AN INTRODUCTION TO OPTIMAL DATA COLLECTION FOR GEOPHYSICAL MODEL CALIBRATION PROBLEMS

Ruanui Nicholson[1], Henry Alferink[1], Elliot Paton-Simpson[1], Michael Gravatt[1], John P. O'Sullivan[1], Michael J. O'Sullivan[1], Oliver J. Maclaren[1]

[1]Department of Engineering Science, The University of Auckland, Auckland 1142, New Zealand

ruanui.nicholson@aucklanduni.ac.nz

Keywords: *Calibration, estimation, inversion, Bayesian, data collection, optimization, uncertainty quantification, uncertainty reduction.*

ABSTRACT

Computational modelling and calibration can play a vital role in various geophysical settings, including the management of geothermal reservoirs. The calibration process typically involves perturbing model parameters, such as subsurface permeabilities, so that the resulting simulator outputs match field data, usually consisting of noisy surface or subsurface measurements. In an engineering context, however, model calibration is not a once-off process, but rather an on-going process that often requires careful decision-making about whether to collect new measurements and where these should be collected. Here we present a tutorial-style introduction on how to formulate and solve such decision problems, using two archetypal but straightforward geophysical model problems.

The basic calibration problem can be considered a statistical inference problem. In particular, the Bayesian framework for calibration naturally allows for uncertainties in the data, the parameters, and models to be quantified and incorporated into this process. In this framework, the solution to the calibration (and inference problem) is a so-called posterior probability density, which characterises the remaining uncertainty in the parameters after conditioning on the field data. This uncertainty can then help guide the management of geophysical resources, where, intuitively, reducing (posterior) uncertainty enables higher quality decisions to be made.

Reduction in posterior uncertainty is achievable by a) improving measurement precision at existing locations and hence obtaining similar data with lower noise , or b) collecting additional data at new locations. Here we focus on the second option. The cost of collecting data in new areas can be considerable, however, especially when this data requires e.g. drilling a new observation well or constructing and deploying other expensive measurement equipment. Hence, we consider the design problem of how to determine the reward, in terms of uncertainty reduction, of a potential new observation, before collection.

In addition to illustrating the basic principles of optimal data collection and data worth analysis using simple geophysical problems, we also show how the methodology used can be further developed to include, for example, the possibility of spatially heterogeneous costs of acquiring new data.

1. INTRODUCTION

Computational modelling has become an indispensable tool in many geophysical settings, including the management of geothermal reservoirs (e.g., O'Sullivan, Pruess, and Lippmann (2001)). However, the usefulness of any computational model depends on its ability to accurately predict reality. This requirement naturally requires calibration of the model, a process which entails tuning of model parameters, such as subsurface permeabilities, or boundary conditions, so that the model can match field data.

Over at least the last forty years there has been a concerted effort towards incorporating various sources of uncertainty into the calibration process (e.g., Tarantola and Valette (1982)).The ultimate goal of these efforts is to quantify the certainty, or lack thereof, in the model parameters and model predictions. The Bayesian framework (e.g., Tarantola (2005), Kaipio and Somersalo (2006), Stuart (2010), Aster, Borchers, Thurber (2018)) is particularly well suited for handling various sources of uncertainty, including uncertainty in parameters, noise in the measurements, and modelling errors (Kaipio and Somersalo (2007)). Posing the calibration problem in the Bayesian framework results in a problem of statistical inference, where the goal is calculation of the parameter posterior probability distribution (or simply *the posterior*), i.e., the probability distribution of the parameters conditioned on the field data.

In addition to one-off calibration problems, the Bayesian framework facilitates on-going management of, and decision-making concerning, a geophysical resource. Since, loosely speaking, the level of confidence in, or credibility of, a computational model is inversely proportional to the level of uncertainty, it is thus desirable to reduce uncertainties in parameters and related quantities of interest. Such reductions in uncertainty can generally be achieved by either improving the quality of the data, i.e. reducing the amount of noise in existing measurements, or by taking additional measurements. The latter is the focus of the current paper. In many geophysical settings, however, the acquisition of new data is expensive, this is particularly the case in petroleum and geothermal engineering when new collecting new data can necessitate the drilling of a new observation well. Intuitively then, prior to deciding on where or when to measure, it could be highly beneficial to investigate the insights any further data would provide, i.e., how much will any new data reduce the posterior uncertainty. Carrying out such a study can be further complicated when the cost associated with collecting new data is heterogeneous, i.e., collecting different types of new data may have (very) different costs.

Here we show how to formalize the problem of finding an optimal data collection strategy, using simple geophysical problems as illustrations. First, we briefly review how the basic calibration process can be posed in the Bayesian framework

2. CALIBRATION AS AN INVERSE PROBLEM

As is standard, we will assume that the relationship between the field data, $\boldsymbol{d} \in \mathbb{R}^d$, and the parameters, $\boldsymbol{m} \in \mathbb{R}^m$, can be expressed by

$$\boldsymbol{d} = \boldsymbol{g}(\boldsymbol{m}) + \boldsymbol{\epsilon},$$

where $\boldsymbol{g}: \mathbb{R}^m \to \mathbb{R}^d$, represents *the forward model* (or simulator), and $\boldsymbol{\epsilon} \in \mathbb{R}^d$, denotes any sources of additive errors and/or noise in the measurements. The *forward problem* involves solving the forward model equations given parameters; the *calibration* problem is, in contrast, to find the model parameters based on the (noisy) data.

2.1 Inverse problems in the Bayesian Framework

In the Bayesian framework, all unknown quantities are modelled as random variables; the probability distributions associated with these variables are updated by conditioning on new data as it becomes available. This updating procedure is based on the use of Bayes' formula:

$$\boldsymbol{p}(\boldsymbol{m}|\boldsymbol{d}) \propto \boldsymbol{p}(\boldsymbol{d}|\boldsymbol{m})\boldsymbol{p}(\boldsymbol{m}),$$

where $\boldsymbol{p}(\boldsymbol{d}|\boldsymbol{m})$ is termed *the likelihood*, $\boldsymbol{p}(\boldsymbol{m})$ *the prior*, and $\boldsymbol{p}(\boldsymbol{m}|\boldsymbol{d})$ is the posterior. In the case of additive noise, the likelihood inherits the distribution of the noise. On the other hand, the prior distribution encodes any of our beliefs about the parameter, such as smoothness or bounds, before the collection of any data. If both the prior distribution and distribution of the noise are Gaussian, and the noise is independent of the parameters, the posterior has the form

$$\boldsymbol{p}(\boldsymbol{m}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2}((\boldsymbol{g}(\boldsymbol{m}) - \boldsymbol{d})^T \mathbf{C}_\mathrm{d}^{-1}(\boldsymbol{g}(\boldsymbol{m}) - \boldsymbol{d}) + \left(\boldsymbol{m} - \boldsymbol{m}_\mathrm{prior}\right)^T \mathbf{C}_\mathrm{m}^{-1}(\boldsymbol{m} - \boldsymbol{m}_\mathrm{prior}))\right)$$

where $\mathbf{C}_\mathrm{d}$ is the data covariance matrix, i.e., the covariance of the noise, and $\mathbf{C}_\mathrm{m}$ is the prior covariance matrix. Although this may appear to be a simple Gaussian distribution at first sight, the nonlinearity of $\boldsymbol{g}(\boldsymbol{m})$ prevents this from being Gaussian in $\boldsymbol{m}$.

Fully characterizing the full, nonlinear posterior can be extremely time consuming, especially when the forward model involves solving high dimensional partial differential equations (PDEs). Such computationally intensive forward models are common for problems arising in geophysics. A computationally cheaper alternative which is often used in practice is to construct a local Gaussian (in $\boldsymbol{m}$) approximation to the posterior centred at the *maximum a posteriori* (MAP) estimate, $\boldsymbol{m}_\mathrm{MAP}$, which is the point in parameter space which maximises the posterior probability. While this MAP estimate typically requires nonlinear optimization methods for its estimation, subsequent uncertainty quantification can utilise efficient local linear analysis methods. In particular, the Gaussian approximation is constructed by linearising the forward problem around the MAP estimate, i.e.,

$$\boldsymbol{d} \approx \boldsymbol{g}(\boldsymbol{m}_\mathrm{MAP}) + \boldsymbol{G}(\boldsymbol{m} - \boldsymbol{m}_\mathrm{MAP}) + \boldsymbol{\epsilon},$$

where $\boldsymbol{G}$ is the Jacobian matrix of the forward model, $\boldsymbol{G}_{ij} = \frac{\partial \boldsymbol{g}_i}{\partial \boldsymbol{m}_j}$. This results in a Gaussian approximation to the posterior with mean $\boldsymbol{m}_\mathrm{MAP}$, and approximate posterior covariance $\mathbf{C}_\mathrm{p} = \left(\mathbf{G}^T \mathbf{C}_\mathrm{d}^{-1} \mathbf{G} + \mathbf{C}_\mathrm{m}^{-1}\right)^{-1}$ (e.g., Tarantola (2005), Kaipio and Somersalo (2006)). The next section details how this posterior covariance matrix is fundamental to the optimal data collection problem.

3. MINIMISING POSTERIOR UNCERTAINTY

The goal of devising an optimal data collection strategy is to determine where, when, and what to measure so as to minimise the posterior uncertainty, which is encoded in the posterior covariance matrix. In the case of a single uncertain parameter, i.e., $\boldsymbol{m} \in \mathbb{R}$, the posterior covariance reduces to a single number, i.e., $\mathbf{C}_\mathrm{p} = \boldsymbol{\sigma}^2 \in \mathbb{R}$, and the goal of the resulting optimization problem is obvious. On the other hand, in the general setting $\boldsymbol{m} \in \mathbb{R}^m$, implying $\mathbf{C}_\mathrm{p} \in \mathbb{R}^{m \times m}$, and minimizing the posterior uncertainty becomes ambiguous: What does it mean to *minimise a matrix*? This has led to the introduction of several criteria, including A-optimality and D-optimality, which are discussed below.

- *A-optimality:* The A-optimality criteria is minimisation of the average (hence the name A-optimality) marginal posterior variance. That is
$$\min\ \phi_\mathrm{A} = \min\ \frac{\mathrm{trace}(\mathbf{C}_\mathrm{p})}{m}.$$
- *D-optimality:* The D-optimality criteria is minimisation of the determinant (hence the name D-optimality) of the posterior covariance matrix. That is
$$\min\ \phi_\mathrm{D} = \min\ \det(\mathbf{C}_\mathrm{p}).$$

Other optimality metrics include E-optimality, which minimizes the largest eigenvalue of $\mathbf{C}_\mathrm{p}$, and conditions such as the goal of minimising the posterior uncertainty of particular subsets of the parameter (see for example: Finsterle (2015), Alexanderian (2020)).

As well as specifying the optimality criteria, to design an optimal data collection strategy we must also take into account several other problem specific details. The first of these is recognising whether the problem is posed in a discrete or continuous setting. Specifically, are there a set of possible candidate measurement locations, or are we essentially free to measure anywhere? Secondly, the cost of taking measurements is generally spatially (and temporally) heterogenous.

A straightforward way of incorporating the heterogenous cost of taking new measurements into the optimisation problem is to introduce a penalty term. Naturally, the penalty term should penalise costly measurement strategies more than cheaper measurement strategies (see Figure 1 below). The exact form of the penalty term is dependent on whether the problem is posed in the discrete or continuous setting. If the problem is of the discrete form, then we only require a penalty term defined at the candidate locations. On the other hand, posing the problem in the continuous setting necessitates the definition of the penalty term *everywhere.*

In the current study, we consider stationary (time-independent) problems only but note that for time-varying problems the time variable can be treated essentially as another spatial dimension. Furthermore, without loss of generality we consider measurements of only one type, and thus the optimization is for the location of the new measurement, χ. Finally, for simplicity, the forward problems considered in the current paper are already given in linear form, so that

$$\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{G}\boldsymbol{m}.$$

This can be considered as arising from a linear approximation to a nonlinear problem, or arising from an inherently linear problem.

Independently of the specifics, devising an optimal data collection strategy can be formulated mathematically as solving a minimization problem of the form

$$\chi_I := \min_{x \in X} \phi_I + P(x),$$

where X denotes the space of possible new measurement locations (discrete or continuous), $I \in \{A, D\}$ is the optimality criterion, and $P(x) \geq 0$ denotes the heterogenous cost of measuring at each $x \in X$, i.e., a penalty function.

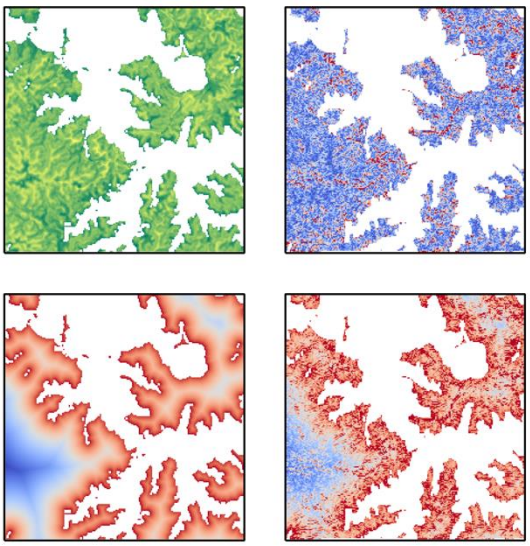

Figure 1: Possible penalty terms for Bay of Islands, New Zealand. Topography of Bay of Islands (top left). A cost term which favours smooth topographies (top right). A cost term which favours being near the water (bottom left), and a combination of the two (bottom right).

In what follows, for ease of communication, we assume the additional sets of measurements are composed of only a single new measurement, the generalization follows trivially. Incorporation of an additional measurement results in an updated posterior covariance matrix. Specifically, we now have

$$\hat{\mathbf{C}}_\mathrm{p} = \left(\hat{\mathbf{G}}^T \hat{\mathbf{C}}_\mathrm{d}^{-1} \hat{\mathbf{G}} + \mathbf{C}_\mathrm{m}^{-1}\right)^{-1}, \quad \hat{\mathbf{G}} = \begin{bmatrix} \mathbf{G} \\ \boldsymbol{g} \end{bmatrix}, \quad \hat{\mathbf{C}}_\mathrm{d}^{-1} = \begin{bmatrix} \mathbf{C}_\mathrm{d}^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma_d^{-2} \end{bmatrix},$$

where $\boldsymbol{g} \in \mathbb{R}^{1 \times m}$ represents the mapping from the parameters, $\boldsymbol{m} \in \mathbb{R}^m$, to the new measurement, and we have assumed the new measurement will also be corrupted by additive noise with standard deviation σ_d. Consequentially, the optimal measurement location will be given by

$$\chi_A = \min_{x \in X} \frac{\mathrm{trace}(\hat{\mathbf{C}}_\mathrm{p})}{m} + P(x),$$

for A-optimality, and

$$\chi_D = \min_{x \in X} \det(\hat{\mathbf{C}}_\mathrm{p}) + P(x),$$

for D-optimality.

In the continuous setting, efficiently solving the optimization problem requires derivative information. That is, we would need to compute at least the gradient,

$$\nabla_x \phi_I + \nabla_x P,$$

and possibly the Hessian,

$$\boldsymbol{H}_{ij} = \frac{\partial \phi_I}{\partial x_i \partial x_j} + \frac{\partial P}{\partial x_i \partial x_j}.$$

Arguably, however, for many geophysical problems, including within the geothermal context, the discrete setting is more applicable. That is, there are often a set of known new measurement candidate locations, denoted here by $\hat{x}_1, \hat{x}_2, \ldots \hat{x}_N$. Then to calculate the optimal measurement location we compute $\phi_I(x_i) + P(x_i)$ for each $i \in \{1,2 \ldots, N\}$, and $I = A$ or $I = D$.

For large-scale problems with more than a few candidate locations, however, recomputing the updated posterior covariance matrix, $\hat{\mathbf{C}}_\mathrm{p}$, N times can easily become prohibitive. To avoid this problem, we first rewrite the updated posterior covariance matrix as

$$\hat{\mathbf{C}}_\mathrm{p} = \left(\mathbf{C}_\mathrm{p}^{-1} + \sigma_d^{-2} \boldsymbol{g}^T \boldsymbol{g}\right)^{-1}$$

Application of the Sherman-Morrison formula then gives

$$\hat{\mathbf{C}}_\mathrm{p} = \mathbf{C}_\mathrm{p} - \frac{\mathbf{C}_\mathrm{p} \boldsymbol{g}^T \boldsymbol{g} \mathbf{C}_\mathrm{p}}{\sigma_d^2 + \boldsymbol{g} \mathbf{C}_\mathrm{p} \boldsymbol{g}^T}.$$

That is, no further matrix inversions are required to update the posterior covariance matrix. Furthermore, the trace, required for ϕ_A, can then be calculated as

$$\mathrm{trace}(\hat{\mathbf{C}}_\mathrm{p}) = \mathrm{trace}(\mathbf{C}_\mathrm{p}) - \lambda \boldsymbol{g} \mathbf{C}_\mathrm{p} \mathbf{C}_\mathrm{p} \boldsymbol{g}^T,$$

where $\lambda = \left(\sigma_d^2 + \boldsymbol{g} \mathbf{C}_\mathrm{p} \boldsymbol{g}^T\right)^{-1}$, and applying the matrix determinant lemma gives

$$\det(\hat{\mathbf{C}}_\mathrm{p}) = \det(\mathbf{C}_\mathrm{p})\left(1 + \lambda \boldsymbol{g} \mathbf{C}_\mathrm{p} \boldsymbol{g}^T\right).$$

4. EXAMPLES

Here we apply the developed methodology to two numerical examples. We only consider linear cases. However, as alluded to previously, the framework considered here relies on reducing the uncertainty in a Gaussian approximation to the posterior, which only relies on the linearization of the forward problem. As such, the treatment of linear problems given here is more broadly applicable than one might expect.

4.1 Example 1: a source reconstruction example

We first consider the problem of determining the time-varying concentration of pollutant being released into a (one dimensional) groundwater aquifer based on well measurements downstream of the concentration levels at a particular time $t = T$. The forward problem which maps the

concentration at the point of release, $\boldsymbol{m} = C_{in}(t)$, to the measurements downstream, at location x, is given by

$$\boldsymbol{G}\boldsymbol{m} = C(x,T) = \int_0^T f(x, T-t)\boldsymbol{m}dt,$$

where $f(x,t) = \frac{x}{2\sqrt{\pi D t^3}} e^{\frac{(x-vt)^2}{4Dt}}$, with the diffusion coefficient, D and flowrate v taken as $D = v = 1$. For more details on the set up see Neupauer, Borchers, and Wilson (2000) and Aster, Borchers, and Thurber (2018). After discretization, the linear forward operator can be written as $\boldsymbol{G}_{ij} = f(x_i, T - t_j)\Delta t$.

We represent our prior assumptions about the time-varying concentration of pollutant being released into stream, $\boldsymbol{m}$, using a prior incorporating smoothness (e.g. due to the expectation of correlations in concentration over time). As such, we take a prior covariance matrix of the form

$$\{\mathbf{C}_{\mathrm{m}}\}_{ij} = \sigma_{\mathrm{m}}^2 e^{\frac{(t_i - t_j)^2}{2\ell^2}} + \gamma \boldsymbol{I},$$

where we take the prior variance as $\sigma_{\mathrm{m}}^2 = 1.5$, the characteristic length scale as $\ell = 10$, and taking $\gamma = 10^{-4}$ to ensure positive definiteness, with $\boldsymbol{I}$ the identity matrix. In Figure 2 we show the prior mean, prior marginal variance, several draws from the prior, along with the truth. We also report the values of $\mathrm{trace}(\mathbf{C}_{\mathrm{m}})/m$ and $-\log(\det(\mathbf{C}_{\mathrm{m}}))$

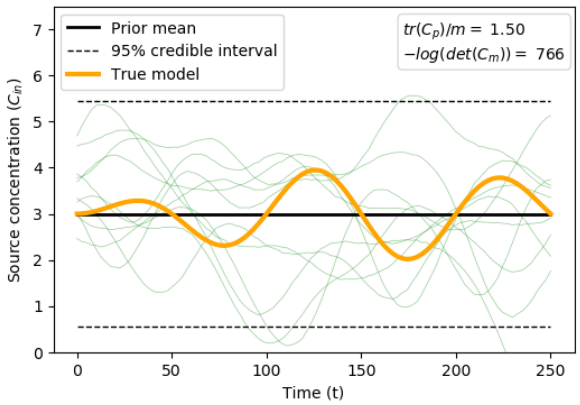

Figure 2: The prior distribution for Example 1. Prior mean, the prior credibility intervals, several samples from the prior showing the smoothness (faint green lines), and the truth.

In terms of the different potential new data, we consider three scenarios where each involve taking four new measurements, as shown in Figure 3. We assume the cost of taking the measurements is the same for each scenario, and thus do away with the penalty term, $P(t)$. In Figure 3 we also show the collected data along with the underlying model.

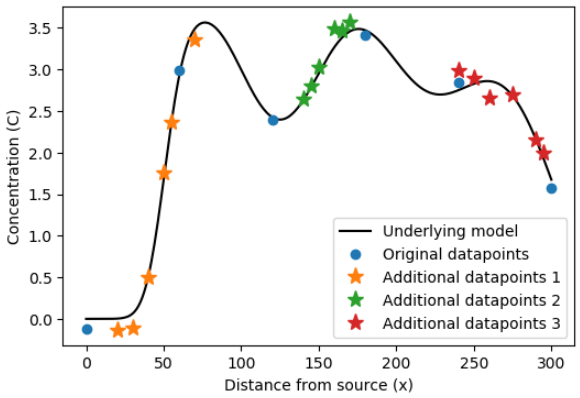

Figure 3: The data for Example 1. The collected data, and each of the proposed new measurement scenarios, as well as the underlying true model output.

In Figure 4 we show the posterior found using only the original field data, and report the values of $\mathrm{trace}(\mathbf{C}_{\mathrm{p}})/m$ and $-\log(\det(\mathbf{C}_{\mathrm{p}}))$. It is clear that, as hoped, the uncertainty in the posterior is significantly reduced compared to the prior uncertainty. This posterior is not of primary interest in this paper, but instead is used as a *reference posterior*, to compare to for the potential posteriors based on the possible new measurements. In Figures 5-7 we show the resulting posterior for new measurement scenarios 1-3, respectively. It is worth pointing out that the resulting posterior uncertainty can be calculated before taking any new measurements in the linear case (Kaipio and Somersalo (2006), Stuart (2010)). The MAP estimates are shown only for completeness. For each of the scenarios we report the values of $\mathrm{trace}(\mathbf{C}_{\mathrm{p}})/m$ and $-\log(\det(\mathbf{C}_{\mathrm{p}}))$

In this case we can see that the second new data collection scenario is the best by both of the measures we use.

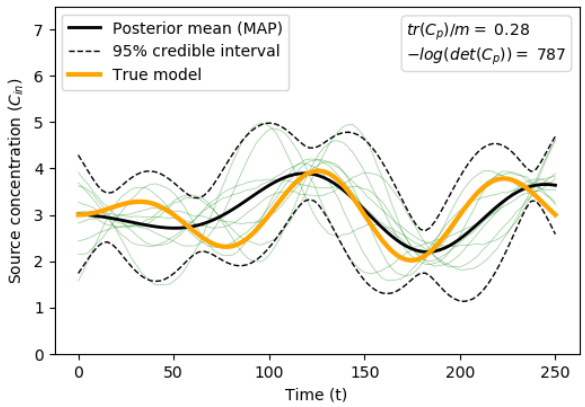

Figure 4: The reference posterior for Example 1. The MAP estimate, credibility intervals, several samples from the posterior (faint green lines), and the truth.

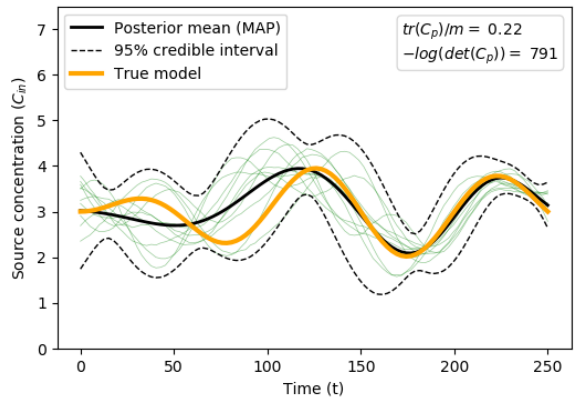

Figure 5: The implied posterior for Scenario 1 of Example 1. The MAP estimate, credibility intervals, several samples from the posterior (faint green lines), and the truth.

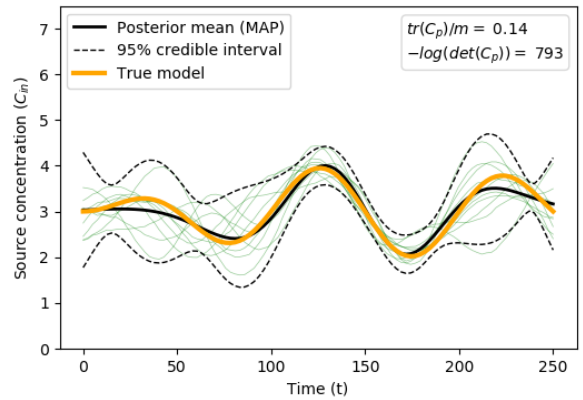

Figure 6: The implied posterior for Scenario 2 of Example 1. The MAP estimate, credibility intervals, several samples from the posterior (faint green lines), and the truth.

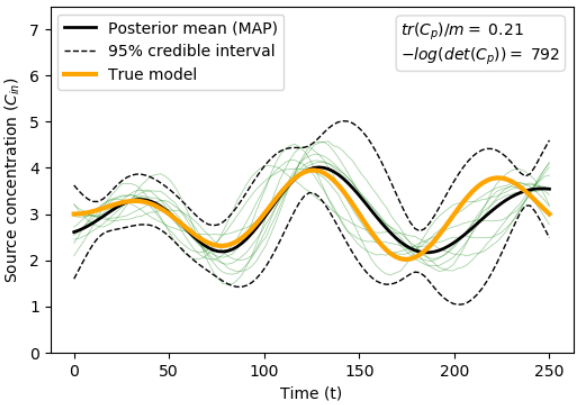

Figure 7: The implied posterior for Scenario 3 of Example 1. The MAP estimate, credibility intervals, several samples from the posterior (faint green lines), and the truth.

4.2 Example 2: a subsurface anomaly detection

The second example we consider arises naturally in the field of geophysics, and serves as an archetypal case of subsurface measurement problems. Specifically, we consider problem of identifying anomalies in the subsurface density, denoted generically as $\boldsymbol{m}$, based on noisy point-wise surface measurements of the local gravitational acceleration, $\boldsymbol{d}$. The forward problem can be derived using Newton's law of gravitation, and can be expressed as

$$\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m} + \boldsymbol{\epsilon},$$

where $\boldsymbol{G}$ can be written, when using a finite element method (FEM) as $\boldsymbol{G} = \boldsymbol{B}\boldsymbol{K}^{-1}$, where $\boldsymbol{K}$ is the so-called stiffness matrix,

$$\boldsymbol{K}_{ij} = \int_{\Omega} \nabla\psi_i^T \nabla\psi_j \, dx\,, \quad i,j = 1,2,\dots,m$$

where Ω is the computational domain, and $\boldsymbol{B}$ is the observation (or restriction) operator, which maps points throughout the domain to only the measurement locations. For more on the gravitational problem, see, for example, Li and Oldenburg (1998).

The prior covariance matrix is taken to be $\mathbf{C}_{\mathrm{m}} = (\boldsymbol{K}^T\boldsymbol{K})^{-1}$, where $\boldsymbol{K}$ is as above, but equipped with Dirichlet boundary conditions. This choice of prior encodes smoothness (though far less smoothness than that of the prior used in the previous example) and is sometimes referred to as a PDE-based prior, due to its link to the discretization of a PDE (see for example: (Kaipio, Kolehmainen, Vauhkonen, et al. (1999), Bardsley (2013)). The true anomalies in density are two spheres of te same size, while the reference measurements consist of 16 equally space noise measurements on the top of the domain, as shown in the top left of Figure 8. Shown at the top right of Figure 8 are the potential new scenarios for data collection, each comprised of 8 colinear points on the top surface. Shown on the same image is the heterogeneous penalty term, $P(x)$. In this case we have manufactured a penalty function which favours taking measurements near existing measurements, and towards the bottom right of the top surface. In the bottom row of Figure 8 we show on the left the MAP estimate, $\boldsymbol{m}_{\mathrm{MAP}}$, along a cross-section, and on the right the marginal posterior standard deviation along the same cross section.

In Figure 9 we show the resulting marginal posterior standard deviation for each of the potential scenarios, for reference we also show the reference marginal posterior standard deviation (top right). For each of the new scenario posteriors we also report the values of $\mathrm{trace}(\mathbf{C}_{\mathrm{m}})/m + P$. Note that for this example we only consider the A-optimality criteria. In this example we see that scenario 2 is the optimal measurement strategy based on the value of $\mathrm{trace}(\mathbf{C}_{\mathrm{m}})/m + P$.

For completeness we also show the resulting MAP estimates (along the same cross-section) in Figure 10. The main contribution from all cases of the potential new measurements is a reduction in the parameter uncertainty between the two spheres (see Figure 9). This in turn results in a clear distinction between the two anomalies in the updated MAP estimates compared to the reference MAP estimate, which is fairly ambiguous as to whether there are one or two anomalies. Finally, it is also evident (see Figure 10) that the MAP estimate found using scenario 2 is the best representation of the true anomalies out of all the MAP estimates computed.

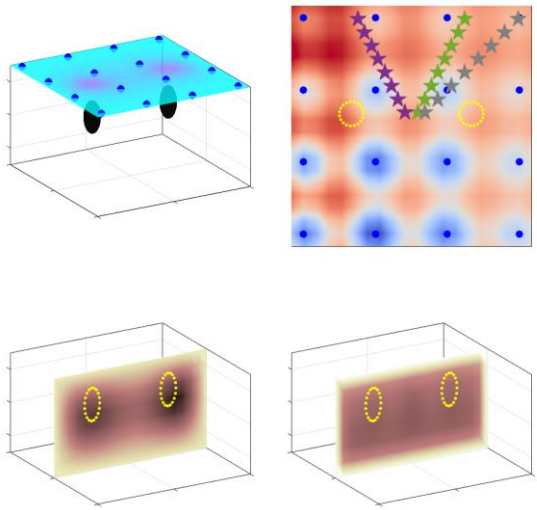

Figure 8: Set up and reference posterior for Example 2. The true spherical density anomalies (black), gravitational force at top surface (violet indicates larger force), and data locations (top left). The penalty function, $P(x)$, (red indicates higher cost) the data locations and each of the proposed new measurement scenarios (Scenario 1: green stars, Scenario 2: purple stars, Scenario 3: grey stars) (top right). Cross section of the MAP estimate (darker brown represents higher density) (bottom left), and the marginal posterior standard deviation (darker brown represents higher uncertainty) (bottom right).

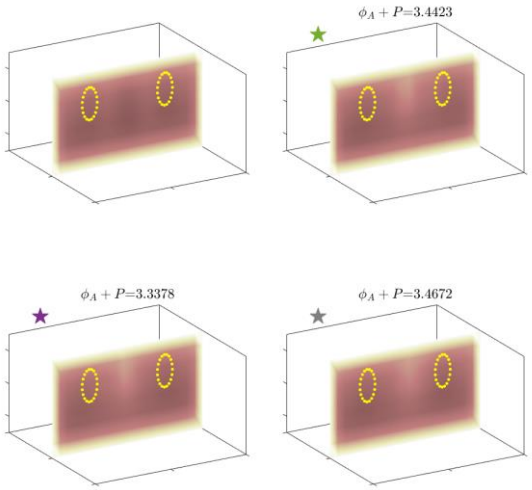

Figure 9: Marginal posterior standard deviations for each scenario for Example 2. The reference (top left), Scenario 1 (top right), Scenario 2 (bottom left), and Scenario 3 (bottom right). In each image the yellow dotted lines indicate the true location of the anomalies (along the cross-section), while darker brown indicates higher uncertainty.

DISCUSSION AND CONCLUSIONS

In this article, we have given a brief introduction to how to build on Bayesian model calibration methodologies to take into account the collection of future data and associated decision problems when managing geophysical resources. We illustrated this methodology on relatively straightforward but representative geophysical model problems: source history reconstruction and subsurface (gravity) anomaly detection. The methodology considered simultaneously attempts to maximise the reduction in uncertainty in the underlying unknown parameters, while minimising the cost associated with carrying out the proposed field measurements. Although here illustrated on relatively simple problems, the approach is immediately applicable to the large-scale setting, given a model linearised around the best fit (MAP) estimate. Given such a linearisation, the optimality criteria discussed can be (very) cheaply evaluated.

We have recently applied this approach to practical, large-scale geothermal problems. Future research will focus on further developing, implementing, and reporting this methodology in the context of such large-scale, real-world geothermal models, where data is particularly sparse and can be very costly to collect.

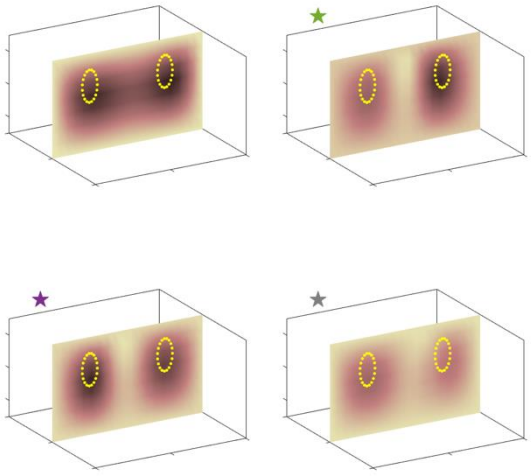

Figure 10: Cross-sections of the MAP estimates for Example 2. The reference (top left), Scenario 1 (top right), Scenario 2 (bottom right), and Scenario 3 (bottom right). In each image the yellow dotted lines indicate the true location of the anomalies (along the cross-section), while darker brown indicates higher uncertainty.

REFERENCES

Alexanderian, A.: Optimal Experimental Design for Bayesian Inverse Problems Governed by PDEs: A Review. *arXiv preprint* arXiv:2005.12998 (2020).

Aster, R. C., Borchers, B., & Thurber, C. H.: *Parameter estimation and inverse problems*. Elsevier. (2018).

Bardsley, J. M.: Gaussian Markov random field priors for inverse problems. *Inverse Problems & Imaging*, 7(2), 397 (2013).

Finsterle, S.: Practical notes on local data-worth analysis. *Water Resources Research*, 51(12), 9904-9924 (2015).

Kaipio, J. P., Kolehmainen, V., Vauhkonen, M., & Somersalo, E.: Inverse problems with structural prior information. *Inverse problems*, 15(3), 713 (1999).

Kaipio, J., & Somersalo, E.: Statistical and computational inverse problems (Vol. 160). *Springer Science & Business Media* (2006).

Kaipio, J., Somersalo, E.: Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of computational and applied mathematics*, *198*(2), 493–504 (2007).

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

Neupauer, R. M., Borchers, B., & Wilson, J. L.: Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. *Water Resources Research*, 36(9), 2469-2475 (2000).

O'Sullivan, M. J., Pruess, K., & Lippmann, M. J.: State of the art of geothermal reservoir simulation. *Geothermics*, 30(4), 395-429 (2001).

Stuart, A. M.: Inverse problems: a Bayesian perspective. Acta numerica, 19, 451 (2010).

Tarantola, A.: *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (2005).

Tarantola, A., & Valette, B.: Inverse problems- quest for information. *Journal of Geophysics*| IF 32.18, 50(1), 159-170 (1982).