INVERSE MODELLING OF GEOTHERMAL RESERVOIRS - A HIERARCHICAL BAYESIAN APPROACH

Oliver J. Maclaren[1], Elvar Bjarkason[1], John O'Sullivan and Michael J. O'Sullivan[1]

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

oliver.maclaren@auckland.ac.nz

Keywords: *Inverse problems, TOUGH2, geothermal modelling, Bayesian inference, MCMC*

ABSTRACT

The proper management of geothermal resources relies heavily on computational and mathematical modelling. Given a sufficiently complete description of the physical processes governing a geothermal field we can, in principle, predict future values of the field's measurable outputs. This is typically called the 'forward problem'. All models, especially complex simulation models, are subject to uncertainty, errors and under-determination, however. This naturally leads to the so-called 'inverse problem' - using those (typically few) quantities that can be reliably measured to determine and quantify the uncertainty in other model parameters and predictions.

While inverse modelling techniques are common in many fields of engineering and physics, geothermal reservoir engineering presents a particularly difficult area of application since it involves nonlinear, coupled multiphase flows, along with possible phase transitions. Here we present preliminary results from our work on developing a unified Bayesian inverse modelling approach for computational models of geothermal fields. To address the multi-scale and coupled nature of geothermal problems we further advocate a so-called 'hierarchical' Bayesian approach; in this approach any assumptions on the separation, composition and reduction of different model components can be naturally expressed using standard conditional probability calculations.

We illustrate how, after formulating a geothermal simulation as a hierarchical Bayesian model, statistical sampling procedures such as Markov Chain Monte Carlo (MCMC) can be used to carry out calibration, prediction and/or uncertainty quantification. The only distinction between these tasks in our framework lies in which particular conditional probability distribution is of interest. We briefly discuss how further extensions, such as the use of reduced-order stochastic process approximations to speed up computation, relate to the hierarchical framework. This commonality further illustrates the benefits of adopting an explicit, unified inverse modelling approach to geothermal reservoir simulation.

1. INTRODUCTION

Computer modelling has a long history in the management of New Zealand's geothermal resources. O'Sullivan *et al.* (2009) describe an interesting history of one such geothermal reservoir simulation model. A number of these models have been developed at the University of Auckland using AUTOUGH2 (Yeh *et al*., 2012), a customised version of the industry-standard TOUGH2 simulation code (Pruess 1999).

Such simulation models have now advanced to the point of becoming a central tool for planning and management. This naturally raises the question of how these models are calibrated and how various potential sources of uncertainty are accounted for. For example, how are, and/or should, uncertainties in the available measurements, the model parameters and the models themselves be considered and incorporated into simulations and associated planning recommendations? How should new information be combined with past information? While there has been increasing interest in better and more systematic automation of these procedures (e.g. Cui *et al*. 2011; Moon *et al.* 2014 and Bjarkason *et al*. 2016), these tasks are still often carried out manually. Not only is this time-consuming, but it is potentially unreliable and even infeasible for humans to carry this out for realistic models.

Many of these challenges can be considered to belong to the technical field of 'inverse problems'. In these problems the goal is to determine information concerning unknown parameters (e.g. model inputs) from measurements carried out on other model quantities (e.g. measured outputs). This terminology arises since the goal is to proceed from information concerning model outputs 'backwards' to information concerning compatible model inputs. The usual input-to-output direction is then termed the 'forward problem'. Tarantola (2006) gives a good overview of the general nature of inverse problems.

Here we consider the so-called Bayesian inference/Bayesian statistics perspective on inverse problems. Kaipio and Somersalo (2006), and Stuart (2010) discuss the Bayesian approach to inverse problems in detail; an essentially equivalent probabilistic framework for inverse problems is discussed by Tarantola (2005). Iglesias and Stuart (2014) give an accessible general overview. A distinguishing aspect of the Bayesian approach is that all sources of uncertainty are represented by probability models. Thus the main challenge is typically in properly formulating a collection of (conditional) probability models. For this and other reasons we further advocate the so-called hierarchical Bayesian framework.

The key feature of this hierarchical approach, which facilitates the management of model complexity, is a systematic decomposition of a full general joint probability model of the overall system of interest into simpler conditional probability components. The hierarchical approach is particularly popular for tackling uncertainty quantification and inverse problems in complex areas with strong connections to geothermal modelling, such as environmental and geophysical science (Berliner 2003; Wikle 2003). Berliner (1996) gave one of the first systematic presentations of this approach and Cressie and Wikle (2011), Wikle (2015) present modern and general perspectives on hierarchical modelling. Berliner (2012) gives a more general overview.

Our primary goal in the present work is to emphasise the hierarchical Bayesian approach as a powerful yet convenient modelling and organisational framework for handling complex models and inverse problems, rather than on details of the particular sampling methods used to evaluate these models. It is interesting to note, however, that the interplay between model formulation and particular numerical (sampling) schemes naturally arises within the hierarchical approach, which we will highlight. Though we do not consider this here, it would be interesting to compare this perspective with recent developments in sampling methods for similar simulation models, such as that presented by Cui *et al*. (2011).

In what follows we first present a simple test inverse problem, based on the work of Bjarkason *et al*. (2016), discuss the challenges it presents, describe an appropriate hierarchical Bayesian model for addressing these challenges, and then present results from Markov Chain Monte Carlo sampling used to 'solve' the inverse problem within our framework. We finish with some discussion about possible future extensions.

2. TEST PROBLEM

In the present work we consider a relatively simple two-dimensional, natural state test problem based on that considered by Bjarkason *et al*. (2016). Natural state problems correspond to pre-production conditions where a potential geothermal reservoir is at equilibrium.

The model geometry consists of a rectangular vertical domain, 1,600 m deep and 2,000 m long in the horizontal direction. There are six rock types each with its own vertical and horizontal permeabilities (see Figure 1). Apart from their permeabilities, all six rock types have the same parameters - a porosity of 10%, a rock grain density of 2,500 kg/m^3, a thermal conductivity of 2.5 W/(m·K) and a rock specific heat of 1.0 kJ/(kg·K).

For the present we limit discussion to a relatively computationally 'cheap' class of models, implemented in AUTOUGH2 with the EOS1 equation of state module (Yeh *et al*., 2012). The numerical grid has 90 elements, 80 of which are cubes (200 m × 200 m × 200 m). The remaining 10 elements are large boundary blocks at the top of the model implementing constant pressure and temperature boundary conditions. The top boundary has a constant pressure of 1 atm and is at a constant temperature of 15°C. At the base of the model there is a constant heat flux of 80 mW/m^2, except at the leftmost bottom boundary block where 1.5 kg/s of 1,200 kJ/kg enthalpy fluid is introduced into the system. The side boundaries are closed.

These models typically only require between 0.1 seconds to 1.0 seconds per forward simulation, making full sampling schemes feasible without further approximation. In later sections we will, however, discuss how the need to speed up simulations and/or introduce and account for approximation errors might be addressed within our framework in future work. This will require formulating approximations to either the forward simulation model or the sampling scheme (or both). Again, we feel that a formal hierarchical approach offers potential benefits for formalising and clarifying the required assumptions.

As indicated, the key parameters of interest in the present two-dimensional test problem consist of a set of 12 permeabilities – 6 horizontal and 6 vertical permeabilities. If we were given these permeabilities, along with an appropriate set of initial conditions (and other default settings), the forward simulation model produces temperatures like those shown in Figure 2. The inverse problem corresponds to the reverse case – we are given measured (or, here, synthetically generated) temperature profiles and the goal is to determine which permeability parameters produced these.

Just as in the case of inverting a simple function of one variable, in general the inverse problem is under-determined/ill-posed – e.g. multiple permeability combinations may produce the same or effectively indistinguishable temperature profiles. Again, just as in the case of a simple function like f(x), perhaps the most general solution approach in principle is (at least for deterministic problems) to simply report all parameter combinations consistent with the observed output y_0 (say) i.e. the inverse set image $f^{-1}\{y_0\} = \{x|f(x) = y_0\}$. This exists even if the point inverse image does not.

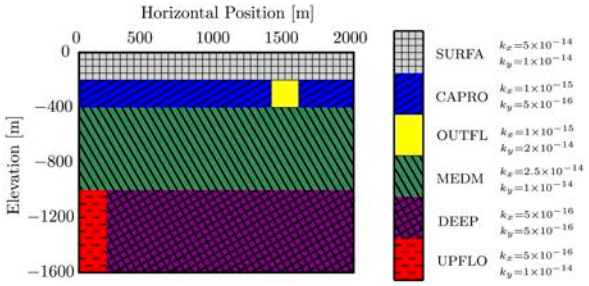

Figure 1: Underlying rock types. Each rock type has, in general, differing vertical and horizontal permeabilities. Modified from (Bjarkason *et al*. 2016).

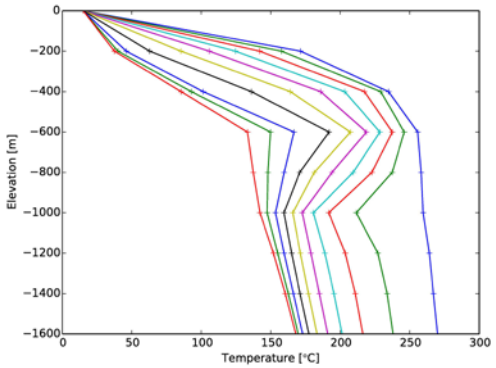

Figure 2: Output (observed) temperature field. Each of the ten lines shows the down-hole temperature recordings at a given (synthetic) 'observation' well.

Two issues (at least) then arise. Firstly, there is typically additional external information available, which further constrains which inputs (here permeabilities) are physically realistic. For example, we require permeabilities to be positive quantities. We may further wish to restrict attention to permeabilities lying within some range of values, based on past experimental/field measurements and/or other expert experience. It is usually desirable and even necessary to incorporate such information in order to obtain more stable, reliable or properly constrained problems. Secondly, all given observations (e.g. temperature measurements) are

subject to uncertainty related to limitations in measurement precision.

In addition to the above, the model itself is of course subject to uncertainty. This may be accounted for in the measurement and/or parameter models or may also be considered as a third distinct source of error. Discretisation and other approximation errors may be considered to fit under this category.

We discuss next how these sources of uncertainty can be naturally formulated within a hierarchical Bayesian framework.

3. HIERARCHICAL BAYESIAN FRAMEWORK

As discussed above, in the Bayesian inference framework all sources of uncertainty, whether physical, epistemic or belonging to some other category, are modelled by probability distributions (Bernardo and Smith 2009; Gelman *et al*. 2014; Tarantola 2005). In particular, this means that all variables entering a problem, including those representing observations and those representing parameters, are taken to be random variables.

While specifying a full general joint probability distribution over all of these variables is typically intractable, this general framework motivates a further useful simplification, resulting in what is typically known as a hierarchical Bayesian model. This is simply the result of making further conditional probability assumptions leading to a factorisation scheme of the three-stage form (Berliner 1996):

data | process variable, observation parameters

process variable | process parameters

process and observation parameters

where '|' indicates 'given' (i.e. probabilistic conditioning). Here the full joint probability distribution is decomposed into three conditional components – in order: a measurement model, a process model and a prior parameter model.

This deceptively simple scheme provides a powerful modelling framework, one that also motivates particular solution or algorithmic approaches that naturally incorporate uncertainty (Berliner 2012 gives a particularly accessible overview and motivation for this framework and related extensions). For more technical presentations of related factorisations see e.g. (Berliner 1996; Tarantola 2005; Cressie and Wikle 2011; Wikle 2015).

In our particular case, we can write our initial probability model as

$$\mathrm{p}\left(\mathbf{y}^{meas}, \mathbf{y}, \mathbf{k} \middle| \boldsymbol{\sigma}_{\boldsymbol{m}}, \mathbf{I}, \boldsymbol{\sigma}_{\boldsymbol{p}}, \mathbf{B}\right) = \mathrm{p}(\mathbf{y}^{meas}|\mathbf{y}, \boldsymbol{\sigma}_{\boldsymbol{m}})\mathrm{p}(\mathbf{y}|\mathbf{k}, \mathbf{I})\mathrm{p}\left(\mathbf{k}\middle|\boldsymbol{\sigma}_{\boldsymbol{p}}, \mathbf{B}\right), \quad (1)$$

where p(x) denotes the probability of x, and the three right-hand side distributions correspond to the measurement, process and parameter distributions introduced above. We discuss the full definitions and interpretations of each of the quantities appearing in the probability distributions above in the context of their respective model components next. We will assume that these quantities can be represented as finite-dimensional random vectors defined with respect to appropriate measurement or simulation grid scales.

3.1 Measurement model

The measurement model in the above is written as

$$\mathrm{p}(\mathbf{y}^{meas}|\mathbf{y}, \boldsymbol{\sigma}_{\boldsymbol{m}}). \quad (2)$$

Here we use $\mathbf{y}^{meas}$ to denote the output vector that would 'actually' be measured by a measuring device, with precision determined by the quantity (here standard deviation) $\boldsymbol{\sigma}_{\boldsymbol{m}}$, when the 'true' underlying vector produced by the process (model) is **y.** In practice, this often incorporates aspects of model (structural) uncertainty rather than 'pure' (random) measurement model uncertainty. That is, we effectively choose to only 'measure' or match the model output to measurements to a reduced tolerance since we know that our model is only an approximation to reality. This is particularly true in the case of black-box simulation models in which it can be difficult to account for structural uncertainty directly.

Here we adopt the following specific functional form corresponding to Normal independent and identically distributed (i.i.d.) errors

$$\mathrm{p}(\mathbf{y}^{meas}|\mathbf{y}, \boldsymbol{\sigma}_{\boldsymbol{m}}) = \prod_{\mathrm{i}=1}^{\mathrm{N}} \frac{1}{\sigma_{m,\mathrm{i}}\sqrt{2\pi}} \exp\left(-\frac{(\mathrm{y}_{\mathrm{i}}^{meas} - \mathrm{y}_{\mathrm{i}})^2}{2\sigma_{m,\mathrm{i}}^2}\right) \quad (3)$$

where N is the number of observations.

We discuss the choice of standard deviations in the 'Sampling and Computational Methods' section below.

3.2 Process model

In principle, the process model is written probabilistically as

$$\mathrm{p}(\mathbf{y}|\mathbf{k}, \mathbf{I}). \quad (4)$$

This indicates that the connection between the output **y**, the parameters of interest **k** (permeabilities) and other 'background' parameters **I** that are not of direct interest (e.g. initial conditions and other settings taken as fixed for the moment), is not necessarily deterministic. This is particularly true when we wish to account for model structural uncertainty, e.g. by modelling structural error as a random process such as a Gaussian process (see e.g. Berliner 2012, Stuart 2012, Tarantola 2005). As discussed above, however, model uncertainty is often pragmatically incorporated into the measurement model, especially when the model is a black-box simulation code with limited access to internal workings. We will adopt this pragmatic approach in the present work but note that most of the promising future directions for improving both modelling and solution methods likely centre on considering this component more carefully. We return to this topic later.

Given the present approach, we take the simulation code to provide a deterministic function for **y** given input permeabilities **k** and other background simulation settings, such as initial conditions, collectively denoted by **I.** Formally, this means that we can write the process model distribution as a Dirac delta function

$$\mathrm{p}(\mathbf{y}|\mathbf{k},\mathbf{I}) = \delta(\mathbf{y} - \mathbf{f}(\mathbf{k},\mathbf{I})), \quad (5)$$

which is centred at the deterministic functional relation $\mathbf{f}(\mathbf{k},\mathbf{I})$. This represents the simulator considered as a black box function relating output $\mathbf{y}$ with inputs $\mathbf{k},\mathbf{I}$. In practice this means that we can replace all instances of $\mathbf{y}$ with $\mathbf{f}(\mathbf{k},\mathbf{I})$.

3.3 Parameter Model

The input parameter distribution of this model

$$\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B}\big) \quad (6)$$

is usually called the prior or parameter prior. Similarly to the measurement model, this involves a probability of a model random variable – here the permeability $\mathbf{k}$ – given a 'precision' measure, here a standard deviation $\boldsymbol{\sigma}_{\boldsymbol{p}}$ and additional background assumptions or parameters represented by $\mathbf{B}$.

The prior distribution (6) represents additional constraints or information about parameters of interest that are not directly contained in the model or data. Minimal prior information might consist of lower and upper bounds and a corresponding uniform distribution. More generally, desired 'regularising' information, designed to penalise unrealistic or undesirable solutions or to enforce e.g. model parsimony is incorporated via this distribution.

Since permeability is a positive quantity, and naturally varies over orders of magnitude, we will use a log-normal prior. In this case the log of the M permeabilities is assumed to be Normal i.i.d. centred about suitable reference values giving, expressed for convenience on the log scale:

$$\mathrm{p}\big(\log(\mathbf{k})\,\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B}\big) = \prod_{\mathrm{i}=1}^{\mathrm{M}} \frac{1}{\sigma_{p,\mathrm{i}}\sqrt{2\pi}} \exp\left(-\frac{\left[\log(\mathrm{k_i}) - \log\left(\mathrm{k_i}^{ref}\right)\right]^2}{2\sigma_{p,\mathrm{i}}^2}\right) \quad (7)$$

Mathematically, this may be considered as imposing a quadratic penalty on solution complexity, i.e. a form of Tikhonov regularisation (Aster *et al*., 2005; Kaipio and Somersalo, 2006). We discuss the choice of reference values and standard deviations in a later section on 'Sampling and Computational Methods'

3.4 Inference and Bayes' Theorem

Our goal is to 'update our information' concerning the unknown permeabilities when given new observed temperature data. This means we wish to update from our prior parameter distribution

$$\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B}\big), \quad (8)$$

discussed previously, to the so-called posterior parameter distribution, here

$$\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B},\mathbf{y}^{meas} = \mathbf{y}^{obs}\big). \quad (9)$$

The difference between the two represents the addition of the information that the measurable model output $\mathbf{y}^{meas}$equals the observed (given) temperature data $\mathbf{y}^{obs}$. Specifically, this connection between the posterior and prior parameter distributions is mediated by the so-called likelihood function $\mathcal{L}(\mathbf{k})$; this connection is formalised by Bayes' theorem written in the form

$$\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B},\mathbf{y}^{meas} = \mathbf{y}^{obs}\big) \propto \mathcal{L}(\mathbf{k})\ \mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B}\big). \quad (10)$$

Here the likelihood function

$$\mathcal{L}(\mathbf{k}) := \mathrm{p}\big(\mathbf{y}^{meas} = \mathbf{y}^{obs}\big|\mathbf{k},\boldsymbol{\sigma}_{\boldsymbol{m}},\mathbf{I}\big), \quad (11)$$

is equal to

$$\int \mathrm{p}\big(\mathbf{y}^{meas} = \mathbf{y}^{obs}\big|\mathbf{y},\boldsymbol{\sigma}_{\boldsymbol{m}}\big)\mathrm{p}(\mathbf{y}|\mathbf{k},\mathbf{I})\mathrm{d}\mathbf{y}, \quad (12)$$

which reduces to

$$\mathrm{p}\big(\mathbf{y}^{meas} = \mathbf{y}^{obs}\big|\mathbf{f}(\mathbf{k},\mathbf{I}),\boldsymbol{\sigma}_{\boldsymbol{m}}\big) \quad (13)$$

since we take the model simulator to be deterministic. The use of a proportionality relation in our form of Bayes' theorem follows from the need to compute a normalising factor (sometimes called the 'evidence') for the posterior distribution; however this factor is typically computed as a by-product of the sampling scheme so the factor is omitted here for convenience. Similarly, a proportionality term is sometimes included in the definition of the likelihood.

The goal then is to produce samples from the posterior parameter distribution for permeabilities given temperature observations. This will be accomplished in this study using Markov Chain Monte Carlo (MCMC) sampling, as discussed in the 'Sampling and Computational Methods' section below.

3.5 Predictive distributions

We note here that for each of the (prior and posterior) 'input' parameter distributions defined above there is a corresponding induced 'output' probability distribution. That is, there is a so-called prior predictive distribution defined by

$$\mathrm{p}(\mathbf{y}^{pred}) := \int \mathrm{p}(\mathbf{y}^{pred}|\mathrm{f}(\mathbf{k},\mathbf{I}),\boldsymbol{\sigma}_{\boldsymbol{m}})\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B}\big)\mathrm{d}\mathbf{k}, \quad (14)$$

and a posterior predictive distribution defined by

$$\mathrm{p}(\mathbf{y}^{pred}|\mathbf{y}^{obs}) := \int \mathrm{p}\big(\mathbf{y}^{pred}\big|\mathrm{f}(\mathbf{k},\mathbf{I}),\boldsymbol{\sigma}_{\boldsymbol{m}}\big)\mathrm{p}\big(\mathbf{k}\big|\boldsymbol{\sigma}_{\boldsymbol{p}},\mathbf{B},\mathbf{y}^{meas} = \mathbf{y}^{obs}\big)\mathrm{d}\mathbf{k}. \quad (15)$$

Each of these can be interpreted as the (prior/posterior) distribution over model outputs that results from carrying out forward simulations using a collection of input parameters sampled from a given (prior/posterior) parameter distribution. These are not only useful for making future predictions, but also for checking model fit/model adequacy. While perhaps obvious, this is important since if a model doesn't fit the data well then any parameter estimates are suspect regardless of any apparent precision. Such checks can be carried out by re-simulating data under the appropriate predictive distribution as described above. This effectively defines 'replicated data', under the conditions defined by the chosen parameter distribution, which are then compared to the given measurements (this is discussed in detail by Gelman (2003) and Gelman *et al*. (2013). We will only briefly consider such checks in the present work, but their importance should not be underestimated.

Note also that forward predictions of just the underlying latent ('noise-free') model component **y**, which here is only uncertain to the extent that the parameters are uncertain, can be presented rather than $\mathbf{y}^{pred}$, which contains both measurement and parameter uncertainty. This is useful for checking that the process model component is behaving as desired. This checking proceeds in the same manner as above, i.e. by drawing a series of samples $\mathbf{k}^i$ from the prior or posterior parameter distribution and computing $\mathbf{y}^i = \mathbf{f}(\mathbf{k}^i, \mathbf{I})$ for each parameter sample. This is the main checking procedure that we will consider in the present work.

4. SAMPLING AND COMPUTATIONAL METHODS

There are at least two related but distinct aspects to the (hierarchical) Bayesian approach. Firstly, it represents a framework for model formulation and the representation of uncertainty. We have outlined this aspect above. Secondly, the probabilistic nature of the framework means that full solutions are almost always in the form of probability distributions, requiring computational sampling methods to evaluate in general.

These two aspects are related – for example approximations employed in a sampling scheme can be directly incorporated into the model structure or vice-versa. They are also, however, distinct – for example, one may seek a 'best single' solution (e.g. MAP – maximum *a posteriori* probability solution) as a point estimate of the full distribution, using optimisation methods rather than sampling methods (see e.g. Aster *et al*., 2005; Tarantola, 2005; Kaipio and Somersalo, 2006). In these cases there may still be a conceptual benefit in formulating the full model in a hierarchical manner, while using potentially more efficient algorithms to obtain fast solutions consisting of less than a full probability distribution.

With this in mind we used the following model estimation procedures to evaluate the output from our model formulation above. These may of course be improved on from both a modelling and an algorithmic perspective, but it can be helpful to distinguish which of the two is of most interest.

4.1 MCMC Sampling For Determining Permeabilities

In the present study the primary parameters of interest are the twelve rock permeabilities contained in the vector **k**. Given the so-called hyper-parameters and other background information, contained in the vectors $\boldsymbol{\sigma}_m, \mathbf{I}, \boldsymbol{\sigma}_p, \mathbf{B}$, we used Markov Chain Monte Carlo sampling (MCMC) to generate samples from the posterior distribution for **k**. Introductions to MCMC methods for inverse problems can be found in, for example, (Aster *et al*., 2005; Tarantola, 2005; Kaipio and Somersalo, 2006).

While a number of innovative sampling schemes have been proposed for similar inverse problems, such as the adaptive scheme described by Cui et al. (2011), in the present case we used a relatively generic method of sampling. This was possible due to the relatively simple test problem considered. In particular, we used the Python package 'emcee' (described in Foreman-Mackey *et al*. (2012)). This package implements an 'affine invariant ensemble sampler' and has the benefit of being easy to implement for arbitrary user-defined models. It also allows for easy communication with the PyTOUGH Python interface (Croucher 2011) to AUTOUGH2 (Yeh et al. 2012) ,which is the University of Auckland's version of the industry-standard TOUGH2 simulation code (Pruess 1999).

We note, however, that some concerns have been raised about the performance in high dimensions of the particular sampling scheme used by 'emcee' (Huijser et al. 2015). With this in mind, future work may require alternative sampling methods to be adopted.

4.2 Parameter Paths And Optimisation For Choosing Hyper-parameters

In principle, the hyper-parameters, such as standard deviations, can be estimated from the data as part of a fully specified Bayesian model with hyper-priors over these parameters. For the present work we adopted a simpler starting point. Firstly, we chose a 'typical' value of $\log_{10}(k^{ref})$ = -15 to centre candidate solutions about. This is consistent with known physical values and those chosen by Bjarkason *et al*. (2016).

Next, it is well known that under typical modelling assumptions the ratio of the standard deviations of the measurement and prior models is directly related to the so-called regularisation parameter of classical inverse problem methods (Aster *et al*. 2005). With this in mind, we first set the measurement standard deviations to be between 10-20% of the observation data (set to 30 °C in the simulations shown here). We then determined an appropriate scale for the prior by first finding an approximate MAP (maximum *a posteriori* probability solution) – i.e. the mode of the posterior parameter distribution (9) - via optimisation for a grid of prior standard deviations, then choosing a point on the 'parameter path' or 'trade-off' curve (the log probability of the MAP solution vs the prior standard deviation) where the effect of the choice of prior standard deviation begins to tail off. Thus this can be considered a method of choosing a weak, but model-parsimony-promoting form of regularisation (see e.g. Aster *et al.* 2005, Hastie *et al*. 2015). To implement this we used the standard Nelder-Mead routine that is part of the Python SciPy (www.scipy.org/) optimisation toolbox.

The resulting standard deviation for the log permeability (base 10) was 1.5, which corresponds well to a reasonable *a priori* range of permeabilities. For example plus or minus two standard deviations about our reference value gives a range for $\log_{10}(k^{candidate})$ of [-12,-18]. This corresponds to a sufficiently broad yet physically constrained range. As noted, however, future work would be to incorporate this choice of 'regularisation scale' into the hierarchical model itself as a parameter to be estimated or informed by the available data in a more probabilistic manner. The present approach can be considered a convenient pragmatic method but one which will lead to underestimation of the uncertainty that would be incorporated into a more fully-Bayesian approach. See the Discussion and Future Directions section below for more on this.

5. RESULTS

Using MCMC, results from 150,000 samples (an ensemble of 300 'walkers' taking 500 samples each),were recorded after an initial 30,000 'burn-in' samples were first discarded, are shown in Figures 3 and 4. The total runtime for the combined 180,000 samples was just under 7 hours, where in the present rather simple problem each model evaluation took only approximately 0.1 seconds. For these simulations we used a relatively coarse grid of only 90

blocks. Here the given or 'observed' temperature data correspond to synthetic data generated under particular 'true' permeability parameter values. The aim was to recover the permeabilities given the temperature observations.

Figure 3 shows posterior predictive model checking, where the model was re-run for a sequence of draws for the posterior permeability probability distribution and compared to the data to which the model was fit. For clarity, only the latent (measurement noise-free) model simulations are shown. Figure 4 shows the inferred posterior probability distributions for underlying permeabilities, visualised in terms of univariate and bivariate marginal distributions. The distributions can be seen to cover the 'true' permeabilities (used to generate the data) well.

For clarity, Figure 4 is presented on the last page, after the references.

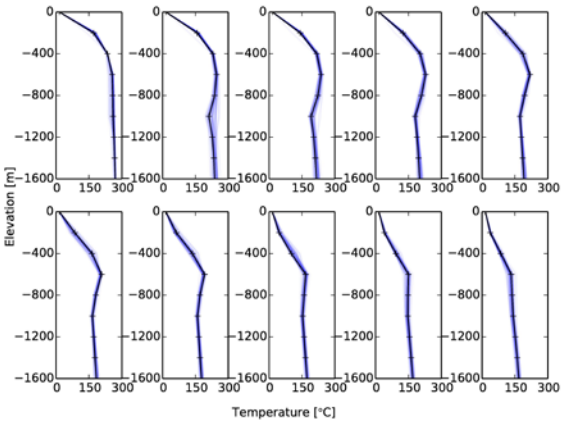

Figure 3: Predictive checks of temperature fit. These 10 profiles correspond to the 10 shown in Figure 2. 60 samples from the posterior predictive distribution for the latent predictions are shown in blue and the 'true' temperatures in black.

6. DISCUSSION AND FUTURE DIRECTIONS

One of the most appealing aspects of the Bayesian approach and, we would argue, the hierarchical Bayesian approach in particular, is that it provides a unified language and framework for formulating and thinking about tasks such as prediction, inference, calibration, approximation error, measurement error, uncertainty quantification, regularisation and sampling (see Berliner (2012), and Iglesias and Stuart (2014) for overviews). These are often treated as distinct tasks but need not be. Within the hierarchical Bayesian framework, each of these may be thought of as related to a particular conditional distribution contained in the overall full joint distribution. We have illustrated some ways in which this perspective works, but there are a number of aspects that we have barely touched on.

An immediate example is the choice of regularisation parameters and/or hyper-parameters. These are often set to a single value on the basis of (frequently ad-hoc) trade-off curves produced by a series of optimisations. This was the approach adopted in this work. It is possible, however, to incorporate this additional uncertainty concerning hyper-parameter/regularisation parameter choice within the hierarchical model itself and thus estimate marginal and jointly varying distributions over these parameters as part of the sampling scheme (Gelman *et al*. (2013) provide a general statistical viewpoint, Hastie *et al.* (2015) give a nice comparison of optimisation and Bayesian hyper-parameter selection in the context of seeking 'sparse' reduced-dimension solutions). This would provide a more 'formal' and automated way of choosing hyper-parameters while also providing more than single point estimates.

In principle this is straightforward; however, it does add another dimension requiring sampling, which increases computational cost. Further case studies focusing on geothermal models would be useful to help address the feasibility of this approach.

Similarly, this perspective highlights the natural relations between model/structural errors, measurement errors and sampling schemes. For example, while it may be desirable to sample from the full distribution determined by a fine-scale simulation model this is typically infeasible. As noted above, here we have focused on a relatively computational cheap simulation model. One alternative, or complement, to focusing on improving sampling algorithms targeted at directly at large and expensive models (as in e.g. Cui *et al.* 2011) is to more carefully consider the relation between these 'fine-scale' models and 'coarser' or 'approximate' (and hence cheaper) versions of these. This generally involves careful consideration of the nature of model structure error and a typical tool is to use stochastic process approximations of this (Berliner, 2012; Iglesias and Stuart, 2014; Kaipio and Somersalo, 2006; Stuart 2010).

Related to models of approximation error are topics such as 'surrogate' or 'emulator' models (Berliner 2012, Vidal 2016), 'Synthetic Likelihoods' (Wood 2010) and 'Approximate Bayesian Computation' (ABC) (Marin et al. 2012). While a full discussion of these topics is beyond the present scope, we note that a number of authors discuss how these various procedures can also be naturally understood and organised using the perspective of hierarchical Bayesian modelling (see e.g. Berliner (1996), Berliner (2012), Iglesias and Stuart (2014), Kaipio and Somersalo (2006), Stuart (2010), and Wilkinson (2013) for fuller discussions).

A natural next step would be to consider to what extent we can use the hierarchical perspective to identify and utilise coarser, but still physically-based, models to approximate fine-scale models, taking into account this approximation error probabilistically. By formalising and identifying when these approximations work, and under what conditions, we can better allocate computational resources to those models that unavoidably require full-scale simulations.

7. CONCLUSIONS

We have discussed the nature of inverse problems in geothermal reservoir modelling and focused on a simple test problem. A hierarchical Bayesian framework was outlined and initial results, obtained using Markov Chain Monte Carlo sampling, were presented. We were able to recover the 'true' parameters used to generate 'observed' synthetic data. Both parameter and predictive distributions were illustrated, and shown to give a good match to their target values. We briefly outlined how our hierarchical framework might be used to tackle more computationally-intensive problems, as well as help resolve difficulties of hyper-parameter/regularisation parameter choices.

Throughout our presentation, we emphasised the unified and intuitive nature of the hierarchical Bayesian approach. In particular, we have described how most questions and challenges can - in principle - be framed in terms of appropriate conditional probability distributions. Furthermore, we have emphasised that this frequently leads directly to a practical interplay between modelling and solution procedures.

ACKNOWLEDGEMENTS

The authors appreciate the contribution of the NZ Ministry of Business, Innovation and Employment for funding parts of this work through the grant: C05X1306, "Geothermal Supermodels". OJM would like to thank Emily Clearwater for help with getting started using AUTOUGH2 and PyTOUGH.

REFERENCES

Aster, R. C., Borchers, B., & Thurber, C. Parameter estimation and inverse problems. (2005). Elsevier Academic Press: Burlington.

Berliner, L. M. (1996). "Hierarchical Bayesian Time Series Models." In Maximum Entropy and Bayesian Methods, 15–22. Fundamental Theories of Physics. Springer Netherlands.

Berliner, L. M. (2003). "Physical-Statistical Modeling in Geophysics." *J. Geophys. Res.* 108 (D24).

Berliner, L. M. (2012). Statistical Approaches to Combining Models and Observations. *SIAM News*, *45*(7).

Bernardo, José M, and Adrian F M Smith. (2009). *Bayesian Theory*. Vol. 405. John Wiley & Sons.

Bjarkason, E. K., O'Sullivan, M. J., O'Sullivan, J. P., & Yeh, A. (2016). Accelerating Calibration of Natural State Geothermal Models. *Proc. 41st Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford California.

Cressie, Noel, and Christopher K Wikle. 2011. *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Croucher, A. (2011). PyTOUGH: a Python scripting library for automating TOUGH2 simulations. In *Proc. the New Zealand Geothermal Workshop* (Vol. 21, p. e23).

Cui, T., Fox, C., & O'Sullivan, M. J. (2011). Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resources Research*, *47*(10).

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, *125*(925), 306.

Foreman-Mackey, D., Vousden, W., Price-Whelan, A., et al. (2016), corner.py: corner.py v1.0.2, doi:10.5281/zenodo.45906

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. (2013). Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Gelman, A. (2003). A Bayesian Formulation of Exploratory Data Analysis and Goodness of fit Testing. *International Statistical Review*, *71*(2), 369-382.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC Press.

Huijser, D., Goodman, J., & Brewer, B. J. (2015). Properties of the Affine Invariant Ensemble Sampler in high dimensions. *arXiv preprint arXiv:1509.02230*.

Iglesias, M, and A. M. Stuart. 2014. "Inverse Problems and Uncertainty Quantification." In *SIAM News*. Vol. July/August. SIAM.

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

Marin, J. M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, *22*(6), 1167-1180.

Moon, H., Clearwater, J., Franz, P., Wallis, I., & Azwar, L. (2014). Sensitivity analysis, parameter estimation and uncertainty propagation in a numerical model of the Ngatamariki geothermal field, New Zealand. *Proc. 39th Workshop on Geothermal Reservoir Engineering.* Stanford University, Stanford, Ca.

O'Sullivan, M. J., Yeh, A., & Mannington, W. I. (2009). A history of numerical modelling of the Wairakei geothermal field. *Geothermics*, *38*(1), 155-168.

Pruess, K., Oldenburg, C. M., & Moridis, G. J. (1999). TOUGH2 user's guide version 2. *Lawrence Berkeley National Laboratory*.

Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta Numerica*, *19*, 451-559.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, *2*(8), 49

Vidal, A., & Archer, R. (2016). Calibration of a Geothermal Reservoir Model using a Global Method based on Surrogate Modeling.

Wikle, C. K. (2003). "Hierarchical Models in Environmental Science." *Int. Stat. Rev.* 71 (2). International Statistical Institute (ISI): 181–99.

Wikle, C.K. (2015). "Modern Perspectives on Statistics for Spatio-Temporal Data." *WIREs Comput Stat* 7 (1). John Wiley & Sons, Inc.: 86–98.

Wilkinson, R. D. (2013). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, *12*(2), 129-141.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, *466*(7310), 1102-1104.

Yeh, A., Croucher, A. E., & O'Sullivan, M. J. (2012). Recent developments in the AUTOUGH2 simulator. In *Proc, TOUGH Symposium, Lawrence Berkeley National Laboratory, Berkeley*.

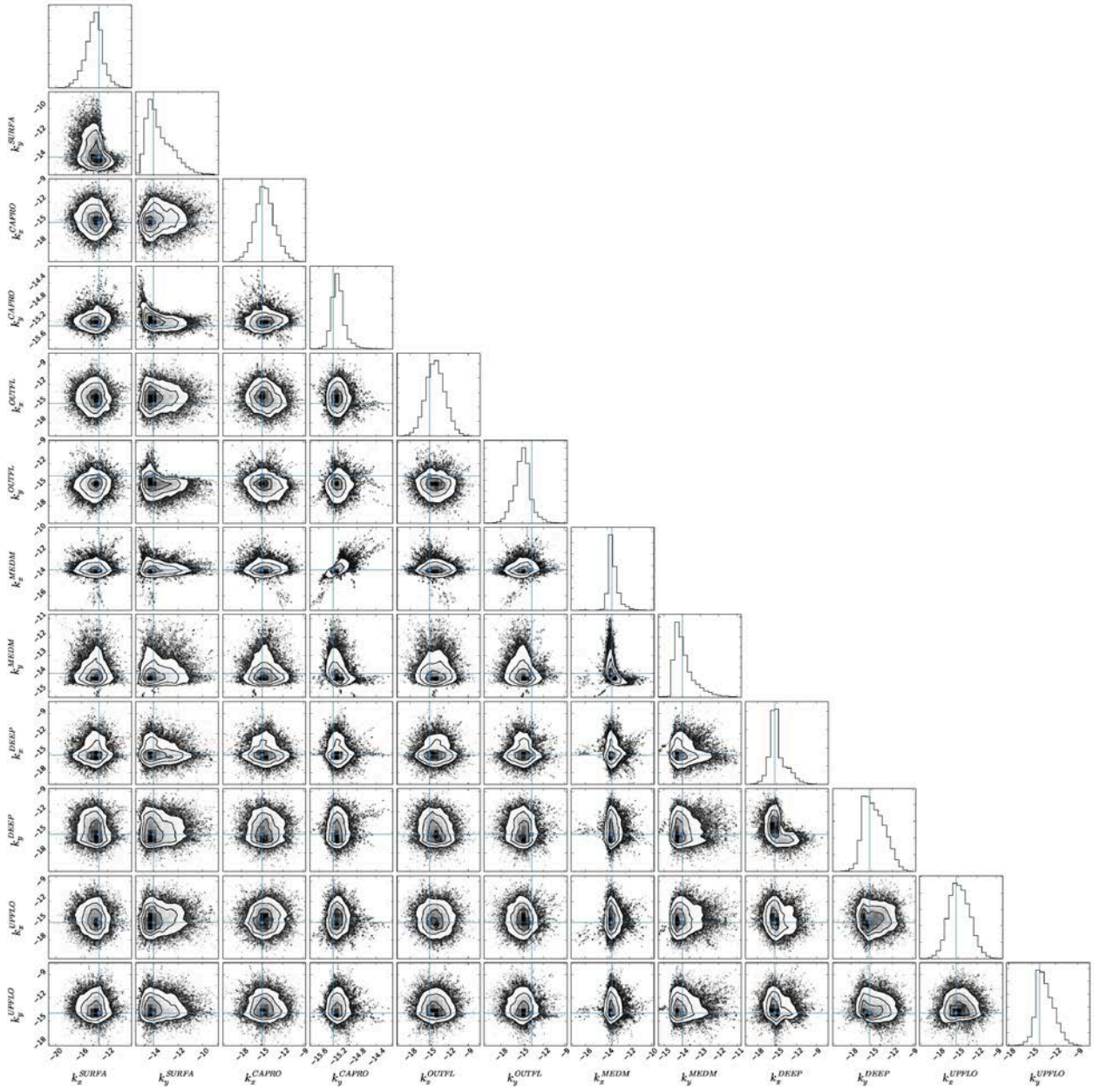

Figure 4: Inferred Permeabilities. The blue vertical lines represent the 'true' target values used to generate the temperature data used for fitting and the histograms represent (marginal) posterior probability distributions for the inferred parameters. The upper diagonal represents the marginal distributions for each permeability parameter when averaging over all other permeabilities. The plots below the diagonal show bivariate marginal distributions illustrating pairwise associations after averaging over all other permeabilities. These visualisations are a way of understanding the full joint posterior distribution which is twelve-dimensional in full generality. The permeabilities are all displayed on a $\log_{10}$ scale. This plot was created using the package 'corner.py' described in (Foreman-Mackey et al., 2016).