Multi-Objective Bayesian Optimisation and Joint Inversion for Active Sensor Fusion

Sebastian Haan¹, Fabio Ramos ² and R. Dietmar Müller ³

ABSTRACT
A critical decision process in data acquisition for mineral and energy resource exploration is how to efficiently combine a variety of sensor types and to minimize total cost. We propose a probabilistic framework for multi-objective optimisation and inverse problems given an expensive cost function for allocating new measurements. This new method is devised to jointly solve multi-linear forward models of 2D-sensor data and 3D-geophysical properties using sparse Gaussian Process kernels while taking into account the cross-variances of different parameters. Multiple optimisation strategies are tested and evaluated on a set of synthetic and real geophysical data. We demonstrate the advantages on a specific example of a joint inverse problem, recommending where to place new drill-core measurements given 2D gravity and magnetic sensor data, the same approach can be applied to a variety of remote sensing problems with linear forward models - ranging from constraints limiting surface access for data acquisition to adaptive multi-sensor positioning.

INTRODUCTION
One of the most important aspects when dealing with incomplete information is where to collect new observations. This may include drill-site placement or the layout of a particular geophysical survey, in particular if measurements are very costly or resources are limited. Bayesian Optimisation (BO) solves this decision-making problem by finding global optimal solutions using a probabilistic framework given multiple measurements, prior knowledge, and model uncertainties. This offers solutions for wide range of sensor fusion problems and allocation problems such as: What is the optimal mixture and placement of different sensors? Where to sample if the cost function is incomplete or if there are large uncertainties in future total budget allocations? How to position sensor grids over time if the model state is dynamic or has a moving target?

Typically the function(s) involved in BO are unknown or very expensive to evaluate, and are therefore surrogated by prior models, such as Gaussian process (GP) models for

¹The University of Sydney, Sydney Informatics Hub, Sydney, NSW, 2008, Australia. E-mail: sebastian.haan@sydney.edu.au (corresponding author)
²The University of Sydney, School of Computer Science, Sydney, NSW, 2006, Australia. Email: fabio.ramos@sydney.edu.au
³The University of Sydney, School of Geosciences, Sydney, NSW, 2006, Australia. Email: dietmar.muller@sydney.edu.au
tractability (see Rasmussen and Williams, 2006, for an introduction to GPs). Once new data are available to evaluate the objective function, the prior is updated to form the posterior distribution over the function space.

While BO research began by focusing on single-objective bound-constrained optimisation (see for earliest work Kushner, 1964; Mockus, 1975), e.g., using the Expected Improvement (EI) criterion (Mockus, 1975; Jones et al., 1998), it has rapidly developed to address multi-objective problems (see, e.g., Emmerich, 2005; Jeong and Obayashi, 2005). Many criteria have been proposed for finding the optimal placements of sensors, where the goal is to maximize the cumulative reward for placing sensors by optimally balancing exploration and exploitation with as few evaluations: from the multi-armed bandit paradigm (Robbins, 1952; Srinivas et al., 2009) to maximizing the information gain and the mutual information between the chosen and unselected locations (Caselton and Zidek, 1984; Krause et al., 2008), and other value of information criteria (see, e.g., Eidsvik et al., 2015; Mukerji et al., 2001; Osborne et al., 2008). Today multi-Objective BO is recognised as a robust statistical method in handling multiple constraints and to find the optimal solution for competing objectives and multi-task problems.

The set of objectives is defined in an acquisition function. While there exist many approaches in the literature to define and optimize acquisition functions (for an overview see, e.g., Brochu et al., 2010), the main challenge still lies in constructing a prior model 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, such as a 3D model of a portion of the continental crust; this is also known as an inverse problem.

In geology and geophysics, inverse problems occur whenever the goal is to reconstruct the geological conditions, i.e. the distribution of physical rock properties, that give rise to a set of geophysical observations. Since the number of possible geological configurations is typically greater than the number of observational constraints, the problem is nearly always under-determined. A typical example is gravity anomalies remotely measured at the Earth’s surface, which reflect the 3D density distribution within the Earth. A given gravity anomaly at the surface may be caused by a given density anomaly near the surface or by a much larger density anomaly more deeply buried. Surface gravity anomaly data cannot distinguish between such alternative scenarios, but may provide constraints on geological conditions if certain assumptions are provided, such as a maximal and minimal density variation from the mean density. One important tool to solve the inverse problem are Monte Carlo (MC) methods (see for a review Mosegaard and Sambridge, 2002), which provide a probabilistic approach by obtaining a posterior distribution of possible solutions. MC methods are in particular useful when no analytical forward model is available or if linearization of the relation between data and model is not possible (e.g. for seismic wave models). However, MC methods are computational very expensive and typically require prior defined knowledge about the underlying geological structure and fine-tuned parameter distribution.

Another promising way to infer the geophysical properties is by simultaneously interpreting multiple sensor measurements using a single model. The motivation behind this joint inversion is to provide a better constrained joint solution of multiple, and often distinct sensor types, rather than taking individual solutions that only satisfy their aspect of data on their own. In case these multiple sensors are sensitive to different aspects of the
geology, e.g., density and magnetic susceptibility, we can take advantage of the statistical properties of a model that simultaneously need to satisfy two or more independent measurements. The resulting ‘boost’ in information is dependent on the strength of the correlation between the different physical properties that are measured, but if statistically properly taken into account, this approach can offer a much greater value than the ‘sum of their parts’.

Multiple methods have been developed to tackle the joint inverse problems: Bosch et al. (2006) have used Monte Carlo methods for joint gravity and magnetic 3D inversion. Fedi and Rapolla (1999) apply a 3D method that provides gravity and magnetic field information along the vertical direction for retrieving deep sources. More recently, 3D stochastic joint inversion of gravity and magnetic data has been developed by Shamsipour et al. (2012), which is based on cokriging and applied to estimate density and magnetic susceptibility distributions from field data. An alternative approach is to use a GP, which is a powerful non-parametric Bayesian method to solve joint inverse problems (Reid et al., 2013). The main practical advantage of a probabilistic (Bayesian) approach is that predictions are described by posterior distributions for each location, 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 et al., 2003; Moorkamp et al., 2011).

In this paper we propose a probabilistic framework that combines coherent joint inverse problems and multi-objective optimisation given an expensive cost function. This technique is applied to jointly solve multi-linear forward models of distinct sensor data to 3D-geophysical properties using sparse Gaussian Process kernels while taking into account the correlations between distinct geological aspects. In general, the method provides a very flexible approach for a large range of sensor fusion problems and is computationally efficient enough to rapidly update a model with new data while solving multiple measurement decision problems simultaneously. Our data driven approach does not attempt to include prior knowledge of any geological structure and returns the probability distribution of 3D reconstructed geophysical properties based on sensor data input only. The resulting 3D cubes and visualizations provide a suitable basis to guide further exploration or to generate subsequent models using additional geological constrains.

The outline of this paper is as follows. Section 2 describes the underlying probabilistic framework and computational methods for geological inversion and strategies for Bayesian optimisation. The method is then evaluated on a set of synthetic data as described in Section 3 and tested on real case scenario in Section 4. The results, limitations, and further applications are discussed in Section 5.

**METHODOLOGY**

Our approach for joint inversion and optimisation builds upon a fully probabilistic model, including likelihood objectives for model selection and complete uncertainty propagation through all stages of data processing: from prior selection and input, forward model inversion, to the final sensor optimisation and real world model output. Figure 1 shows a graphical model of the developed probabilistic framework.

In general the relationship between a physical system (or its model parameters) $\Phi$ and
Figure 1: Graphical model 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) maximization of acquisition function (objectives) to allocate optimal new sensor location and type, (4) New acquired data are combined with existing data; process repeats until maximum number of iterations is achieved.

The observed sensor data $y$ is described by a forward model

$$y = G(\Phi)$$  \hspace{1cm} (1)

given the transformation function $G$. For a linear system such as, e.g., the gravitational or magnetic field, the problem reduces to

$$y = G\Phi$$  \hspace{1cm} (2)

where $G$ is the transformation operator (often described as matrix). To infer the underlying parameters of the physical system, the inverse operator, $G^{-1}$, is required. However, direct matrix inversion is only possible for a limited range of cases ($G$ must be a square matrix and non-singular, i.e. $\text{det}(|G|) \neq 0$), and in most applications $G$ is not even a square matrix. To solve the inverse problem for the more general case, we deploy a Bayesian framework with GP priors over the physical properties $\Phi$ of interest (e.g., rock properties). An important advantage of the Bayesian method is that it generates a predictive distribution with a mean $\mu(\Phi)$ and variance $\sigma^2(\Phi)$, which are prerequisites for Bayesian optimisation with respect to, e.g., information gain from a new measurement.

**Linear Forward Models**

In this work we apply linear forward models to infer geological rock properties (i.e., density and magnetic susceptibility) in 3D (Latitude, Longitude, Depth) from 2D (Latitude, Longitude) sensor-mesh data, which are typically placed on top of the earth surface or from
aerial surveys. The sensors measure the variation of the gravitational and magnetic field: the gravitational field variations are driven by the density variations of the material, while magnetic field variations are described by the superposition of Earth’s magnetic field and the induced secondary fields caused by magnetisation of materials, which depends upon their magnetic susceptibility and is linear as first-order approximation (neglecting residual remanent magnetisation).

In practice, the magnetic field information is reduced to the residual between natural and induced field — defined as anomalous magnetic field component. The gravitational and magnetic forward model, \( T_{\text{grav}} \) and \( T_{\text{magn}} \) can be determined analytically by using Li’s tractable approximation (Li and Oldenburg, 1996, 1998) for a 3D field of prisms of constant susceptibility and density, and we apply this prism shape model to compute the corresponding sensor sensitivity for gravity and anomalous magnetic field related to each prism cell (see, e.g., Nagy et al., 2000; Reid et al., 2013).

**Gaussian Processes**

GPs are a flexible, probabilistic approach using kernel machines with non-parametric priors, and are successfully used in a large range of machine learning problems (see Rasmussen and Williams, 2006 for more details). An important advantage of using GPs for data processing and machine learning is their ability to propagate consistently uncertainties from input to output space under the Bayesian formalism, which is an essential requirement for Bayesian optimisation. A GP \( f(x) \sim \mathcal{GP}(\mu(x), k(x,x')) \) is completely determined by its mean \( \mu(x) = \mathbb{E}[f(x)] \) and covariance \( k(x,x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x'))] \), essentially placing a multivariate Gaussian distribution over the space of functions that map the input to the output, or informally, to measure the similarity between points as function of, e.g., their distance (the kernel function) and to predict a Gaussian distribution over \( f(x^*) \) (i.e., a mean value and variance) at any new sampling location (unseen point) \( x^* \) from training data.

GPs are mathematically closely related to many well-known models, including spline models, Bayesian linear models, and, to some extent, support vector machines and deep neural networks (under certain conditions, i.e., a certain covariance kernel and infinite hidden layers). Another advantage of GPs is, as for other kernel methods, that their marginal likelihood function is well defined by the values of their hyper-parameters, and can thus be optimized. This reasoning about functions under uncertainty and their well-tuned interpolation character allows GPs to work extremely well for sparse data, but can come with an expensive computational cost function for large data (\( N > 10000 \)) as it scales with \( \mathcal{O}(N^3) \). However, recent advances in approximate (variational) inference as well as using sparse covariance functions make GPs also applicable for big data problems by reducing the complexity to \( \mathcal{O}(N^2) \) and \( \mathcal{O}(M \log M) \), respectively, where \( M \) is the number of inducing points with \( M << N \). For comparison, the computational time to train deep neural networks scales typically with \( \mathcal{O}(CN) \), where \( C \) is the number of layers and \( N \) the average number of perceptrons per layer.
Choice of Covariance Functions

The choice for an appropriate covariance function is important, as the GP’s output directly depends on it. There are many stationary (invariant to translation in input space) and non-stationary covariance functions (for a review see Rasmussen and Williams, 2006); the former are typically defined by radial basis functions (RBF) such as the squared-exponential covariance function,
\[
k(x, x') = \sigma_0^2 \exp\left(-\frac{1}{2l^2} (x - x')^2\right)
\]
which depends on the radial distance between points \(x\) and \(x'\) and at least two parameters, one defining the length-scale \(l\) (multiple dimensions can have multiple length-scales) and the other the signal variance \(\sigma_0^2\). These parameters of the covariance function are referred to as the hyperparameters \(\Theta\) of the GP, which can be either given by a fixed covariance scale and noise, or learned from data by optimising the marginal likelihood. The latter allows us also to marginalise over a range of GP hyperparameters such as the length-scale parameters \(l\). Computationally this can be done by defining a grid of hyperparameter values and then weight the posterior distribution with the corresponding marginalised likelihood values. In general, it is advisable to compute the joint inversion at least for a range of \(l\) parameters and to compare the reconstructed models against each other.

To handle the computational problem of inverting a large covariance matrix, Melkumyan and Ramos (2009) proposed an intrinsically sparse covariance function \(K(r, l, \sigma_0) =\)
\[
\begin{cases}
K(r, l, \sigma_0) = \sigma_0 \left[\frac{1}{3}(2 + \cos(2\pi \frac{r}{l}))(1 - \frac{r}{l}) + \frac{1}{2\pi} \sin(2\pi \frac{r}{l})\right], & \text{if } r < l. \\
0, & \text{if } r \geq l.
\end{cases}
\]
where \(\sigma_0 > 0\) is a constant coefficient, \(l > 0\) is a given scale and \(r\) is the distance between the points: \(r = |x - x'|\).

Multi-Kernel Covariance Functions

To take fully into account cross-covariances between multiple model parameters (e.g., rock density and magnetic susceptibility), we construct cross-covariance terms between all kernel pairs. One important requirement for constructing cross-covariance terms is that they must be defined to be both positive semi-definite and informative; for an overview how to construct such a matrix in detail see Melkumyan and Ramos (2011). Unless explicitly mentioned, all subsequent analysis is based on sparse covariances. Following Melkumyan...
sparse-sparse cross-covariance terms are defined as

\[
k_{S_1 \times S_2}(r; l_1, l_2) = \frac{2}{3\sqrt{l_1l_2}} \left[ l_{\text{min}} + \frac{l_3^3}{\pi^2 l_{\text{max}}^2 - l_{\text{min}}^2} \right] \times \sin \left( \pi \frac{l_{\text{min}}}{l_{\text{max}}} \right) \cos \left( \frac{2\pi r}{l_{\text{max}}} \right)
\]

if \( r \leq \frac{|l_2 - l_1|}{2} \)

\[
k_{S_1 \times S_2}(r; l_1, l_2) = \frac{2}{3\sqrt{l_1l_2}} \left[ l_{\text{mean}} - r + \frac{l_3^3}{2\pi(l_1^2 - l_2^2)} \sin \left( \frac{\pi}{2}(l_1^2 - l_2^2) \right) \right]
\]

if \( \frac{|l_2 - l_1|}{2} \leq r \leq \frac{|l_1 + l_2|}{2} \)

\[
k_{S_1 \times S_2}(r; l_1, l_2) = 0 \quad \text{if } r \geq \frac{l_1 + l_2}{2}
\]

with \( l_{\text{mean}} = (l_1 + l_2)/2, l_{\text{min}} = \min(l_1, l_2), \) and \( l_{\text{max}} = \max(l_1, l_2). \) The full covariance matrix can then be constructed by setting sparse covariance functions on all diagonal block elements and sparse-sparse cross-covariance functions on all non-diagonal block elements, e.g., for three cross-correlated covariance functions, the complete covariance matrix is given by

\[
K = \begin{bmatrix}
k_{S_1} & w_{1,2}k_{S_1 \times S_2} & w_{1,3}k_{S_1 \times S_3} \\
w_{1,2}k_{S_2 \times S_1} & k_{S_2} & w_{2,3}k_{S_2 \times S_3} \\
w_{1,3}k_{S_3 \times S_1} & w_{2,3}k_{S_3 \times S_2} & k_{S_3}
\end{bmatrix}
\]

(6)

with the cross-covariance amplitude terms, \( w_{1,2}, w_{1,3}, w_{2,3}, \) which can be determined by optimising the marginal log-likelihood (see equation (15)) in conjunction with all other GP hyperparameters (amplitude, kernel length-scale). While these correlation terms between different geophysical properties can be also measured independently, the joint optimization has the advantage to find the best possible solution over the complete parameter space.

### Joint Inversion

Joint Inversion within a Bayesian framework corresponds to computing the posterior distribution of an ensemble of quantities of interest \( \phi_{n=1}^{N} \) (e.g. \( N \) distinct rock properties) at the locations of interest \( x_{m=1}^{M} \) given \( M \) sensor observations \( y \). If we assume a GP prior over the functions \( \phi \) and an isotropic likelihood model:

\[
\phi(x) \sim \mathcal{GP}(0, K_{\phi,\phi'}(x, x'))
\]

\[
y = f + \eta \quad \text{with} \quad \eta \sim \mathcal{N}(0, \sigma_s^2 I)
\]

(7)

where \( K_{\phi,\phi'} \) is the covariance function in the parameter space including cross-covariances between distinct quantities of interest \( (\phi, \phi') \), and \( \sigma_s^2 \) is the noise variance of the sensor observations. This translates to a Gaussian prior over the quantities of interest:

\[
\phi \sim \mathcal{N}(\phi|0, K_{\phi,\phi'}),
\]

(8)
where $K_{\phi\phi'}$ is the $(N \times N) \times (M \times M)$ covariance matrix obtained by evaluating the kernel $K_{\Phi,\Phi'}(x, x')$ at the locations $(x, x')$ of interest and the covariance between distinct quantities $\Phi$ and $\Phi'$. The covariance structure in the sensor space is computed as:

$$K_{y\phi} = G K_{\phi\phi'}$$

$$K_{yy} = G K_{\phi\phi'} G^T + \sigma_s^2 I,$$  \hspace{1cm} (9, 10)

where the joint forward model $G$ is an operator matrix with diagonal block elements representing the distinct forward model operators and non-diagonal set to zero. For instance a joint forward model with three different forward models ($G_{1,2,3}$) — such as gravity, magnetic, and drill-cores — is represented by the matrix

$$G = \begin{bmatrix}
[G_1]_{j,m} & [0]_{k,m} & [0]_{l,m} \\
[0]_{j,m} & [G_2]_{k,m} & [0]_{l,m} \\
[0]_{j,m} & [0]_{k,m} & [G_3]_{l,m}
\end{bmatrix},$$

(11)

with $M$ locations of interest and three sensor grids with $J$, $K$, and $L$ number of sensors, respectively. Note that the uncertainty $\sigma$ of the sensor observations $y$ is included in equation and can be either a constant $\sigma$ (same uncertainty for all sensors) or a vector that is given by the individual uncertainty of each sensor observation. In case the sensor uncertainty is unknown, it can be estimated in conjunction with the other GP hyperparameters by maximizing the marginal log-likelihood (see equation 15).

The predictive distribution for all quantities of interest $\Phi$ is given by

$$\phi|G, y \sim N(\mu_{\phi|y}, \Sigma_{\phi|y}) \text{ with:}$$

$$\mu_{\phi|y} = K_{y\phi} K_{yy}^{-1} y,$$

$$\Sigma_{\phi|y} = K_{\phi\phi} - K_{y\phi} K_{yy}^{-1} K_{y\phi'},$$

(12, 13, 14)

where $\mu$ is the mean at each location $m$ and for each quantity $n$ (matrix with $M \times N$ elements), and $\Sigma_{\phi|y}$ the covariance matrix with $(M \times N)^2$ elements. Thus, different sensor types can be fused and processed jointly by taking advantage of cross-correlating the target quantities via a joint covariance function. This results in multivariate outputs and their respective uncertainties, which are important for subsequent decision-making processes as outlined in the next section. Another advantage of the Bayesian framework is the ability to calculate the marginal likelihood,

$$L(\Theta) = -\frac{1}{2} y^T K_{yy}^{-1} y - \frac{1}{2} \log |K_{yy}| - \frac{\log 2\pi}{2} \frac{M}{2} \sum_{i=1}^{M} N_i,$$

(15)

to optimize the parameters $\Theta$ of the GPs, which include the hyper-parameters of the covariance functions $k(x, x')$, the noise variances $\sigma_s^2$, and optionally the sensor uncertainties and cross-covariance amplitude terms (for more details see section Hyper-Parameter Learning for GP Inversion).
Multi-Objective Bayesian Optimisation

Bayesian Optimisation (BO) is a powerful framework for finding the extrema of objective functions that are noisy, expensive to evaluate, or do not have a closed-form (e.g., black-box functions), or have no accessible derivatives. Bayes’ rule is applied to derive the posterior estimate of the unknown function, given sample observations (e.g., sensor measurements) and prior knowledge. The prior is a GP model of the function, whose mean captures the estimated value of the function, and whose variance determines the level of uncertainty of the prediction in that particular location. The posterior is then used to query for the next most promising point based on objectives that are defined in an acquisition function $g$, which guides the search for a user-defined optimum. This is equivalent of finding the $x^* = \arg \max_x g(x)$. The BO algorithm maximizes this function that quantifies the benefit of choosing a specific location to be sampled. Once the maximum of $g$ is found and new measurements is taken at the proposed position, the model is updated with the new data, and the BO algorithm proposes a new point at each iteration step until the objectives are achieved or a maximum of iterations is reached.

**Acquisition Functions**

The key of BO is the acquisition function $g$, which typically has to balance between (1) exploration, i.e., querying points that maximise the information gain and minimize the uncertainty of a model (e.g., of a geophysical model of a site), (2) exploitation, i.e., querying points that maximise the reward (e.g. concentrating search in the vicinity locations with high value such as minerals), and (3) minimize the number of samples given an expensive cost function for any new measurement (e.g., energy-cost of active sensors, or costly drill-cores). Several acquisition functions have been proposed (for a practical overview, see, e.g., Brochu et al., 2010) that make use of the mean and variance of the posterior, such as the probability of improvement, the expected improvement and the Upper/Lower Confidence Bound (UCB/LCB). The latter evaluates sample locations $x$ in the form

\[
ULCB(x) = \mu(x) \pm \kappa \cdot \sigma(x),
\]

with $+(-)$ indicating the Upper (Lower) CB, the mean value for the prediction $\mu(x)$, the standard deviation $\sigma(x)$ of the GP posterior at $x$, and the parameter $\kappa$ that balances the exploration-exploitation trade-off.

Finding the optimal $\kappa$ is not trivial and may depend on the user-specific objectives, or can be constraint by its cumulative regret in terms of maximal information gain (Srinivas et al., 2009). If the parameter $\kappa$ is not dependent on any user-specific objectives, a simple estimate of $\kappa$ can be obtained under assumption of equal weight between the mean and the variance:

\[
\kappa = \frac{\sigma(\mu(x))^2}{\sigma(\sigma(x))^2}.
\]

Alternatively, it is possible to optimize the UCB acquisition function over a uniform range of $\kappa$ values and to select the most frequent solution. Moreover, various extensions to the acquisition function have been proposed to accommodate the increasing complexity of remote sensing applications (e.g., environmental monitoring, Marchant and Ramos, 2012).
Multiple Objectives: Pareto Front and Expected Hyper-volume Improvement

Typically, multiple objectives are conflicting, and no solution may exist where all objectives can be minimized at once. In this case the task is then to find the set of optimal compromise solutions, where at least one objective improves without reducing any of the other objectives. The optimal set of these solutions is called Pareto set, named after Vilfredo Pareto, who applied this concept to study economic efficiencies (Cirillo, 2012). Graphically it can be described as a frontier of optimal points in a hyper-dimensional space, where each dimension corresponds to one objective. The frontier defines all points of a sample that dominate over others, so that it is not possible to improve an objective without degrading the others.

To query a new sample point, we calculate the Expected Hyper-volume Improvement (EHVI, Emmerich, 2005) with regards to the current optimal Pareto set. The Pareto Front and EHVI is illustrated in Figure 2 for two objectives, which reduces to the problem to query the sample point that maximizes the Area of expected improvement that is spanned by the new Pareto Front. This is straightforward if the number of evaluation points is not too high, otherwise computationally efficient grid-search or variational methods can be applied to speed-up the computation. For larger number of objectives $N_{Obj}$ and evaluation points, the integration of the EHVI becomes computationally very expensive as it increases exponentially with $N_{Obj}^3$ and several decomposition methods have been proposed (Emmerich and Klinkenberg, 2008; Bader and Zitzler, 2011; Feliot et al., 2017) by partitioning the integration domain into hyper-rectangles or using sequential Monte Carlo techniques.

Multi-objective optimisation is particularly important for joint inverse problems, which generate posteriors via a GP prior with multiple outputs such as different geophysical properties. To find the optimal new sampling point, we calculate the EHVI given $N$ objectives. In this paper each objective function $f_i$ is defined by maximizing the UCB

$$UCB_i(x) = \mu_i(x) + \kappa_i \cdot \sigma_i(x) - \gamma_i c(x)$$  \hspace{1cm} (18)

with the mean value for the prediction $\mu_i(x)$, the variance $\sigma_i^2(x)$, and a cost function $c(x)$, which is defined by the cost of obtaining a measurement at the sample point $x$. The parameter $\kappa_i$ and $\gamma_i$ define the trade-off in exploration-to-exploitation and gain-to-cost, respectively. For example, maximizing the mean value can be beneficial if the goal is to sample new data at locations with high density or mineral content, and not only where the uncertainty is high. At each iteration of the BO, new potential sample points are evaluated given their EHVI. The BO algorithm stops if none of the proposed new sample points improves any of the objectives without reducing any of the other.

EXPERIMENTS WITH SIMULATED DATA

Geophysical Forward Models and Sensor Data Characterisation

To test our proposed BO algorithm, we first need to define the set of forward models that transform the localised measurement of a sensor grid into a 3D representation of geophysical properties of a region. Inversion can be solved for linear forward models which follow the relation $y = G\phi$ where $y$ is the vector of observations; $\phi$ is the vector of unknown parameters at the 3D locations; $G$ is a known $N \times M$ sensitivity matrix that relates the values of the geophysical property at different locations to the observations. The most common
geophysical linear forward models are gravity and magnetic forward models as described in the following.

**Gravity Forward Model**

The gravity forward model describes the relation between the earth’s density distribution and the gravitational force measured by sensors above the surface by simulating Newton’s law. The gravity forward model is defined by using Li’s tractable approximation for a 3D field of constant density prisms (Li and Oldenburg [1998] and can be determined analytically (Nagy et al., 2000). A detailed form of the sensitivity matrix calculation is given in Appendix A.

**Magnetic Forward Model**

Magnetic susceptibility depends on the Earth’s magnetic mineral content such as iron-rich minerals including magnetite and haematite below the surface and is measured by the response of their magnetic dipoles induced by the Earth’s magnetic field. The induced magnetic field calculation uses Li’s tractable approximation for a 3D field of prisms of constant susceptibility (neglecting residual permanent magnetism), and Earth’s local magnetic field. The joint GP inversion takes into account a covariance that exists between density and magnetic susceptibility, which does provide additional insight into the region’s geophysical properties. For instance, magnetite has a high magnetic susceptibility and density, while haematite has a similar density to magnetite, but a much lower magnetic susceptibility.
BO sampling

We define the goal of the BO experiment to find the optimal location where to place a new sample measurement as given by a set of acquisition functions. The experiments are carried out in the following steps:

1. Generate a simulated 3D voxel cube with geophysical structures given by their density and magnetic susceptibility (described by the vector $\phi$).

2. Calculate response $y$ of gravity and magnetic sensors at the surface, and of pre-existing drill-core samples.

3. Solve joint GP inversion to calculate the posterior distribution of geophysical properties in 3D as defined by their mean and variance for each voxel (equations 13 and 14).

4. Find the $x, y$ location for new drill-core that maximizes the BO acquisition function(s) (e.g., given by UCB, equation 18 or Pareto optimisation).

5. Add selected drill-core data to measurements and repeat Steps 2 to 4 until the maximum of iterations (here 25) or final objective is achieved.

While more drill-core parameters can be added (drill length, dip-, and azimuth-angle), we assume for simplicity in this study vertical drill-holes with constant length equal to the cube height.

Synthetic Model Setup

To test the joint GP inversion and the proposed BO algorithm, a set of four different synthetic geophysical models are created as shown in Figure 3:

1. Even Cylinder Model: Two even cylinder dipping bodies, which consist of slabs of dense, magnetically susceptible material, and are surrounded by a uniform low density and magnetic medium.

2. Uneven Cylinder Model: Similar to the two even cylinder dipping bodies as described above, but with density ratio 2:1 and different density-to-magnetic susceptibility correlation coefficients.

3. Layer model: Multiple-layer folded structure, with the central layer consisting of dense, magnetically susceptible material surrounded by two layers of material with lower density and magnetic susceptibility.

4. Clumpy model: Four spherical high density and magnetic bodies, at different depths and with different density-to-magnetic susceptibility correlation coefficients.

All of the simulated environments include three sensor types: (1) 2D surface gravity measurements, (2) 2D surface magnetic sensors, and (3) vertical drill-cores for direct local measurements of rock density.
Figure 3: The 3D synthetic models’ density distribution for the four different models: two-even cylindrical body model (Model 1), two-uneven cylindrical body model (Model 2), three-layer model (Model 3), and clumpy model (Model 4). The magnetic susceptibility follows a similar distribution with varying correlation coefficients (see text). Each of the four models shows (a) the volumetric 3D density across an area of of 20 × 20km with 10km depth, (b) the mean projected density, and (c) the core model object bodies.

Geometry and Voxelisation

The region’s geophysical properties are reconstructed in a cube with an extent of 20 km × 20 km in horizontal and 10 km in vertical direction. The gravitational and magnetic sensors are distributed over a 20 × 20 uniform distributed sensor grid. The final reconstructed scene is represented by a voxel grid of size 20 × 20 (horizontal) × 10 (vertical), resulting in total of 4000 voxels. For visual simplification the voxel centre positions $x, y$ are aligned with the
positions of the surface sensor grid and we assume that the sensor grid is on a flat surface at constant surface height.

**Hyper-Parameter Learning for GP Inversion**

As given by equation [15], the GP inversion framework provides a calculation of the marginal likelihood, which allows us to optimize the GP hyperparameters, i.e., the length-scale \( l \) and the noise variance \( \sigma_s^2 \) as regularisation term. As optimisation strategy we apply a global optimisation (e.g., NLopt) as starting point for subsequent local optimisation (e.g., COBYLA [Powell, 1994]) to obtain an optimum to a greater accuracy. We assume the same hyperparameter length-scale for the entire cube (see Discussion for including stationary dependence of hyperparameters). Using the synthetic test bodies as described in the previous section, we have tested a variety of GP non-stationary kernels, including the squared exponential, Matern32, as well as sparse kernels. The selection of the optimal kernel and GP hyperparameters is based on their marginal likelihood and the comparison between reconstructed and true geophysical properties. Overall, the sparse covariance kernel performed best given the most robust solution over a wide range of synthetic bodies, with an optimal GP length-scale \( l \) of five voxel lengths (equal one quarter of the region’s length) and \( \sigma_s^2 \) of half the standard deviation of the normalised sensor measurements. The correlation amplitude \( w_{D-M} \) between density and magnetic susceptibility depends on the geochemistry of the rock type, and we have included in our simulation a range of correlation coefficients from 0.0 (no correlation at all) to 1.0 (complete correlation).

**Results on Dipping Body Model**

The synthetic model consists of two dipping bodies which are two separate, horizontally aligned cylinders of dense, magnetically susceptible material as shown in Figure 3. The surrounding medium is uniform with low density and magnetic susceptibility. In the first scenario the two dipping bodies have the same mass and same rock chemistry with \( w_{D-M} = 1 \). The second scenario consists of two dipping bodies with different mass and distinct rock chemistry with \( w_{D-M} = 1 \) for one cylinder and \( w_{D-M} = 0.5 \) for the other. The gravity and magnetic sensitivities are calculated from their forward models and are shown in Figure 4. An example solution of the joint inversion is shown in Figure 5 given only one central vertical drill-hole.

We simulate exploration by drilling yielding a total of 25 drill-cores. The posterior of the geophysical properties, density and magnetic susceptibility, is updated and evaluated after each step of new drill-core data added. Figure 6 shows the reconstructed density model after 2, 8, 16, and 25 drill-cores. A video of the complete sequence is attached in the supplementary material.

**Dependence on Exploration-Exploitation Parameter \( \kappa \)**

We test the BO algorithm for a range of \( \kappa \) values (see Equation [15]) in terms of their speed to converge towards the true model, as shown in Figure 7. The difference between reconstructed and true model is given as Root Mean Square Error (RMSE) and becomes
Figure 4: (a) The gravitational anomaly field and (b) magnetic anomaly field on the surface as measured by sensors for scenario 1. (c) The gravitational anomaly field and (d) magnetic anomaly field for scenario 2, see text.

smaller with increasing number of drill-cores. The best results are typically achieved with a $\kappa$ of 1-3. While we kept $\kappa$ fixed for each BO experiment, in principle it is possible to employ an adaptive $\kappa$ that changes with the number of drill-cores (e.g., to, change from exploration to exploitation after a certain threshold of information gain is reached). Note that any parameter can be updated during the exploration. Moreover, it is possible to marginalize the posterior of the reconstructed model over a wide parameter range of GP length-scales, either using a uniform space or weighted by the corresponding marginalized log likelihood.

Including Non-uniform Drill-Core Cost Function

In the following we test how our optimisation approach performs if we add a non-uniform cost-function that depends on the location price where the drill-core is sampled. Figure 8 shows an example of a spatial distribution of a cost-function, which can in practice, e.g., represent different drill-costs for farm land and unused land. We compare the results of the BO UCB method (equation 18) for two different weightings of costs ($\gamma = 0.5$ and $\gamma = 1.0$) to a standard non-BO method that relies on random drill-core sampling that is probabilistically weighted with the measured gravitation field divided by the cost at this location. In addition we show the average of ten such random drill-core samples. The
Figure 5: The reconstructed density (Density Distribution) and magnetic susceptibility model (MagSus Distribution) based on only gravity, magnetic sensors and one central drill-core (red line), shown without correlation between density and magnetic susceptibility (No correlation, $w_{D-M} = 0$) and with $w_{D-M} = 1$ (100% correlation) between density and magnetic susceptibility. Sub-panels (a,b,c) are defined as in Figure 3.

The experiment shows that the BO method with full cost weighting significantly outperforms the standard non-BO method at roughly the same cumulative costs of drilling (see Figure 8). The BO method with full cost weight shows only a slow increase in cost until a larger jump in cost is necessary to significantly reduce the RMSE further (BO step 17-18 and 20-22). If the objective function has a smaller weight on costs ($\gamma = 0.5$), the RMSE reduces faster but at a higher price.
Figure 6: The reconstructed density after 2, 8, 16, and 25 BO steps for drill-core measurements based on the UCB method with $\kappa = 1$. Sub-panels (a,b,c) are defined as in Figure 3.

Simulation Results: Multi-Layer Model

This synthetic model simulates a folded geological formation that consists of three layers with different density and magnetic susceptibility. The three layers are surrounded by a low density and magnetic medium. The layers are orientated parallel to the surface while being folded halfway in y-direction as shown in the middle panel of Figure 3. Given that there is no clear separated concentration in density, the BO algorithm proposes drill-cores fairly uniformly, starting at the cube side with the lowest layer depth-to-surface and propagates drill-cores towards the other half of the cube. The reconstructed density distribution is shown in Figure 9.
Simulation Results: Multi-Clump Model

The clumpy model consists of four spherical high density and magnetic clumps that are well separated from each other as shown in bottom panel of Figure 3. Two of the four clumps are placed lower in vertical height than the other two. These two pairs are subdivided again in having different rock chemistry ($w_{D-M} = 1.0$ or $w_{D-M} = 0.5$). The best reconstructed model, as based on the final RMSE after 25 drill-core sampling, is achieved using the UCB BO method. The reconstructed density distribution of this model is shown in Figure 10.
Evaluation of Optimisation Strategies

To evaluate the efficiency of the proposed BO methods, we show in Figure 11 the results of our simulated experiments for a sampling series of 25 drill-core locations. The RMSE values are calculated for the four different synthetic models (see model description above), and the performance of multiple BO functions are compared to two mock traditional non-BO approaches: The first one samples from a random selection of 25 uniform evenly spaced drill-core locations across the entire model space. The second non-BO method selects drill-core locations from a random distribution that is probabilistically weighted with the gravity and magnetic field sensor measurement. These two methods are compared to our BO using (1) Pareto EHVI, (2) the sum of the two UCB function over density and magnetic properties given as

\[
UCB(x) = \mu_{\text{density}}(x) + \kappa \cdot \sigma_{\text{density}}(x) \\
+ \mu_{\text{magnetic}}(x) + \kappa \cdot \sigma_{\text{magnetic}}(x),
\]  

(19)
Figure 10: The reconstructed density model of the clumpy model after 25 drill-core BO samples. Sub-panels (a,b,c) are defined as in Figure 3.

(3) the expected improvement (EI, Mockus et al., 1978, Jones et al. (1998), which is defined by maximizing the probability of improvement (PI) over the incumbent $f_{\text{max}}(x_{\text{drill}})$ with

$$PI(x) = P(f(x) \geq f_{\text{max}}(x_{\text{drill}})) = \text{NormCDF} \left[ \frac{\mu(x) - f_{\text{max}}(x_{\text{drill}})}{\sigma(x)} \right],$$

(20)
given the normal cumulative distribution function $\text{NormCDF}$, and (4) optimising only the variance to minimize the total uncertainty of the reconstructed cube.

The results of these experiments are presented in Figure 11, which shows that the two BO strategies Pareto and UCB consistently outperform the non-BO methods by roughly a factor of two in final accuracy and speed of convergence towards the true model. Only in case that the model consists of continuous smooth layers, the Variance-only BO method has the fastest convergence speed, but eventually converges for larger drill-core samples towards the same RMSE values as Pareto and UCB.
Multi-Objective Bayesian Optimisation

Figure 11: Results of the RMSE values between reconstructed and true model as measured for different synthetic models: (a) even and (b) uneven cylinder model, (c) layer model, and (d) clumpy model, as described in text. Multiple BO drill-core selection strategies are compared against each other in terms of their performance for selecting additional drill-cores: 1) Pareto, 2) UCB, 3) Expected Improvement (EI), and 4) Variance only. Additionally, two non-BO standard methods are shown, which are 5) randomly sampling drill-cores from a uniform evenly spaced grid (Random - Uniform) and 6) random sampling with probabilistic sensor data weighting (Random - Weighted). In all four test-sets, Pareto and UCB converge faster towards the true model than the two standard methods.

CASE STUDY ON THE MOOMBA REGION

Data Measurements and Geometry

To test the proposed joint BO method for a real geophysical formation scenario, we selected as test-set a basement in the vicinity of the Moomba gas field in NE South Australia. The survey data was collected over an area of 36 km × 36 km by 12 km deep, bounded by 410000-445000E, 6886000-6921000N. Regional sensor data are available in the form of Bouguer anomaly gravity and total magnetic field anomaly at a number of discrete locations and is interpolated to a uniform grid as shown in Figure 12. The model cube is divided into nx=24 by ny=24 by nz=8 cubes of dimension (1500m)³ with a total number of 4608 voxels.
Figure 12: (a) The interpolated gravity Bouguer and (b) magnetic anomaly field and sensor locations (circles) for the Moomba region. (c) the average vertical-profile of density and (d) magnetic susceptibility for the Moomba region as obtained from the mean of the Obsidian posterior distribution. Note that the central granite intrusion is less dense than the surrounding sediments and also has a lower magnetic susceptibility.

Test Setup and Existing Geophysical Models

To simulate and evaluate our bayesian optimisation strategy, we need to add drill-hole data after each BO step and evaluate the inversion results to the ground truth that holds the 3D geophysical properties. Since in practice it is not possible to obtain a complete large-scale 3D measurement cube of geophysical properties, we use instead the best available model as approximation of the true geophysical properties for the Moomba region in the southwestern portion of the Cooper Basin in South Australia. The portion of the Moomba region modeled here is characterized by a granite intrusion. Granites in this region are of economic interest for their geothermal energy potential and the associated sandstone-hosted Uranium deposits (see Huston and van der Wielen [2011], Khair et al., [2015]). A 3D model for the Moomba region has been generated previously with the Bayesian inference tool Obsidian (see McCalman et al., 2014, Beardsmore et al., 2016), which uses Markov Chain Monte Carlo (MCMC) methods to sample from a joint probability distribution of rock properties for a set of parameterized layered models. Obsidian includes a range of forward models such as gravity, magnetic, 2D seismic reflection profiles as well as petroleum wells providing local constraints on formation depths and temperatures. It can incorporate prior knowledge of the geological structure and assumes a certain number of layers, whose rock properties and boundaries are constrained by Gaussian priors. Overall, it is currently the best 3D model available for this site.

For this test we take the mean of the final Obsidian posterior distribution over the 3D space as a proxy of the "true" basement model, i.e a granite intrusion surrounded by
sedimentary deposits, from which we can sample drill-core measurements. Note that the true basement model is not seen by the BO selection algorithm and that only selected drill-hole data from this model are added to the inversion task along with the real gravity and magnetic anomaly field measurements for the Moomba region as shown in Figure 12.

**Experiment Results**

The simulated BO exploration experiment queries locations for new drill-cores given the gravity magnetic data and drill-core measurements obtained from the 3D Obsidian model. In total, 25 vertical drill-cores are sampled iteratively before the experiment stops. The sampling scheme and process of reconstructing the density and magnetic properties as function of drill-core is shown as videos in the supplementary material. The final reconstructed density and magnetic properties are shown in Figure 13 after 25 drill-core measurements.

The experiment evaluates multiple optimisation schemes and calculates (1) the total RMSE between reconstructed and the true basement model as gauge for accuracy, and (2) the predictive log-likelihood, which considers also the predicted uncertainty $\sigma_i$ (square root of diagonal elements of the predicted covariance matrix, see equation 14),

$$\ln L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp \left( \frac{-(y_i - y_i^*)^2}{2\sigma_i^2} \right)$$

$$= -0.5 \sum_{i=1}^{N} \left[ \ln(2\pi \sigma_i^2) + \frac{(y_i - y_i^*)^2}{\sigma_i^2} \right],$$

where $(y_i - y_i^*)$ is the residual between ground truth and the reconstructed mean predictive value for each voxel $i$. The performance evaluation is shown in Figure 14 for the two BO methods, Pareto and UCB, and compared to the two random based sampling methods. The results show that both BO methods outperform the non-BO methods in terms of accuracy, effectiveness, and maximizing the predictive log-likelihood.

**DISCUSSION**

In this paper we develop a probabilistic framework for joint inversion and multi-objective BO to select new sensor measurements, and demonstrate its capabilities on a test-set of synthetic geological models as well as real world scenario by solving the example problem of allocating iteratively new expensive drill-core samples. Our results show that this method can accurately reconstruct from sparse 2D sensor data a 3D model of multiple geophysical properties and their associated uncertainties. Multiple BO sampling schemes are evaluated in terms of accuracy and effectiveness for a wide range of geological models and compared to traditional non-BO sampling schemes. The results demonstrate that BO can significantly outperform traditional non-BO sampling schemes, which may ultimately lead to more efficient and cost-saving sensor selection.

A key feature of the applied method is to consider all cross-variances between geophysical properties by solving simultaneously multiple forward models and to fuse jointly multiple and distinct sensors. Thus, this approach can take full advantage of additional sensor
Figure 13: The reconstructed density (Density Distribution) and magnetic susceptibility (MagSus Distribution) after 25 drill-core sampling. The distributions are shown after subtracting their mean value. Sub-panels (a,b,c) are defined as in Figure 3. The geometry is highlighted by the relatively high densities (d) and magnetic susceptibilities (e), forming a U-shape, reflect the sediments surrounding the central granite intrusion in this region. The average vertical-profiles are shown for the reconstructed density (f) and magnetic susceptibility (g).

information to reveal properties of the reconstructed geophysical model at higher accuracy. Moreover, by using non-parametric Gaussian Process priors, the solution of the inverse problem provides a complete posterior distribution for all geophysical properties under consideration at any location or grid resolution. The reconstructed geophysical properties are described by a mean and variance value at each location (voxel), which are the two key-ingredients for Bayesian optimisation.

The proposed model allows applications to specify priors as additional options: (1) the
Figure 14: Comparison between BO methods and random sampling as function of sampled drill-core number for the Moomba region: (a) total RMSE, (b) predictive log-likelihood (see equation 22 which describes the sum over all voxel probabilities considering the difference between observed and predicted mean values and uncertainties) for density, and (c) the predictive log-likelihood for magnetic susceptibility.

typical length-scale of the geological structure or the spatial resolution of the reconstructed model via GP hyperparameters, (2) the gain/cost parameters in the BO acquisition function, and (3) the correlation coefficients between different geophysical properties. If the application is purely data driven (default option), GP hyperparameters and correlation coefficients are optimized given equation 15. Within the BO sampling iteration process, all parameters are adaptive towards possible changes in objectives over time. Alternatively, in case the aim is to reduce the dependency on certain prior parameters, it is straightforward to marginalize over the corresponding BO or GP hyperparameters, which increases the computational time but can offer a more accurate posterior distribution.

Further improvement in accuracy and computational performance can be achieved by dividing the cube in multiple cube elements that are larger than the maximal spatial correlation length-scale. Solving the inversion step for individual cube elements has the following advantages: (1) to optimise the GP hyperparameter length-scale specifically for each stationary dependent cube element (e.g., for tracing different geological structures), and (2) to update only the cube element(s) where new data are added. Note that the BO will still select the optimal new sensor data from the combined cube but the computational time for optimizing the acquisition function is much smaller in comparison to computing the GP inversion.
While the proposed method of inversion via GP is limited to linear forward models, such as gravity and magnetics, the reconstructed 3D distribution of the geophysical properties can be easily refined further by other probabilistic methods. This subsequent modeling can, e.g., take into account site-specific geological knowledge as well as other measurements that require non-linear forward models (e.g., seismic data), which rely on sampling from model priors or a distribution of possible models rather than solving the inversion directly. Such sampling models (e.g., Obsidian or GemPy, see Shamsipour et al., 2012; Beardsmore et al., 2016, respectively) are computationally very expensive, even after considering MCMC sampling or variational inference, given their need to sample over an almost infinite range of possible models, and are therefore in practice typically constrained by certain assumptions about the underlying geological formation and prior parameterisation. In these cases, the proposed GP method in this paper may significantly help to boost the sampling efficiency by providing the probability distribution of 3D reconstructed geophysical properties as probabilistic sampling constrain or guideline to speed up the sampling scheme.

The BO method as discussed here is not limited to geophysical inversion and drill-core sampling optimization alone, but can be applied to a wide range of sensor fusion problems and allocation problems with expensive cost functions for any new measurement (e.g., costly drill cores). This offers solutions for possible questions such as: Which additional sensor type would provide the maximum gain? What is the optimal mixture of sensors? What is the minimum amount of sensors to be activated given the total energy constraints and how should they be distributed? How to position sensor grids over time if the model state is dynamic or has a moving target? Future research may address these problems in more detail, ideally with a network of probabilistic modeling components where each node is capable of complete propagation of uncertainties from input node to output node.

CONCLUSIONS

One of the most important aspects when dealing with incomplete information is where to acquire new observations, in particular if measurements are very costly or resources are limited. The probabilistic framework presented here combines inverse modeling and multi-objective Bayesian optimisation to find the global optimal solution for allocating new measurement. The joint inversion model uses sparse Gaussian Process kernels and multi-linear forward models to construct 3D-geophysical properties from 2D-sensor data. One key advantage of this method is to solve simultaneously multiple sensor measurements by taking into account the correlations between distinct geological aspect, e.g., density and magnetic susceptibility. Furthermore, the solution of the inverse problem provides a complete posterior distribution for all geophysical properties under consideration at any location or grid resolution.

Multiple optimisation strategies have been evaluated on a set of synthetic and real geophysical data by addressing a specific example of a joint inverse problem, i.e., where to place new drill-core measurements given 2D gravity and magnetic sensor data, while taking into account potential drill-core costs and any pre-existing drill-core data. We conclude that the proposed method can accurately reconstruct 3D geophysical properties and their associated uncertainties, and significantly outperform traditional non-bayesian optimisation sampling schemes, which ultimately leads to more efficient sensor selection, higher information gain, and reduced cumulative costs. Moreover, we find that the model is flexible enough to accom-
modate a variety of prior knowledge, such as the characteristic length-scales of geological structures and the correlation coefficients between different geophysical properties. We propose that the same model can be applied to a wide range of sensor fusion problems and allocation problems - ranging from constraints limiting surface access for data collection to adaptive multi-sensor positioning.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments to improve and clarify this manuscript. This research was supported by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.
APPENDIX A

GRAVITY FORWARD MODEL

The values of the matrix $G_{\text{const}}$ can be determined analytically for a prism by its origin and end coordinates $z^0$ and $z^e$, and the corresponding sensitivity at location $x$ is given by (Nagy et al., 2000; Reid et al., 2013)

$$T_{x,z} = G_{\text{const}} g(z) \left| \frac{z_1 z_2}{z_1^2 + z_2^2 + z_3^2} \right|,$$

where $G_{\text{const}}$ is the gravitational constant, and

$$g(z) = z_1 \log (z_2 + r) + z_2 \log (z_1 + r) - z_3 \arctan \frac{z_1 z_2}{z_3^2};$$

$$z^o = z^o - x; z^e = z^e - x; \text{ and } r = \sqrt{z_1^2 + z_2^2 + z_3^2}.$$  

REFERENCES

Bader, J., and E. Zitzler, 2011, HypE: An algorithm for fast hypervolume-based many-objective optimization: Evolutionary computation, 19, 45–76.

Beardsmore, G., H. Durrant-Whyte, L. McCalman, S. O’Callaghan, and A. Reid, 2016, A Bayesian inference tool for geophysical joint inversions: ASEG Extended Abstracts, 2016, 1–10.

Bosch, M., R. Meza, R. Jiménez, and A. Höning, 2006, Joint gravity and magnetic inversion in 3D using Monte Carlo methods: Geophysics, 71, 4, G153–G156.

Brochu, E., V. M. Cora, and N. 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.

Caselton, W. F., and J. V. Zidek, 1984, Optimal monitoring network designs: Statistics & Probability Letters, 2, 223–227.

Cirillo, R., 2012, The Economics of Vilfredo Pareto: Routledge.

Eidsvik, J., T. Mukerji, and D. Bhattacharjya, 2015, Value of information in the earth sciences: Integrating spatial modeling and decision analysis: Cambridge University Press.

Emmerich, M., 2005, Single- and multi-objective evolutionary design optimization assisted by gaussian random field metamodels: Dissertation, Universität Dortmund.

Emmerich, M., and J.-w. Klinkenberg, 2008, The computation of the expected improvement in dominated hypervolume of Pareto front approximations: Rapport technique, Leiden University, 34.

Fedi, M., and A. Rapolla, 1999, 3-D inversion of gravity and magnetic data with depth resolution: Geophysics, 64, 452–460.

Feliot, P., J. Bect, and E. Vazquez, 2017, A Bayesian approach to constrained single- and multi-objective optimization: Journal of Global Optimization, 67, 97–133.

Gallardo-Delgado, L. A., M. A. Pérez-Flores, and E. Gómez-Treviño, 2003, A versatile algorithm for joint 3D inversion of gravity and magnetic data: Geophysics, 68, 949–959.

Huston, D., and S. van der Wielen, 2011, An assessment of the uranium and geothermal prospectivity of east-central South Australia: Geoscience Australia.

Jeong, S., and S. Obayashi, 2005, Efficient global optimization (EGO) for multi-objective
problem and data mining: Evolutionary Computation, 2005. The 2005 IEEE Congress on, IEEE, 2138–2145.

Jones, D. R., M. Schonlau, and W. J. Welch, 1998, Efficient global optimization of expensive black-box functions: Journal of Global optimization, 13, 455–492.

Khair, H. A., D. Cooke, and M. Hand, 2015, Seismic mapping and geomechanical analyses of faults within deep hot granites, a workflow for enhanced geothermal system projects: Geothermics, 53, 46–56.

Krause, A., A. Singh, and C. Guestrin, 2008, Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies: Journal of Machine Learning Research, 9, 235–284.

Kushner, H. J., 1964, A new method of locating the maximum point of an arbitrary multi-peak curve in the presence of noise: Journal of Basic Engineering, 86, 97–106.

Li, Y., and D. W. Oldenburg, 1996, 3-D inversion of magnetic data: Geophysics, 61, 394–408.

Li, Y., 1998, 3D-inversion of gravity data: Geophysics, 63, 109–119.

Marchant, R., and F. Ramos, 2012, Bayesian optimisation for intelligent environmental monitoring: 2012 IEEE/RSJ international conference on intelligent robots and systems, IEEE, 2242–2249.

McCalman, L., S. T. O’Callaghan, A. Reid, D. Shen, S. Carter, L. Krieger, G. Beardsmore, E. V. Bonilla, and F. T. Ramos, 2014, Distributed Bayesian geophysical inversions: Thirty-Ninth Stanford Geothermal Workshop, 1–11.

Melkuymyan, A., and F. Ramos, 2009, A sparse covariance function for exact Gaussian process inference in large datasets.: International Joint Conferences on Artificial Intelligence, 1936–1942.

Melkuymyan, A., and F. Ramos, 2011, Multi-kernel Gaussian processes: International Joint Conferences on Artificial Intelligence, 1408.

Moćkus, J., 1975, On Bayesian methods for seeking the extremum: Optimization Techniques IFIP Technical Conference, Springer, 400–404.

Moćkus, J., 1975, On Bayesian methods of optimization: Towards Global Optimization, 166–181.

Mockus, J., V. Tiesis, and A. Zilinskas, 1978, The application of Bayesian methods for seeking the extremum: Towards global optimization, 2, 2.

Moorkamp, M., B. Heincke, M. Jegen, A. W. Roberts, and R. W. Hobbs, 2011, A framework for 3-D joint inversion of MT, gravity and seismic refraction data: Geophysical Journal International, 184, 477–493.

Mosegaard, K., and M. Sambridge, 2002, Monte Carlo analysis of inverse problems: Inverse problems, 18, R29–R54.

Mukerji, T., P. Avseth, G. Mavko, I. Takahashi, and E. F. González, 2001, Statistical rock physics: Combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization: The Leading Edge, 20, 313–319.

Nagy, D., G. Papp, and J. Benedek, 2000, The gravitational potential and its derivatives for the prism: Journal of Geodesy, 74, 552–560.

Osborne, M. A., S. J. Roberts, A. Rogers, S. D. Ramchurn, and N. R. Jennings, 2008, Towards real-time information processing of sensor network data using computationally efficient multi-output Gaussian processes: 2008 international conference on information processing in sensor networks (IPSN 2008), IEEE, 109–120.

Powell, M. J., 1994, A direct search optimization method that models the objective and constraint functions by linear interpolation, in Advances in optimization and numerical analysis: Springer, 51–67.
Rasmussen, C. E., and C. K. Williams, 2006, Gaussian process for machine learning: MIT press.

Reid, A., O. Simon Timothy, E. V. Bonilla, L. McCalman, T. Rawling, and F. Ramos, 2013, Bayesian joint inversions for the exploration of earth resources.: International Joint Conferences on Artificial Intelligence, 2877–2884.

Robbins, H., 1952, Some aspects of the sequential design of experiments: Bulletin of the American Mathematical Society, 58, 527–535.

Shamsipour, P., D. Marcotte, and M. Chouteau, 2012, 3D stochastic joint inversion of gravity and magnetic data: Journal of Applied Geophysics, 79, 27–37.

Srinivas, N., A. Krause, S. M. Kakade, and M. Seeger, 2009, Gaussian process optimization in the bandit setting: No regret and experimental design: arXiv preprint arXiv:0912.3995.

Zeyen, H., and J. Pous, 1993, 3-D joint inversion of magnetic and gravimetric data with a priori information: Geophysical Journal International, 112, 244–256.