimal dipoles. We employ the results presented by Li and Oldenburg [7] for analytically integrating the contributions to the TMA of the magnetic fields from these dipoles over a regular prism of constant susceptibility as a function of its position relative to the sensor (while neglecting any secondary permanent magnetism present in the rocks).

1.4.3 Reflection Seismic

Reflection seismology images the subsurface by generating compression pulses at ground level, and then using the timing of pulse reflections to infer locations of discontinuity in wave velocity. Seismic surveys are typically run along a line, in which an array of geophones simultaneously measures reflections from a single 'shot'. The results for multiple shots are often combined additively to increase the signal-to-noise ratio. We model seismic reflections using a combination of fast-marching wavefront methods and full-waveform finite-difference methods. We are currently developing a likelihood function based on extracting major reflectors from the signal.

1.4.4 Magnetotellurics

Magnetotellurics (MT) uses the naturally occurring electromagnetic fields at the surface of Earth to determine the subsurface resistivity structure. Data are analysed in the frequency domain, where the E- and B- fields are connected via the relation E=Z*B. For each frequency, the transfer function Z is a complex valued 2 x 2 matrix with elements comprised of the ratios of the electric and magnetic fields i.e., E_x/B_x, E_x/B_y, E_y/B_x and E_y/B_y. Similar to magnetism and gravity likelihood models, a Gaussian function is employed to determine the probability of the survey data given a particular geological instantiation [10]. The predicted MT response for a specific set of parameters is determined using the 2-D finite element forward model proposed by Wannamaker et al. [11].

1.4.5 Temperature

To model the production and redistribution of heat in the Earth's crust we consider the 3D heat transfer equation, which is a second-order linear differential equation of the form: $\nabla(\square(x,y,z)\nabla T) = -A(x,y,z)$, where ∇ is the gradient operator; T is the temperature; $\square(x,y,z)$ is the thermal conductivity; and A is the heat production rate. This assumes that there is no temporal change of the temperature profile (i.e. thermal steady state) and ignores heat transfer due to advection. Although there are some specific cases for which the heat transfer equation can be solved analytically, we solve this equation for temperature using finite

differences (see e.g. [12]). The boundary conditions are: (a) constant temperature at the surface; (b) constant heat flow at the bottom; and (c) zero heat flow on the sides of the volume under study. This forward temperature model provides the basis for our likelihood model, which assume Gaussian observations centered at this forward model with diagonal covariance.

1.5 Joint Likelihood

Putting together the hierarchical prior model and the appropriate sensor likelihood models, the joint distribution is fully defined by:

$$P(\alpha, c, \rho, D) = P(\alpha)P(c)\prod_{\ell=1}^{L} P(\rho_\ell)\prod_{k=1}^{K} P\big(D_k|f_k(g(\alpha, c, \rho))\big),$$

where the specific distributions we have adopted in our experiments are given by

$$\begin{aligned} P(\alpha) &= \text{Normal}(\alpha; \mu_\alpha, \Sigma_\alpha), \quad P(c_\ell) = \text{Multinomial}(c_\ell; \theta_\ell), \\ P(\rho_\ell|r) &= \text{Normal}(\rho_\ell; \mu_r, \Sigma_r), \quad P(D_k|g) = \text{Normal}\big(D_k; f_k(g(\alpha, c, \rho)), \sigma_k^2\mathrm{I}\big). \end{aligned}$$

Next we determine the posterior distribution over geological structures and properties given measurements from each of the K different modalities jointly. The difficulty of computing this posterior depends upon the complexity of the forward models considered. The modalities considered in our experiments range from very simple linear forward models for sensors such as gravity to quite elaborate non-linear forward models such as those involved in seismic data.

A graphical representation of this dependency structure is illustrated in Figure 2 (right). This formulation differs from the "traditional" geophysical inversion approach, illustrated in Figure 2 (left), [7, 8]. This traditional model contains no notion of geological units with different prior relationships between geophysical properties. As such, correlations between properties are constant spatially. Our structure function c allows for different property distributions for different units. Geological information such as formation age and composition can therefore be explicitly encoded.

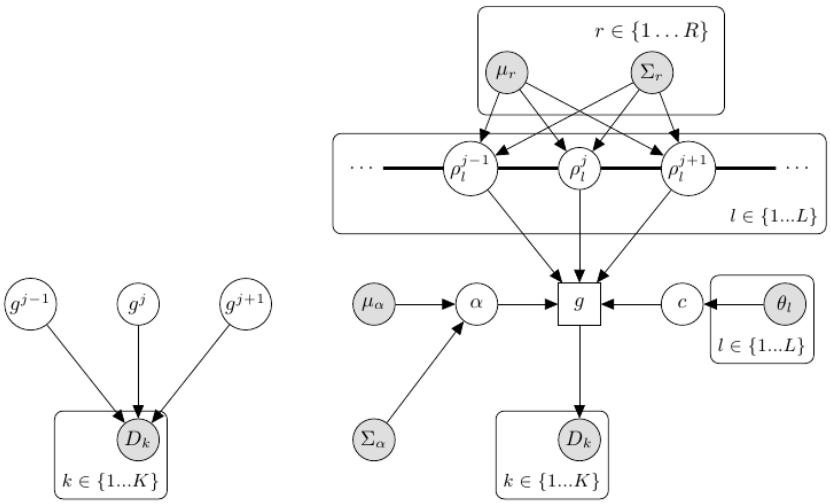


Figure 2: Probabilistic graphical model representations of traditional survey techniques (left) and the fusion procedure implemented in our inference engine (right). The traditional approach contains no notion of geological units with different prior relationships between geophysical properties, so correlations between properties are constant spatially. Our structure function c allows for different property distributions for different units. Geological structure can therefore be explicitly encoded. Additional dependencies are modeled by considered a joint prior P (ρ) over rock properties.

2. INFERENCE COMPUTATION

Nonlinear geophysical inversion problems typically involve probabilistic sampling from multimodal density functions. The ill-posed nature of the problem frequently leads to numerous diverse yet probable solutions all of which can satisfy the conditions imposed by the data. Two approaches have been explored in the project. The first was an analytical solution for the posterior,

leveraging a number of simplifying assumptions regarding the space of structural geologies considered and the suite of geophysical sensors employed (one rock type, linear forward models) [9]. The analytic approach (based on Gaussian process regression) provides tractability advantages, but its assumptions preclude the possibility of both complex geological features and the fusion of several highly informative, non-linear datasets such as seismography and magnetotellurics that require numerical simulation of partial differential equations to compute the forward relationship $f_k(g)$.

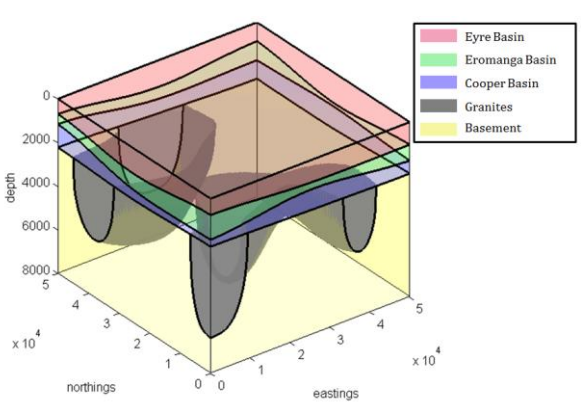
Consequently, a transition to a more flexible inference methods has been followed. In this case, a prototype inference engine based on parallel tempering MCMC inference has been implemented. This approach is able to use complex forward sensor models, and leverage the strong knowledge of geology coded into the parameterised priors.
Creating proposal distributions that enable Markov chains to efficiently traverse distributions with isolated modes is nontrivial [4]. The trade-off between a sufficiently broad proposal distribution to bridge the regions of low probabilities between modes and a tolerable acceptance rate of proposed states is difficult to achieve with standard MCMC techniques such as Metropolis Hastings (MH) [1]. Consequently, we employ a meta-algorithm known as parallel tempering [3] to aid with convergence. Intuitively, parallel tempering runs multiple chains of an arbitrary MH updated MCMC algorithm simultaneously at different 'sampling energies $\Psi = 1$ and $\Psi > 1$. A modified MH acceptance probability function $p_{\text{accept}} = \min\left(1, \exp\left[\left(\varphi(\theta_c) - \varphi\left(\theta_p\right)\right)/\Psi\right]\right)$ is used, where $\varphi(\theta_c)$ is the unnormalized log probability of the target distribution at the current state θ_c and θ_p is the proposed state, provides a mechanism for chains at higher energies to more easily move between modes due to the relaxed acceptance criteria. Lower energy chains enable more precise local sampling but are susceptible to becoming trapped in modes. Consequently, parallel tempering also incorporates an interchange procedure whereby two chains, i and j, at different energies can swap location in the state space with the probability defined as: $p_{\text{swap}} = \min\left(1, \exp\left[\left(\varphi\left(\theta_j\right)/\Psi_i - \varphi(\theta_i)/\Psi_j\right]/\exp\left[\left(\varphi(\theta_i)/\Psi_i - \varphi\left(\theta_j\right)/\Psi_j\right]\right)\right.$ which satisfies the detailed balance constraint of MCMC.

One final benefit of this inference method is its parallel architecture. Individual chains can be run almost independently across a large cluster of computing cores. Low bandwidth communication between chains is required only when a proposed swap in location is scheduled.

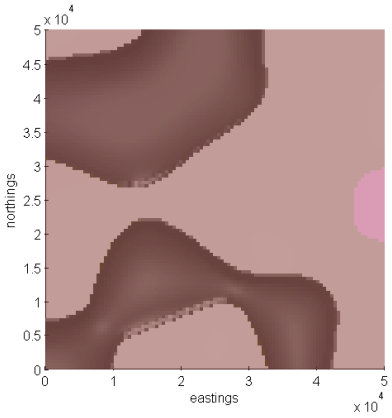
3. RESULTS TO DATE

This section presents an illustrative synthetic example of the capabilities of the inference engine. A 3D world model is defined based on prior knowledge of the region stratigraphy (Fig. 3A). Sensors can be defined on a 3D, 2D or 1D slice of this model.

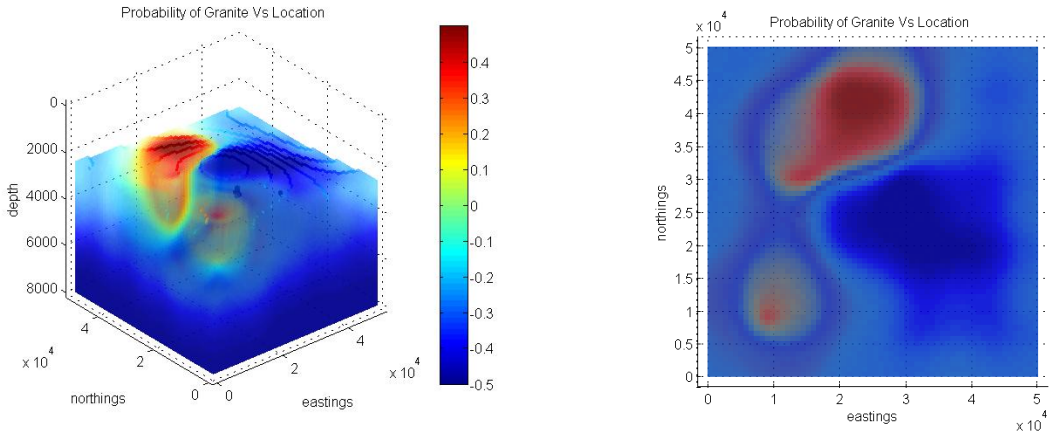
Information from several sensor modalities was generated to condition our method on including gravitational and magnetic field measurements, core samples and temperature readings from exploratory drill-holes as well as a 1-D seismic survey. The MCMC algorithm was then run to collect a large dataset of samples from the inversion posterior. As the size of the sample dataset becomes large, the distribution of its samples resemble the true Bayesian predictive distribution. Thus it is possible to both inspect individual samples from this posterior (Fig. 3B), which are each explanations of the data of varying probability density, or most importantly ask probabilistic questions of the posterior. The resulting set of samples can be used to answer several key questions typically encountered in geothermal exploration. In this case we ask "what is the probability of granite rock type occurring over space", and can answer this question by marginalising (integrating) over all the samples to obtain a "heat map" (Fig. 3C).



(a)



(b)



(c) (d)

Figure 3: a: Sample geological stratigraphy drawn from a prior distribution, p(c). The colors represent different geological units labelled by the classifier. b: A plan view of the same geology. The grey area indicates the outline of the granite targets. c & d: The result of marginalizing over the posterior distribution to produce the probability of granite versus location at an angled and plan view, respectively.

In this example, the parallel tempering algorithm used to estimate the posterior consisted of 4 chains with temperatures scaled logarithmically from T = 1 to T = 120. Each chain's state vector had 18 dimensions which controlled the shape of the geological units (parameters of c) and the estimated bulk geophysical properties, g. They were allowed to run for 400,000 iterations of which 20,000 were retained after a burn-in phase of 100,000 samples and a 25:1 sub-sampling regime to reduce dependencies between draws. Fig 3C and 3D show the probability of granite versus location after marginalizing out all other factors. Red and blue regions represent areas of high and low probabilities of granites, respectively. Transparency is used to reflect the level of uncertainty in our prediction at a given location. The probability of granites in the Cooper, Eromanga and Eyre Basins is not shown here for visualisation reasons.

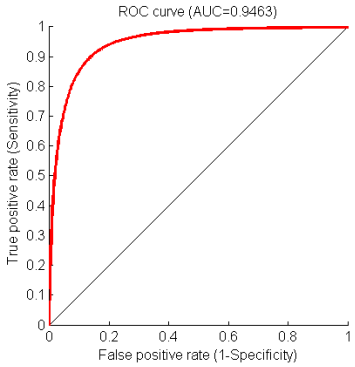


Figure 4: A receiver operator curve measuring the accuracy of the granite classifier. Area under the curve = 0.9463.

This bears a close resemblance to the ground truth and the classifying power of this prediction has been shown against the truth using an ROC curve in Fig 4 - the large fraction of the graph area below the curve (~95%) indicates this is a strong probabilistic predictor.

4. DISCUSSION

We have developed a Bayesian inference procedure to the exploration of geothermal energy resources. Our formulation enables inversions of multi-modal geophysical data by defining a joint distribution over rock material properties in conjunction with an explicit prior model for the sub-surface rock geometry. This approach has a high potential impact on the industry, as it will allow experts to make risk-based exploration decisions, naturally leading to machine learning strategies such as active sampling and planning to trade exploration against exploitation in selecting drilling locations.

Upcoming research will largely address the prior selection and tractability aspects of the problem. Appropriate selection of priors requires domain knowledge elicitation together with analysis of lab data acquired for the project, while the high dimensional model space is expected to motivate new variational or MCMC approaches.

Another line of research is the development of strategies to model fractures and faulting directly as this is central to geothermal exploration in particular tenements. Fractures can be observed from drill-hole data but are much more difficult to be perceived from remote geophysics surveys. A hypothesis we plan to investigate is the fusion of seismic and MT data to improve the prediction in areas where drilling information is too sparse.

Further, a very important outcome of the project will be a tool for the categorisation of geothermal targets, such that the industry can have easy access to the algorithms and data. We will investigate two alternatives for the tool: 1) software that can directly connect to the data portal, request data and perform inversions in a desktop machine, and 2) a more sophisticated web-based service. In this case, users can submit their data to the portal which will be fused with public data, to generate new inversion results. The interface with the system will be through a web-browser and the algorithms will be executed in on-demand cloud computing services. We are looking forward to the next steps of the project and appreciate comments and suggestions from the geothermal community into our developments.

REFERENCES

[1] Brooks S., Gelman A., Jones G., and Meng X.L.: Handbook of Markov Chain Monte Carlo. *Chapman & Hall/CRC* Handbooks of Modern Statistical Methods. Taylor & Francis, 2011.

[2] Pachauri R.K. and Reisinger A.: Climate change 2007: Synthesis report, 2007.

[3] Earl D.J. and Deem M.W.: Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys*., 7:3910–3916, 2005.

[4] Gilks W. R.: Markov Chain Monte Carlo in Practice. *Chapman and Hall/CRC*, 1999.

[5] Huenges E. and Ledru P. editors. Geothermal Energy Systems: Exploration, Development, and Utilization. *John Wiley & Sons*, 2010.

[6] Kearey P., Brooks M., and Hill I.: An Introduction to Geophysical Exploration. *Blackwell Science, 3rd edition*, 2002.

[7] Li Y. and Oldenburg D.W.: 3D inversion of magnetic data. *Geophysics*, 61(1):394– 408, 1996.

[8] Li Y. and Oldenburg D.W.:3D inversion of gravity data. *Geophysics*, 63(1):109– 119, 1998.

[9] Reid A., O'Callaghan S., Bonilla E.V., McCalman, L., Rawling T., and Ramos F.: Bayesian joint inversions for the exploration of earth resources. *In International Joint Conference on Artificial Intelligence*, pages 2877–2884, 2013.

[10] Grandis H., Menvielle M., and Roussignol M.: Bayesian inversion with Markov chains – I. The magnetotelluric one-dimensional case. *Geophysics. J. Int*. (1999) 138, 757-768.

[11] Wannamaker P. E., Stodt J. A. and Rijo L.: A stable finite element solution for two-dimensional magnetotelluric modelling. *Geophysical Journal of the Royal Astronomical Society*, 88: 277–296, 1987.

[12] Beardsmore G. R. and Cull J. P.: Crustal Heat Flow: A Guide to Measurement and Modelling. *Cambridge University Press*, 2001.