PROCEEDINGS, Thirty-Ninth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 24-26, 2014
SGP-TR-202

DISTRIBUTED BAYESIAN GEOPHYSICAL INVERSIONS

Lachlan McCalman[1*], Simon T. O'Callaghan[1*], Alistair Reid[1*], Darren Shen[1,2], Simon Carter[1,5], Lars Krieger[1,5], Graeme Beardsmore[1,6], Edwin V. Bonilla[1,3], Fabio T. Ramos[1,2]

[*]Equal Contribution

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

lachlan.mccalman@nicta.com.au

Keywords: exploration, Bayesian inference, inversion, software, drilling risk reduction, magnetotellurics

ABSTRACT

The quantification of uncertainty in exploration for geothermal targets is central to determining the best locations for drilling. We develop algorithms to perform data fusion and joint inversions from a broad range of independent geophysics surveys to infer rock properties such as density, magnetic susceptibility, seismic velocity, resistivity, thermal conductivity and temperature. Our methods seamlessly integrate surveys including geology, gravity, magnetics, seismic reflection, magnetotellurics and borehole geophysics to predict the most relevant properties for predicting geothermal drilling success. As part of a probabilistic inference procedure, we quantify the uncertainty for each quantity of interest, providing principled strategies for risk minimisation in drilling projects. While developed initially for geothermal exploration in Australia, the methodology is readily adaptable for other geological settings and geothermal drilling targets.

In this work we develop an inference procedure based on parallel tempering Monte Carlo techniques that is well suited to large-scale computation infrastructures such as cloud computing. This enables our probabilistic approach to the inversion problem not only to quantify the risk associated with drilling, but also guide strategies for further pre-drilling exploration, increasing drill target confidence. The Bayesian formulation allows expert knowledge to be added in the form of prior distributions, making explicit the contribution of human judgment to the final result.

We demonstrate our approach in three experiments: 1) a joint magnetics and gravity inversion to infer locations of buried granite outcrops, 2) a joint magnetotelluric and gravity inversion to illustrate the benefits of data fusion and; 3) a performance test comparing our cluster-based parallel inference algorithm to traditional methods.

1. INTRODUCTION

Nonconventional geothermal systems have the potential to unlock vast renewable energy sources by exploiting hot aquifers or granites buried under kilometres of rock. Unfortunately, direct drillhole observations at these depths are both rare and expensive, Huenges (2010). This makes exploring for new reservoirs and characterising existing reservoirs reliant on indirect geophysical measurements. However, such measurements only weakly inform the properties and structure of the underlying rock. One of the central challenges of unconventional geothermal is using this weak information to construct a model with enough certainty to drill.

Determining subsurface properties and structure from a set of geophysical measurements is an example of an inversion problem. Inversion problems in geophysics are typically high dimensional, in that there are an infinite number of plausible geological/geophysical models to consider, and under constrained, in that many different models may produce the same measurements on the surface, Li (1998).

To further increase the complexity of the problem, a single type of sensor may be insufficient to represent the entire dimensionality of the problem. For example, gravity sensors alone provide poor depth resolution but can be very useful when fused with complementary sources of information such as seismic surveys. A number of previous approaches have been proposed to take advantage of this synergy. For example, Fullagar (2007) used alternating optimizations of single sensor inversions on a joint model to find satisfactory geological structures that satisfied multiple sets of observations. Recent work in Shamsipour (2012) carried out stochastic inversions of gravity and magnetic data jointly using cokriging.

A critical property of a geophysical inversion is that a single 'best guess' or *most likely* answer will invariably fail to capture the uncertainty associated with result, Kearey (2002). Many other, possibly quite distinct structures could explain the measurements, and these other possible structures (and their relative likelihood) could be critical to making decisions. Alternatively, Bayesian inversion approaches not only provide principled means to quantify uncertainty but also the integration of relevant, multi-modal information to constrain the solutions. For example, direct samples of rock properties such as density and resistivity enable the definition of a joint prior, in which measurements of one modality (such as gravity) provide information about both density and resistivity.

We present a Bayesian approach to geophysical inversion that directly assess uncertainty and allows observations from different types of sensors to be fused in a principled manner. Our method explicitly computes probabilities over different possible geological structures, and so enables the calculation of risk to guide geothermal exploration, sensing and drilling decisions. This, however, requires significant

computation resources and scalable algorithms that can be deployed in modern parallel hardware. We develop a Bayesian inference engine explicitly designed to enable forward model simulations to be computed on massively parallel hardware such as cloud computing. The algorithm can utilise hundreds of processors simultaneously, and achieve near linear speed ups.

This paper is organised as follows. In Section 2 we explain the methodology of our approach followed by a description of its implementation in Section 3. A set of experiments are discussed in Section 4 which illustrate some of the framework's key characteristics and the main conclusions are presented in Section 5.

2. FORMULATION

A novel Bayesian inference engine has been developed to probabilistically invert multi-sensor geophysical observations. This section outlines the key assumptions of this inference engine, while implementation and computational details are deferred to the following section.

Let $D = D_k\, k = 1 \ldots K$ be the set of observed quantities from *K* distinct sensors that respond to *J* spatially distributed rock properties $g(x) = [g_1(x), \ldots, g_J(x)]$, functions of spatial position *x*. The *forward models*, $f_k(g)$ predict the ideal sensor measurements such that D_k are noisy observations of the underlying physical process modeled by $f_k(g)$. For example, one may observe variations in the gravitational field D_k, which are related to the density of the underlying rocks $g_j(x)$ via a forward model as given by Newton's law of gravitation. Our goal is to reason about these rock properties $g(x)$ given the set observations *D*. For a given data *D*, and a space of models g, our inference engine is defined as a computational application of Bayes' rule:

$$P(g|D) = \frac{P(D|g)P(g)}{\int P(D|g)P(g)\, dg}$$

This rule states that the predictive posterior distribution over possible models can be computed from an interpretable prior belief of which models are expected to occur, and a likelihood function that objectively quantifies the compatibility of the model forward simulations with the observed data. The key advantages of a Bayesian approach are that the system can natively use all the available sensor modalities, model assumptions and biases are made explicit in the formulation, and the results are presented as a posterior distribution over models allowing the end users to reason about risk.

Our design has decomposed the prior and likelihood terms of Bayes' rule above into four modeling components: (i) a prior belief of the structures present in the region under study; (ii) a joint prior over the properties of the rocks found in these structures; (iii) geophysical forward simulations; and (iv) probabilistic likelihood models. Figure 2 shows a graphical model of our approach. These components are introduced below.

2.1 World Model: Geological Structure $\mathbf{P(\alpha)P(c)}$

Given our current knowledge of existing geothermal sites, we have decomposed our geological prior into a prior over the structure of the underlying 3D volume and a prior over the possible *rock types* in that structure. In particular, we have considered a layer-cake or sedimentary basin model where the prior over the structure defines different layers, which are determined by control points $\alpha_i \in \mathbb{R}^3$. We denote this structural prior with $P(\alpha)$.

This representation of the world allows us encode 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, this prior defines 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 and denotes it with $P(c_l)$, with $l = 1,\ldots,L$.

An example of such a world is shown in Figure 1 below. This is generated using a random draw from our prior probability distributions over α and *c*. The model has been configured to (in this particular run) have 3 sedimentary layers, and a basement with granite intrusions (grey). The age sequence of the regions is specified so that younger layers may overlay (and `pinch out') older layers that would occupy the same space.

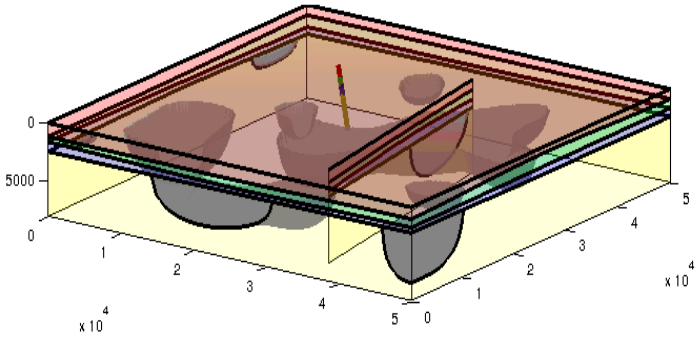

Figure 1: Geometry of an example parameterised world model. The top three layer interfaces are controlled by only 9 parameters each, while 64 parameters are dedicated to the granite intrusions. A drill-hole and a 2D slice section are also depicted within this volume.

2.2 World Model: Joint Prior over rock properties P(**g**)

While we are ultimately interested in predicting reservoir depth, volume, temperature and transmissivity, the ability to infer these parameters from a diverse range of geophysical sensors comes from modeling the joint distribution of a larger set of physical properties. Put simply, if our inference engine is given prior knowledge about how the properties of a given rock type such as density, magnetic susceptibility, seismic velocity, conductivity and transmissivity vary jointly, then it can fuse observations of sensors that depend on any of these properties in a hierarchical fashion, and could even infer partial information about properties that haven't been directly sensed.

For each layer we define a non-spatial joint prior over rock properties $P(\rho_l) = (\rho_1 \dots \rho_L)$. For a given geology as determined by α and c, this defines a joint prior over the corresponding spatial properties g(x). The prior is drawn from a library of R rock types. This allows us to maintain a (non-spatial) low-dimensional distribution $P(\rho_l)$ that, together with the geology, implicitly defines a distribution over the rock properties over the entire space.

The key bodies in the Cooper Basin in South Australia have been coarsely classified into distinguishable features (granites, basins, sandstones). We have begun building the library for these classes, based on new lab test analyses of existing drill-cores in the cooper basin. Joint data in the literature is rare - usually only one or two properties of interest are sampled. In our work we require simultaneous measurements of all properties of interest, so the new core testing data are both critically important to our project and novel.

Modeling down to 5 km deep we cannot expect to predict small scale property variations in the rocks that the very small core samples draw. Short of direct drilling, the relatively low cost sensors we are fusing cannot resolve the small scale variations in rock bodies such as individual fractures. However, it is reasonable to model bulk properties, such as mean density or fracture level over a spatial region.

This data for each class has been analysed using Gaussian Process regression, a nonparametric probabilistic machine learning tool that allows the estimation of both the degree and range of spatial variation in the data, and the first order dependencies between the properties. A different bulk property distribution is learnt for each rock class, and these serve as the material property prior. The collection of distributions serves as a library of rock priors - the world model above utilises this library, as it selects discrete class assignments to the modeled rock regions.

2.3 Geophysical Forward Models

The inference engine employs an explicit understanding of the relationship between the geophysical observations and the underlying geology. This appears in Bayes rule as part of the likelihood term *P(D|g)* - the probability of the data given the model. If all the uncertainty in the modeling and sensing can be propagated through the code probabilistically, then this allows the forward model to generate a likelihood directly. However, we must also consider that the assumptions of the model may prevent the closest representation from being an exact data fit. Because in reality the noise, intrinsic and extrinsic errors in the forward models are difficult to understand and predict, we have split the likelihood into two components - the physics based simulation of ideal forward models, and the likelihood distributions to model their uncertainty. The forward models f_k for the k^{th} sensor type we are considering are summarised in Table 1 below, and map a geology g into an ideal observation $D_k = f_k(g)$*:*

Table 1: Summary of the key sensor forward models in the inference engine.

Sensor	Forward Model $f_k(g)$

Gravity	Newton's gravitational law, using Li's tractable approximation for a 3-D field of constant density prisms, Li (1998).
Total magnetic anomaly	Induced magnetic field using Li's tractable approximation for a 3-D field of prisms of constant susceptibility (neglecting residual permanent magnetism), magnetic field referenced from IGRF tables, Li (1996).
Reflection seismology	Return travel times of major reflectors from the signal extracted from migrated reflection seismic lines (available recorded lines are not raw: multiple shots have been stacked additively to increase the signal-to-noise ratio, preventing raw signal analysis).
Magnetotellurics (MT)	Frequency domain analysis of the electrical conductivity of the Earth's crust, using the 2-D finite element forward model proposed by Wannamaker et al. (1987). MT is expected to respond to fluid and is therefore a key sensor for fracture modeling.
Temperature	Local temperature observations can be used to infer the thermal conductivity structure of the subsurface. The production and redistribution of heat is modeled in 3-D with the second-order linear differential equation: $\nabla(\lambda(x,y,z)\nabla T) = -A(x,y,z)$, where ∇ is the gradient operator; T is the temperature; $\lambda(x,y,z)$ is the thermal conductivity; and A is the heat production rate. We assume.thermal steady state, ignore heat transfer due to advection, and solve this equation using finite differences with the following boundary conditions: (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.
Drillholes	Sparse direct observations of the rock properties and interfaces can be incorporated as a strong localised sensor into our forward models.

2.4 Likelihood Model

The above sensor models are deterministic simulations that compute ideal sensor measurements for a given sensor and world model. The Bayesian formulation instead requires a probabilistic likelihood *P(g|D)*, which extends the forward model to the space of possible measurements. This brings in observation-model fit uncertainties, which are tied to instrument errors, noise levels, data density, sensor resolution and even the assumptions of the model.

As described above, each sensor dataset D_k is constrained by its corresponding forward model $f_k(g)$. We assume that all sensor measurements are conditionally independent given the world model outputg. In other words: $P(D|g) = \prod_{k=1}^{K} P(D_k|\, f_k(g))$. Note that this is not a strong assumption to carry out *joint inversions* because dependencies between observations are still understood via the hierarchical structure of the model and additional statistical dependencies are exploited via the prior.

We assume that the likelihood of observations obtained by each sensor is a Gaussian centred on the forward model:

$$P(D_k|g,{\sigma_k}^2) = Normal(D_k;\, f_k(g),{\sigma_k}^2 I),$$

where ${\sigma_k}^2$is the variance corresponding to the observations of sensor modality D_k. However, we don't have sufficient prior knowledge to determine a priori a sensible valuable for each sensor variance ${\sigma_k}^2$, and we instead place an additional hierarchical prior on the variances of the likelihoods. In particular, we use an *Inverse-Gamma prior*:

$$P({\sigma_k}^2|a,b) = Inverse-Gamma({\sigma_k}^2;\, a_k,b_k),\ a_k > 0 \text{ and } b_k > 0,$$

where $a_k$and $b_k$are the shape parameter and the scale parameter respectively. This states that we don't know what the true noises are, but we have a prior belief about the possible values they might take. Since the Inverse-Gamma prior is *conjugate* to the Normal distribution (given a fixed mean), and we have assumed independence on the conditional likelihood models across the different sensor modalities, we can integrate out the unknown noise variances analytically and obtain a (marginal) likelihood model of the form:

$$P(D_k|g,a_k,b_k) = \int P(D_k|g,{\sigma_k}^2)P({\sigma_k}^2|a_k,b_k)d{\sigma_k}^2 = t_{2a_k}(D_k; f_k(g),\frac{b_k}{a_k}I),$$

where $t_\nu(D_k;m,S)$ denotes a multivariate t-distribution over $D_k$with ν degrees of freedom, location parameter m and scale matrix S.

2.5 Joint Probability Distribution

Putting together the prior model and the likelihood model, we can expand Bayes' rule into the full joint distribution:

$$P(\alpha, c, \rho, D|\mu_\alpha, \Sigma_\alpha, \{\mu_t, \Sigma_r\}, \{\theta_l\}, \{a_k, b_k\}) = P(\alpha|\mu_\alpha, \Sigma_\alpha)P(c|\{\theta_l\}) \prod_{l=1}^{L} P(\rho_l|\{\mu_t, \Sigma_r\}) \prod_{k=1}^{K} P(D_k|g, a_k, b_k)$$

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

$$P(\alpha|\, \mu_\alpha, \Sigma_\alpha) = Normal(\alpha; \mu_\alpha, \Sigma_\alpha) \qquad P(c_l|\theta_l) = Multinomial(c_l; \theta_l)$$

$$P(\rho_{l=r}|\mu_t, \Sigma_r) = Normal(\rho_{l=r};\, \mu_r, \Sigma_r) \qquad P(D_k|g, a_k, b_k) = t_{2a_k}(D_k; f_k(g(\alpha, c, \rho)), \frac{b_k}{a_k} I)$$

Next we determine the posterior distribution over geological structures and properties given measurements from each of the *K* different modalities jointly: $P(\alpha, c, \rho|\, D)$where $D = \{D_1, \ldots D_k\}$ and, for simplicity in the notation, we have omitted the conditioning on the known hyper-parameters. 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 sensor 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), Li and Oldenburg (1996, 1998). This model contains no notion of rock types 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.

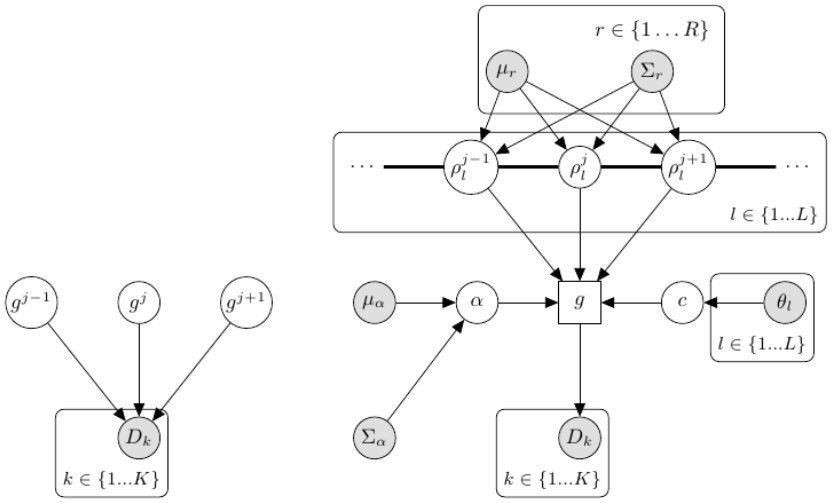

Figure 2: Probabilistic graphical model representations of traditional survey techniques (left) and the fusion procedure implemented in our inference engine (right). The traditional approach contains no notion of geological units with different prior relationships between geophysical properties, so correlations 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.

3. IMPLEMENTATION

An analytical solution for the posterior of interest has been previously identified by leveraging a number of simplifying assumptions regarding the space of structural geologies considered and the suite of geophysical sensors employed (spatially continuous rock properties, linear forward models), Reid (2013). The approach provides analytical tractability advantages, but its assumptions preclude the possibility of both complex geological features and the fusion of several highly informative, nonlinear datasets such as a seismography and magnetotellurics that require numerical simulation of partial differential equations to compute the forward models $f_k(g)$. Crucially, the original approach scales poorly to parallel architectures. Instead, our implementation of the inference engine has transitioned to the more flexible inference method of parallel Markov chain Monte Carlo (MCMC) sampling, Geyer (1991).

3.1 Parallel MCMC

Nonlinear geophysical inversion problems typically involve probabilistic sampling from multi-modal 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. Creating proposal distributions that enable Markov chains to efficiently traverse these distributions with isolated modes is non-

trivial. 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), Brooks (2011).

We employ a meta-algorithm known as parallel tempering (PT) to aid with convergence by utilising multiple coupled MH chains. 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 $p_{\text{accept}} = \min(1, \exp[(\varphi(\theta_c) - \varphi(\theta_p))/\Psi])$ 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: $p_{\text{swap}} = \min(1, \exp[(\varphi(\theta_j)/\Psi_i - \varphi(\theta_i)/\Psi_j]/\exp[(\varphi(\theta_i)/\Psi_i - \varphi(\theta_j)/\Psi_j])$ which satisfies the detailed balance constraint of MCMC.

Convergence is monitored using the potential scale reduction factor. Multiple chains at identical energy levels are run out until the ratio of key inter and intra-chain statistical measurements approaches unity. A different reduction factor is required for each dimension of the state space. Once the chains are deemed to have converged, the samples can be combined to create a long single chain from the stationary distribution.

Only samples produced by the Markov chains with $\psi = 1$ can be considered as draws from the posterior. However, rejection and importance sampling can be employed to recover true posterior samples from the higher energy chains by using $\exp(-\phi(\theta)\psi_i)$ as the proposal distribution and $\exp(-\phi(\theta))$ as the target.

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. We are developing software to exploit this scalability, outlined in the next section.

3.2 GDF Distributed Inversion Software

MCMC Inference enables the use of complex forward models and arbitrary priors and likelihoods. The cost of this flexibility is the need to evaluate a large number of sample points. Geophysical forward models can be particularly expensive to sample, requiring either voxelized integration or numerical simulation of partial differential equations. To keep run-times feasible it is therefore critical that an MCMC-based inversion algorithm be as computationally efficient as possible.

To address the computational challenges of our approach, we are developing a distributed computing platform called Geothermal Data Fusion (GDF) that exploits the natural parallelism of multi-chain and parallel tempering MCMC to run sample evaluations simultaneously on a large number of computers. GDF implements a remote procedure call interface to evaluate forward models concurrently on a set of connected ‘worker’ machines.

Architecture

GDF is designed as a client/server distributed architecture with TCP communication over IP networks. The client and server are designed primarily to be deployed with virtual machines targeting Amazon EC2 instances or HPC clusters. GDF implements a distributed likelihood pool, in which the server selects parameters to evaluate, and worker machines actually compute and return the associated value of the unnormalised posterior.

Forward models

GDF currently has implementations of 3D gravity, magnetics and heat flow, and 2D magnetotellurics and (migrated) seismic. The software also supports input of core sample data and contact points to inform the rock property and structural priors. These models have been implemented in C++, with both CPU and GPU implementations; multi-core parallelism is also supported: A single worker may simultaneously process jobs on multiple CPU cores and GPUs.

4. EXPERIMENTS

4.1 Gravity Inversion Scenario

The focus area for our preliminary testing and evaluation is the Cooper Basin in South Australia. The region's geological stratigraphy is well understood, consisting of a basin-shaped partially metamorphosized basement overlain with multiple sedimentary basins. The basement is intruded by radiogenic granites that have risen from the Earth's upper mantle and act as a heat source, which is insulated to varying degrees by the 3-5 kilometers of crust above. This knowledge of the stratigraphy has been encoded into the synthetic world model prior as shown in Figure 3 (a) where a random draw from the prior distribution has been visualised and will serve as our ‘truth’ for an experiment.

The goal of this scenario is to demonstrate the MCMC-based inference engine’s ability to predict properties of interest such as the location of the granite intrusions in a probabilistic fashion, and how the results can be presented. The parallel tempering algorithm has been configured to consist of 4 chains with energies scaled logarithmically from $\psi = 1$ to $\psi = 120$. The world model’s state vector has 25 dimensions of freedom, so each chain stores a separate state vector θ with 25 entries that control the shape of the geological units (parameters of c) and the estimated bulk geophysical properties, g. They were allowed to run for 2,000,000 iterations of which 150,000 were retained after a burn-in phase of 500,000 samples and a 10:1 sub-sampling regime to reduce dependencies between draws.

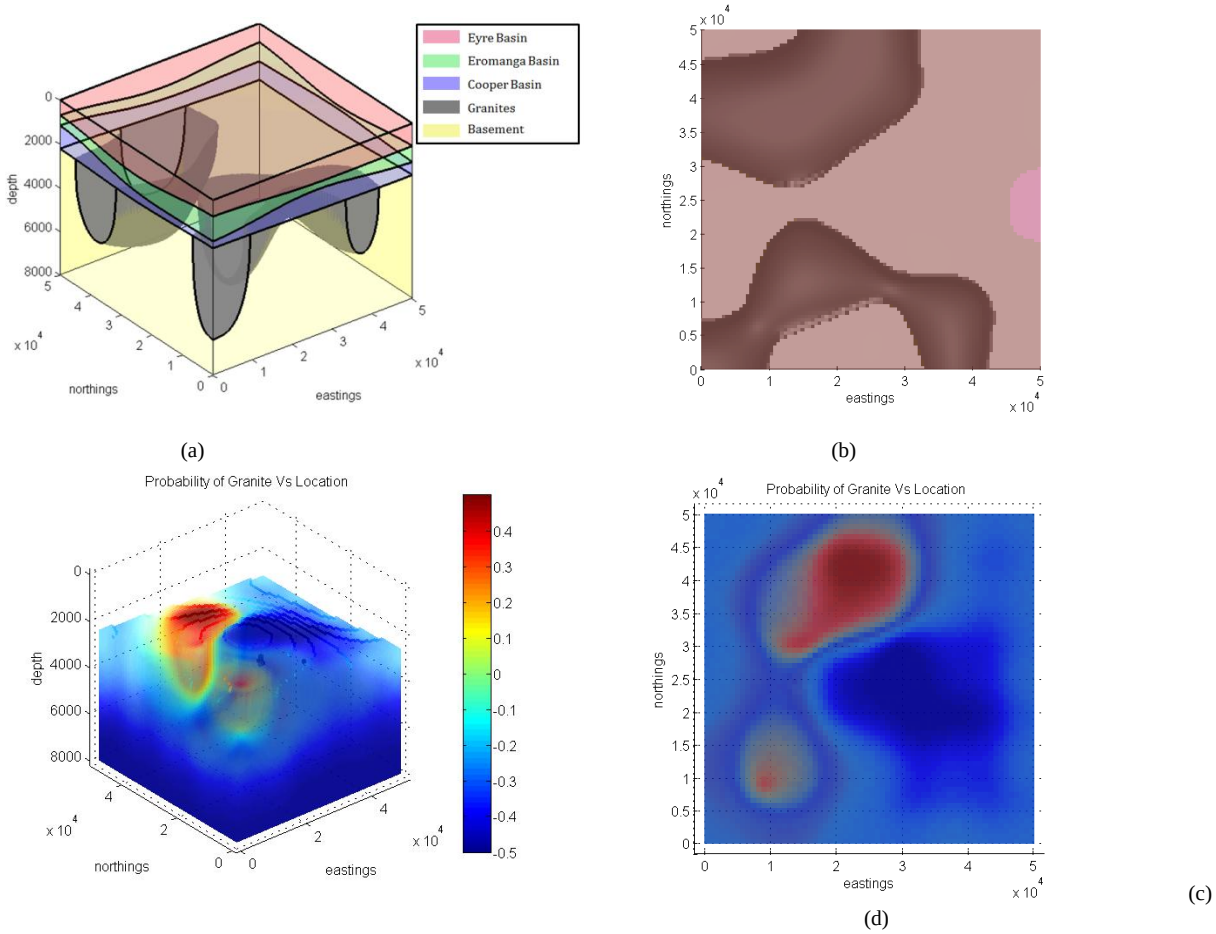

Figure 3: a: Sample geological stratigraphy drawn from a prior distribution, p(c). The colors represent different geological units labelled by the classifier. b: A plan view of the same geology. The grey area indicates the outline of the granite targets. c & d: The result of marginalizing over the posterior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

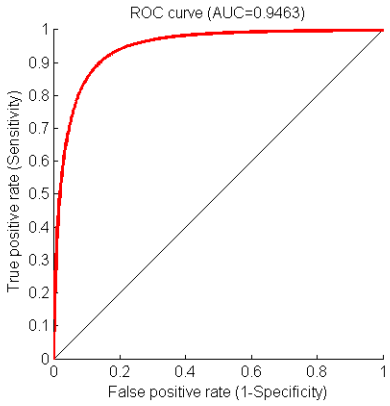

Figure 4: A receiver operator curve measuring the accuracy of the granite classifier. Area under the curve = 0.9463.

The resulting set of samples can be used to answer several key questions typically encountered in geothermal exploration because they represent a distribution over explanations of the data. We may pose questions over these distributions, such as `where has the model placed the granite intrusions', and extract a scalar probability field out of the distribution through marginalisation. Figure 3 (c) and (d) visualise the probability field 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 top 3000 metres is fully transparent to clarify the visualisation. The location and shape of the

granites bear a close resemblance to the ground truth - if we evaluate the output as a probabilistic classifier against the ground truth, a receiver operating characteristic analysis returns a high area under curve score of 0.946 (Figure 4).

4.2 Magnetotelluric-Gravity Joint Inversion

Magnetotellurics (MT) is a natural-source electromagnetic method, used to map local and regional resisitivity structures of the subsurface at various depths scales of a few metres to several hundred kilometres. Such electrical properties are dependent on the amount of fluids present and the connection of fluids, the temperature and chemical composition of the rocks, and whether the rock is molten.

In contrast to potential field methods, MT can resolve directional information about conductivity. This information can be used to resolve the fractured rock pathways which are a target of geothermal projects. MT has been used in classical geothermal exploration of over 20 years and has proven to be promising for Enhanced Geothermal Systems (EGS) as well. While the sensor modality can provide a valuable insight into the subterranean structure it is susceptible to considerable signal strength attenuation in the presence of conductive layers near the surface.

The skin depth equation provides an estimate for the relation between the penetration depth of the electromagnetic MT signal and the product of the signal frequency and the medium's resistivity. It shows that the ability to resolve structures in a given depth depends on the resistivity of overlying layers. If a conductive layer is present close to the surface, a significant portion of high frequency signal content is attenuated within this layer. Therefore, effectively only low-frequency parts of the original electromagnetic signal can sufficiently penetrate the medium down to the target depth. Naturally the resulting spatial resolution of the information obtained from these parts of the signal is decreased.

In the synthetic test case presented below, we model such a scenario: we set up a layered Earth with varying resistivities. A sedimentary basin with resistivity increasing with depth lies on top of an igneous basement at 10 km. A fluid-permeated layer is represented by a highly conductive region at 2.5 km (Figure 5 (a)). We calculate the respective MT signals and then try to image the underlying structure by inverting these signals again with particular emphasis on resolving the depth to basement. First, we perform an inversion using the MT data solely. Secondly, a joint inversion, including gravity information, is applied.

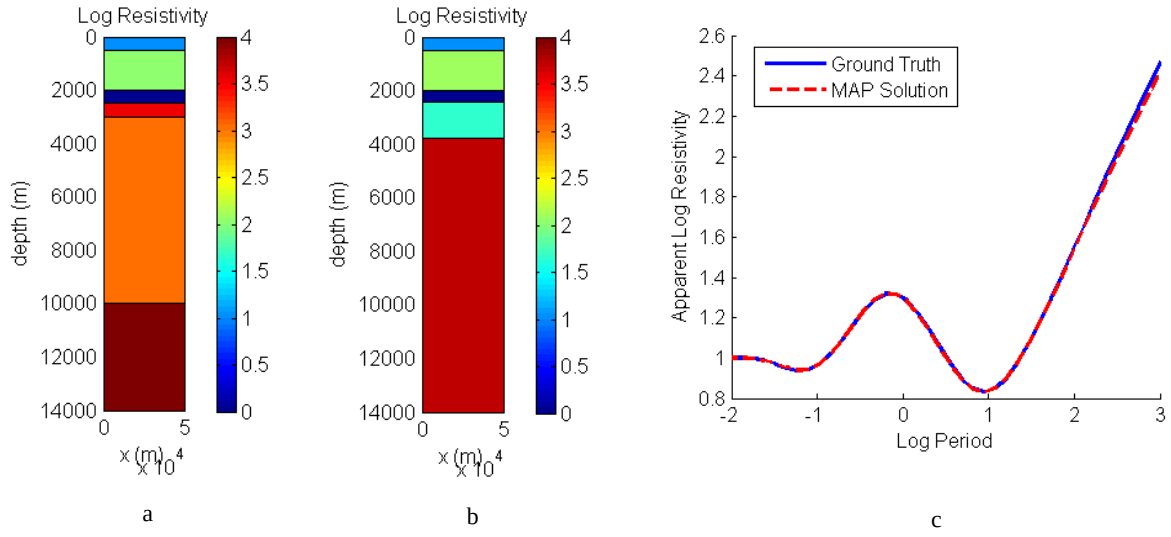

Figure 5: a: The ground truth stratigraphy. b: Predicted stratigraphy after minimising the forward MT model error relative to the observed data. c: Measured apparent resistivity versus predicted apparent resistivity.

The traditional approach used in industry is an optimisation of the model parameters to reduce error in the model's predicted forward simulation versus the observed data. The results of such an approach are shown in Figure 5 (b) and (c). While this method performs well in regions above the permeated region, there is a large discrepancy between the ground truth and prediction below 2.5 km due to the shielding effects of the conductive layer. This is despite the strong correlation between the measured data and the prediction's forward model. Crucially, this method's output lacks an estimate of uncertainty which hinders a robust risk analysis strategy.

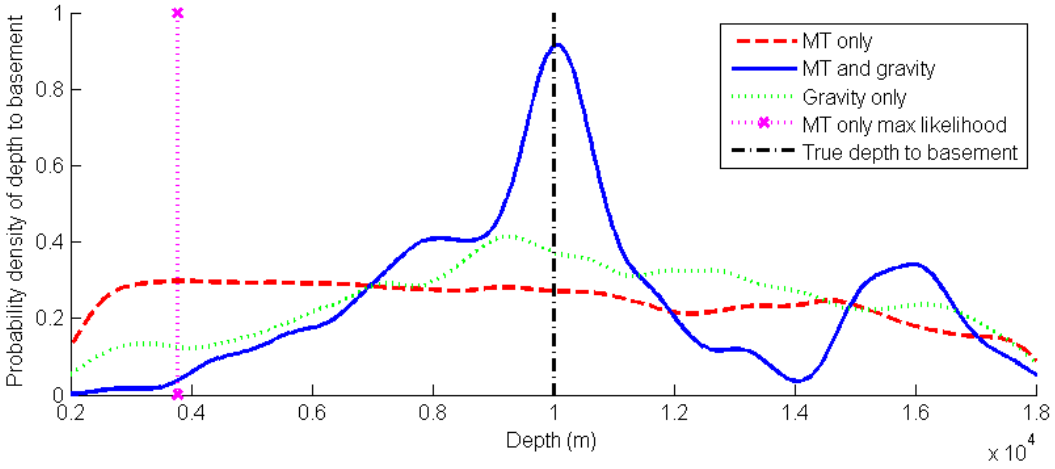

Figure 6: Probability density functions over the depth to basement. The true depth is 10 km with a bulk resistivity of 10000 Ωm. A traditional optimisation approach estimates the basin's depth to be 3.77 km and a basement resistivity of 4981 Ωm (vertical pink dashed line).

Our proposed approach provides a principled measure of the model's uncertainty. The red curve in Figure 6 represents the marginal probability distribution of the depth to basement after condition the parameters on the magnetotellurics data. The influence of the conductive layer's shielding effect can be seen in the distribution's large variance. Fusing additional sensor modalities is relatively straightforward within the proposed Bayesian framework. Using the previous distribution as prior, the blue curve can be recovered after conditioning on a single gravity observation and a subsequent marginalization over the model's other parameters. The combination of both sensor types, which independently provide poor depth resolution for different reasons, yields a more precise estimate of the depth to basement.

4.3 Computational Scaling of Gravity Inversion

In this experiment we tested our parallelisation strategies on a simulated gravity inversion using PT MCMC. To perform the inference the software must evaluate sample likelihoods requested by one or more MCMC chains. The likelihood evaluation requires gravity observations to be computed on a candidate geological structure. The bulk of the computational cost of the problem arises from these forward model evaluations. A parallel implementation of this problem running multiple chains ought to achieve close to linear gains with the number of cores utilized.

The experiment compared three computational approaches to performing the inversion. The first was a traditional, fully-serial implementation on a single CPU. The second was a small cluster running a pre-alpha version of the GDF software. The third was also a GDF cluster, but with multi-threaded workers to leverage multiple physical CPU cores on a single machine. The methods compared were the all running the same gravity forward model implemented in C++ with the Eigen3 linear algebra library, Guennebaud (2010).

For testing purposes, the inversion was stopped after 1000 forward model evaluations. Note that these were not enough samples to achieve convergence, but were sufficient for testing the relative speeds of the computational approaches. For the single-CPU code, this was made up of a single chain of length 1000, whilst the cluster approaches both used 25 parallel chains with 40 samples each. The problem difficulty was adjusted by increasing the resolution of the voxelized world model, from 16x16x16 to 160x160x160 in 10 steps. The cluster contained 4 desktop machines on a local network. The single-CPU was testing on the fastest machine in that cluster, which was an Intel Xeon E7-8830 clocked at 2.13Ghz. The results of the experiment are displayed in Figure 7.

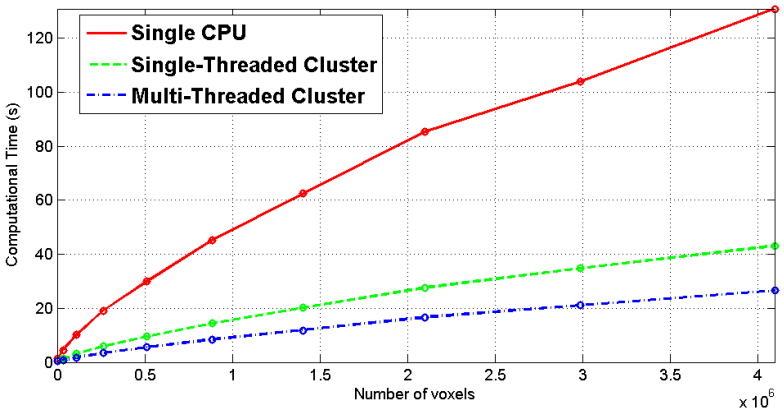

Figure 7: Computation time for 1000 Gravity forward model simulations of a 3D grid, as a function of the number of voxels in the grid and the computational strategy employed

The single-threaded cluster achieved an approximate factor of 3 speed up over the single CPU implementation. As the serial job was run on the fastest machine, this is approximately the factor expected by the additional FLOPS available to the cluster, though clearly there is some computational overhead with the communications system and the use of multiple chains. The multi-threaded cluster achieves another factor of two gain, which corresponds to the 2 threads available per machine. As expected, the multi-core parallelisation does carry a significant overhead.

Note that the apparent sublinear performance of all methods is an artefact of the caching strategy used in the forward model. The sensitivity matrix is pre-computed and saved before any jobs are run. This calculation dominates the runtime for the smaller problems but becomes a progressively smaller fraction of the total computational time.

5. CONCLUSIONS AND FUTURE WORK

We have applied Bayesian inference 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. We also have attacked the computational challenges of this approach by implementing a distributed computation system.

Our 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. Further work on the GDF software system will verify scalability on larger clusters, and provide a scriptable user interface and visualisation capability.

ACKNOWLEDGEMENTS

This research is supported by the Australian Renewable Energy Agency (ARENA) through the measure ``Data Fusion and Machine Learning for Geothermal Target Exploration and Characterization". NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

REFERENCES

1. Beardsmore G. R. and Cull J. P.: Crustal Heat Flow: A Guide to Measurement and Modelling. *Cambridge University Press*, 2001.
2. 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.
3. Pachauri R.K. and Reisinger A.: Climate change 2007: Synthesis report, 2007.
4. Earl D.J. and Deem M.W.: Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys*., 7:3910–3916, 2005.
5. Fullagar, P. K., and G. A. Pears. "Towards geologically realistic inversion."*Proceeding of Exploration*. Vol. 7. 2007.
6. Geyer, C.J. "Markov chain Monte Carlo maximum likelihood". In Keramidas, E.M. *Computing Science and Statistics: Proceedings of the 23rd Symposium of the Interface*. Fairfax Station VA: Interface Foundation. (1991). pp. 156–163.
7. Gilks W. R.: Markov Chain Monte Carlo in Practice. *Chapman and Hall/CRC*, 1999.
8. 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.
9. Guennebaud G. and Jacob B. and others: Eigen v3, http://eigen.tuxfamily.org, 2010

10. Huenges E. and Ledru P. editors. Geothermal Energy Systems: Exploration, Development, and Utilization. *John Wiley & Sons*, 2010.
11. Kearey P., Brooks M., and Hill I.: An Introduction to Geophysical Exploration. *Blackwell Science, 3rd edition*, 2002.
12. Li Y. and Oldenburg D.W.: 3D inversion of magnetic data. *Geophysics*, 61(1):394– 408, 1996.
13. Li Y. and Oldenburg D.W.:3D inversion of gravity data. *Geophysics*, 63(1):109– 119, 1998.
14. 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.
15. Shamsipour, P., Marcotte D., andl Chouteau M., "3D stochastic joint inversion of gravity and magnetic data." Journal of Applied Geophysics 79 (2012): 27-37.
16. 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.