PROCEEDINGS, 48th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 6-8, 2023
SGP-TR-224

# Geologically Consistent Prior Parameter Distributions for Uncertainty Quantification of Geothermal Reservoirs

Alex de Beer, Michael J. Gravatt, Theo Renaud, Ruanui Nicholson, Oliver J. Maclaren, Ken Dekkers, John P. O'Sullivan, Andrew Power, Joris Popineau and Michael J. O'Sullivan

Department of Engineering Science, University of Auckland, 70 Symonds Street, Grafton, Auckland 1010, New Zealand

michael.gravatt@auckland.ac.nz

**Keywords:** Reservoir modelling, geology, Bayesian inference, uncertainty quantification.

**ABSTRACT**

Reservoir modeling is a vital tool in the sustainable management of geothermal reservoirs. A key component of the modeling process is model calibration, in which parameters, including permeabilities and mass upflows, are adjusted until there is an adequate match between the model outputs and collected data.

The Bayesian approach to model calibration is gaining increasing use in the context of reservoir modeling. A key element of the Bayesian approach is the use of a prior distribution, which acts as a mathematical representation of expert knowledge on the likely values of the model parameters prior to data being collected. Once data is collected, it is combined with the prior to form the posterior probability distribution, which defines the solution to the calibration problem.

Characterizing the prior is a subjective process. Common practice in geothermal modelling is to use a log-normal distribution to represent the permeabilities associated with each rock type in each formation within the model. This approach, however, disregards geological principles which dictate what the relative values of the permeabilities of different rock types within a given formation can be. For instance, the permeability of a fault rock type in the direction across the fault may not exceed the permeability of the surrounding formation; a prior constructed in the traditional, naïve manner, however, assigns a non-zero probability to this situation.

In this paper, we demonstrate how a prior that adheres to simple geological principles can be characterized. We then perform several comparisons, by running a natural-state model with parameters sampled from both a geologically consistent prior and a prior constructed in a naïve manner, to determine whether adhering to geological principles when characterizing the prior improves the quality of the model outputs. These tests include quantitative analysis of the agreement between reservoir simulation results and geophysical data such as inferred alteration and downhole temperature profiles, as well as a qualitative analysis of the shapes of the convective plumes produced.

Our analysis suggests that there are differences in the temperature profiles produced by running the model with parameter sets sampled from each type of prior. It is difficult, however, to identify the prior that produces the more realistic set of modeled temperatures based on these comparisons. By contrast, our analysis of the convective plumes produced suggests that the plumes produced using parameters sampled from a geologically consistent prior interact with the fault structures of the system in a way that is more aligned with how we would expect the system to behave in reality. We conclude by discussing some directions for future investigation of the effectiveness of geologically consistent priors.

## 1. INTRODUCTION

Computational modeling is an important and widely used tool in the management of geothermal reservoirs (O'Sullivan, Pruess and Lippmann, 2001). However, the usefulness of a reservoir model depends on the accuracy with which it represents the real geothermal system. This is achieved through a calibration process, which typically involves adjusting model parameters, such as subsurface permeabilities and the strengths and locations of hot mass upflows, until there is an adequate match between the model outputs, such as temperatures and pressures, and collected data. In this process, expert domain knowledge is imposed on the permeability distribution which, in theory, should follow geological principles.

The reservoir model we use in this study is a synthetic model which is representative of a typical volcanic hosted geothermal reservoir in New Zealand. This approach, however, could be applied to a wide range of geothermal fields, such as rifting systems or sedimentary basins. The geology (grey), faults (orange) and alteration (blue) of the model are shown in Figure 1. In this work we focus on how structures, such as faults and mineral alteration, alter the permeability of the rock formations, and encode this into the prior distribution used in the Bayesian approach to calibration.

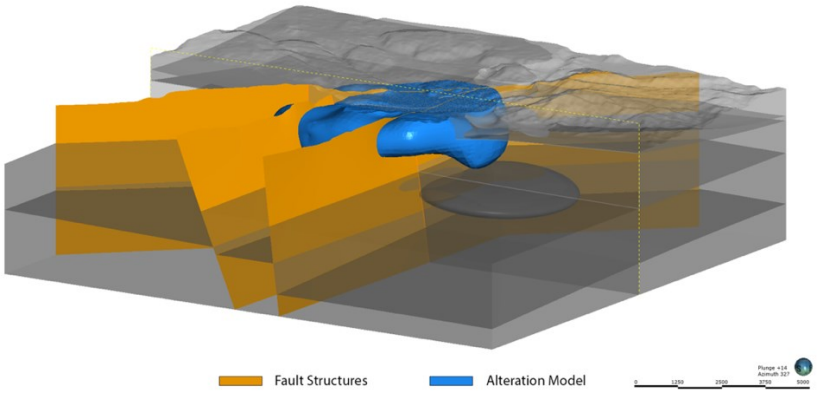


**Figure 1: Fault structures and alteration of the synthetic geothermal reservoir model used. The lithological structure of the geological model is shown in gray.**

### 1.1 Calibration in the Bayesian Framework

Geothermal model calibration can be posed as a statistical inference problem within the Bayesian framework (see e.g. Kaipio and Somersalo, 2005; Aster, Borchers and Thurber, 2018). A key element of the Bayesian approach is that the solution to the calibration problem is not a single set of model parameters; rather, it is an entire probability distribution, referred to as the *posterior*. A second important element of the Bayesian approach is the incorporation of prior knowledge of the model parameters into the problem. This knowledge is represented mathematically as a *prior* distribution. The modeler combines the prior with measured data to form the posterior using Bayes' theorem:

$$p(m|d) \propto p(d|m)p(m) \tag{1}$$

where $m$ represents the unknown model parameters, $d$ represents the observed data, $p(m)$ represents the prior density, $p(d|m)$ represents the likelihood (the conditional density of the observed data given the model parameters) and $p(m|d)$ represents the posterior density (the conditional density of the model parameters given the observed data).

The complex nature of the typical geothermal model makes computation of the posterior a difficult task. A number of methods have been employed within the geothermal context, which differ in terms of the accuracy with which they characterize the posterior and their computational efficiency. These include the construction of a Gaussian approximation to the posterior about the set of parameters with the greatest posterior density (the *maximum-a-posteriori* estimate) (Omagbon et al., 2021), approximate Bayesian computation (Dekkers et al., 2022) and variants of Markov chain Monte Carlo (Cui et al., 2011; Maclaren et al., 2020).

Once characterized, the posterior allows the modeler to describe the system in a probabilistic sense (Aster, Borchers and Thurber, 2018); for instance, one can determine the probability that the upflow in a given region of the reservoir exceeds a specified value. Additionally, the posterior uncertainty in the model parameters can be propagated through to the model forecasts; this allows the modeler to determine, for instance, the probability of the reservoir temperature decreasing below a given value under a future extraction scenario.

### 1.2 Characterization of the Prior

One of the difficulties associated with model calibration using the Bayesian approach is the characterization of the prior. This is a subjective process and typically requires the translation of qualitative information about the system supplied by experts into a quantitative form that can be represented as a probability density. This act of translation is referred to as prior elicitation (see e.g. Garthwaite, Kadane and O'Hagan, 2005; O'Hagan, 2019), and is a widely studied area.

An additional technique often employed to evaluate the quality of a prior is prior predictive checking (Gabry et al., 2018; Gelman et al., 2020). A typical prior predictive check proceeds by running the model using sets of parameters sampled from the prior under consideration and plotting the distribution of the model outputs, or a suitable set of summary statistics. These outputs are then compared with the prior beliefs of experts. If this check reveals issues regarding the outputs, the prior can be re-characterized and the process repeated.

We are particularly interested in the characterization of the prior distribution of the permeabilities of a reservoir model. Common practice is to use a multivariate log-normal distribution to represent each permeability (see e.g. Omagbon et al., 2021; Scott et al., 2022; Dekkers et al., 2022). Priors constructed in this manner typically treat the permeabilities of different rock types as independent, though sometimes correlations are encoded between the permeabilities of individual rock types. Regardless of the correlation structure imposed, however, this type of prior does not account for some important domain knowledge as there exist geological principles that constrain what the permeabilities of different rock types can be. For instance, a fault rock type in a given formation cannot have a higher permeability across the fault than the permeability of the base rock type in the surrounding formation. That is, the effect of the fault is neutral, or it acts as a barrier, impeding the movement of fluid across it. However, a prior constructed using the traditional, naïve approach assigns a non-zero probability to enhanced permeability across the fault. There may be many high-density regions in such a prior within which the permeabilities violate these types of geological principles.

In this paper, we demonstrate how a prior distribution that adheres to these principles can be characterized. We also use several techniques to compare the quality of model outputs associated with samples generated from a prior distribution that adheres to these principles with one constructed using the traditional, naïve approach. Some of these are prior predictive checks, while others involve investigating the match between model outputs and collected data. Our aim is to determine whether constructing the prior in this manner will have an influence on uncertainty quantification of the geothermal model under consideration.

## 2. METHODS

To compare the differences between using geologically consistent and naïve priors, we sampled sets of parameter values for a synthetic model using both types of prior. The model that was used is composed of 11,764 blocks and has the characteristics of a typical New Zealand geothermal system. It includes three faults, six distinct formations and a shallow alteration zone (i.e. a clay cap) (Renaud et al., 2020). Both subsurface permeabilities and hot mass upflows were included when characterizing each prior.

### 2.1 Permeability Samples

Each formation within the model is composed of two regions: one inside the reservoir and one outside inside the reservoir. Each region contains a number of rock types, which fall into four broad categories:

- *Base rock types*: rock that is not part of the clay cap and is not located on a fault.
- *Clay cap rock types*: rock that forms part of the clay cap but is not located on a fault.
- *Fault rock types*: rock that is located on a single fault (inside or outside the clay cap).
- *Intersection rock types*: rock that is located at the intersection of multiple faults (inside or outside the clay cap).

The permeabilities of all 82 distinct subsurface rock types within the model were treated as unknown when characterizing each prior. However, the way in which they were characterized differed.

#### 2.1.1 Geologically Consistent Samples

To generate samples from a geologically consistent prior, we first generated permeability samples for the base rock type of each region within each formation. Figure 2 shows the joint and marginal densities for the permeabilities of a base rock type within the andesite formation of the model.

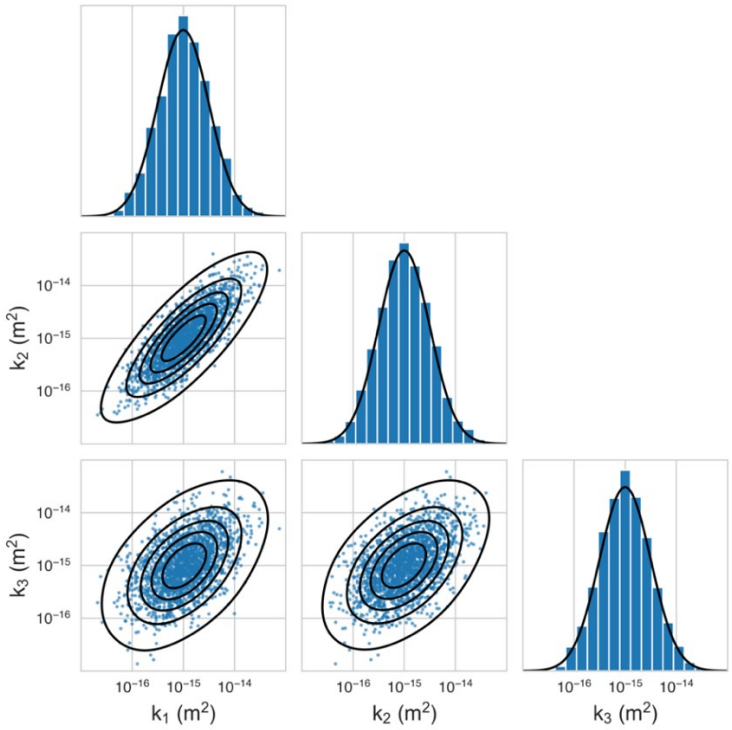


**Figure 2: Joint and marginal densities for the permeabilities of a base rock type within the rhyolite formation of the model (note that the plot uses a log scale). There is a strong correlation between the horizontal permeabilities, and a moderate correlation between each horizontal permeability and the vertical permeability.**

A log-normal distribution was used for each permeability; the mean varied between $10^{-16}$ $m^2$ and $10^{-15}$ $m^2$ depending on the formation and permeability direction, while the standard deviation of the underlying normal distribution was set to 0.5. A correlation of 0.8 was imposed

between the permeabilities in the $k_1$ and $k_2$ (horizontal) directions, and a correlation of 0.5 was imposed between the permeabilities in the $k_1/k_2$ directions and the $k_3$ (vertical) direction.

Permeabilities of the remaining rock types were generated using sets of modifiers that altered the permeabilities of the corresponding base rock type. All sets of modifiers were sampled from (potentially truncated) log-normal distributions; the mean of each distribution varied depending on the formation, permeability direction and rock type under consideration, while the standard deviation of the underlying normal distribution was set to 0.5.

For each region of each formation, sets of multipliers were sampled and multiplied by the permeabilities of the corresponding base rock type to produce the permeabilities of the associated clay cap rock type (in instances where there was no intersection between the region of the formation and the clay cap, this step was skipped). Subsequently, additional set of multipliers were sampled, and multiplied by the permeabilities of the base rock type or clay cap rock type as appropriate, to produce the permeabilities of the fault and intersection rock types inside and outside the clay cap. Figure 3 illustrates the sampling algorithm. All permeabilities were bounded between $10^{-17}$ $m^2$ and $10^{-11}$ $m^2$; if a sampled base permeability violated these constraints, or a sampled modifier produced a permeability outside this range, it was re-sampled. This necessitated the sampling of the base permeabilities and modifiers in a sequential manner, as indicated by the numbers in Figure 3.

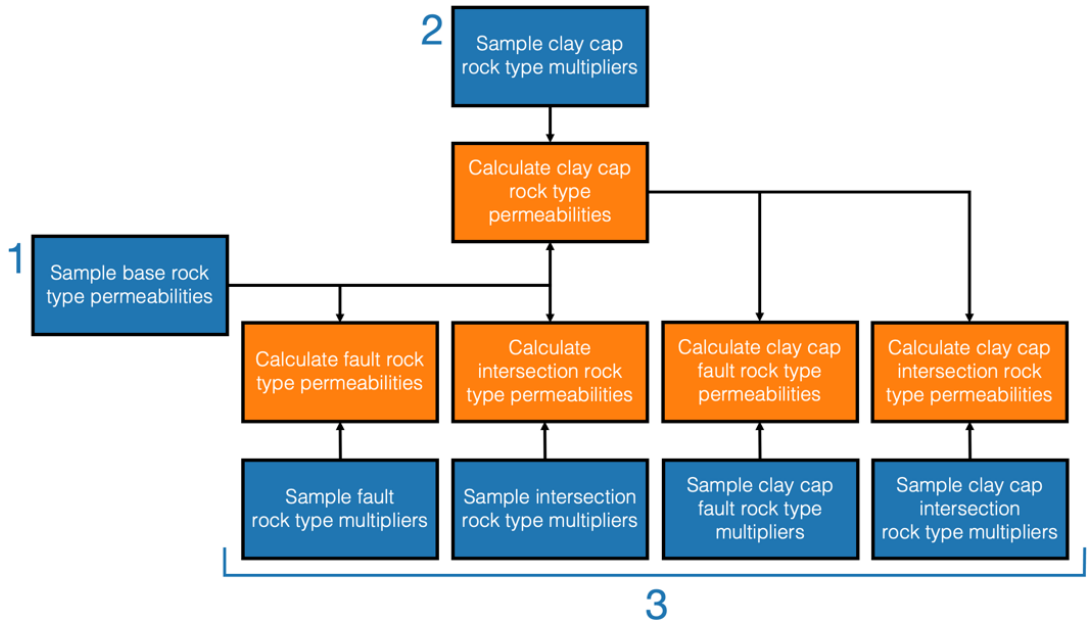


**Figure 3: The algorithm used to sample permeabilities from the geologically consistent prior, starting with each base rock type within each formation. Steps at which quantities were sampled are indicated in blue. Steps at which permeabilities were calculated using sets of modifiers are indicated in orange.**

The distributions of some multipliers were truncated to enforce geological principles. Multipliers used when generating permeabilities of clay cap rock types were constrained to be less than or equal to 1, to prevent the permeability of a formation increasing within the clay cap. Multipliers used when generating permeabilities of fault rock types in the direction along or up the associated fault were constrained to be greater than or equal to 1, to prevent the permeability of a formation decreasing along or up a fault, while modifiers across the fault were constrained to be less than or equal to 1, to prevent the permeability of a formation increasing across a fault. Similarly, multipliers used when generating permeabilities of intersection rock types in the upwards direction were constrained to be greater than or equal to 1.

Figure 4 shows a set of multiplier samples for a fault rock type from the diorite formation of the model, as well as the associated permeability samples of the base rock type, and the permeability samples of the fault rock type (which were generated by taking the product of the multipliers and the base rock type permeability samples).

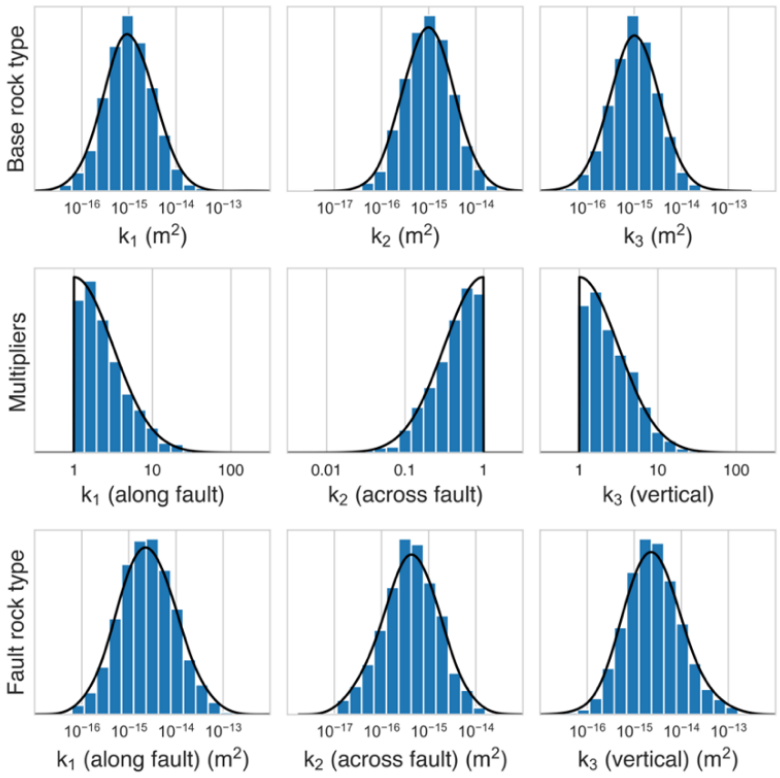


**Figure 4: Samples of the permeabilities of a base rock type (top row) and fault rock type (bottom row) from the diorite formation of the model, and samples of the multipliers (middle row) that were used with the base permeability samples to produce the fault permeability samples.**

2.1.2 Naïve Samples

To generate samples from a naïve prior, we retained the permeability samples of each base rock type generated during the geologically consistent sampling process. However, to generate permeability samples of the remaining rock types, we re-sampled the parameters with replacement from the samples generated using the multipliers. Therefore, while the marginal distributions of each permeability remained approximately log-normally distributed, the permeabilities of different rock types were now independent, meaning that these samples had very similar characteristics to those from a prior constructed in the typical, naïve way. The only substantive difference between the two sets of samples was their correlation structure; this allowed for a fair analysis of the effect of introducing a geologically consistent correlation structure into the prior.

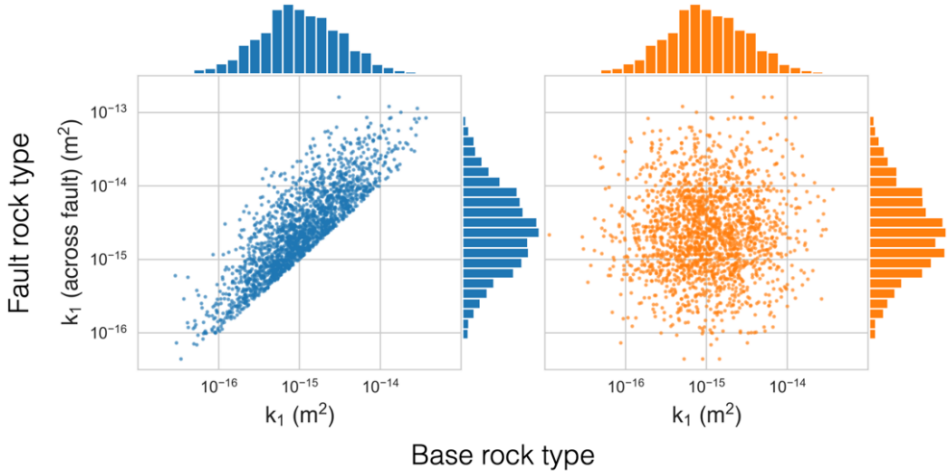


**Figure 5: Permeability samples from geologically consistent (blue, left) and naïve (orange, right) priors, for a base rock type and a corresponding fault rock type from the diorite formation of the model, in the direction across the fault.**

Figure 5 shows an example of the joint and marginal densities of a set of geologically consistent samples, and a set of naïve samples, of a base rock type and a fault rock type of the diorite formation of the model, for the permeability in the direction across the fault. While the marginal distributions are very similar between the two plots, the joint distributions are markedly different. In the geologically consistent samples, the permeabilities of the two rock types are highly correlated, and there are no instances where the permeability of the fault rock type exceeds that of the base rock type. By contrast, there are a large number of naïve samples in which the permeability of the base rock type is lower than that of the fault rock type, violating this geological principle.

### 2.2 Upflow Samples

In addition to the permeabilities, the mass upflows at the base of the model were also treated as unknown. Each of the 24 blocks at the base of the model underneath the clay cap corresponding to a fault or fault intersection rock type was considered as a potential upflow location.

To generate sets of samples, the upflows for each block were sampled from a multivariate Gaussian distribution, truncated at 0. The mean upflows were specified such that they were proportional to the surface area of the top of the corresponding block, and summed to 50 kg/s. An exponential squared covariance matrix of the form described by Nicholson et al. (2021) was used to induce correlations between the upflows of blocks in close proximity along a fault.

The use of a multivariate Gaussian distribution for each potential upflow meant that the distribution of the total upflow within each model was also approximately Gaussian. However, the desired prior distribution for the total upflow was a uniform distribution, with bounds of 30 kg/s and 70 kg/s. We used a probability integral transform to scale the samples such that the totals were uniformly distributed between these bounds.

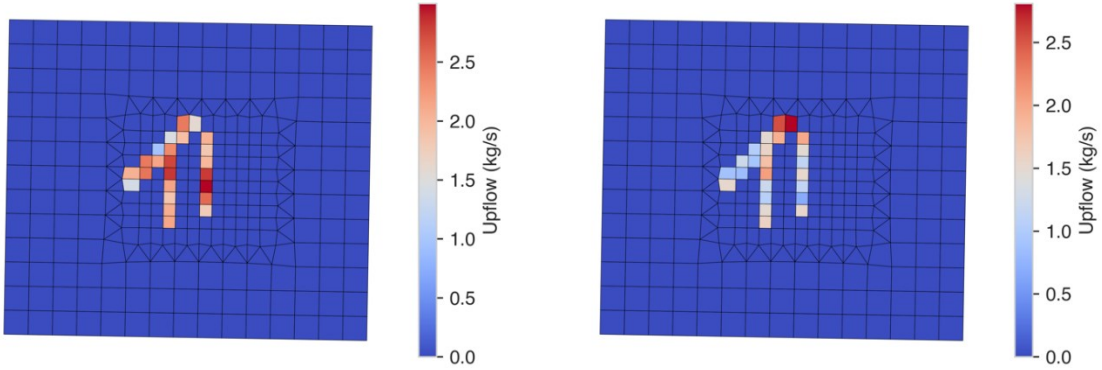


**Figure 6: Sets of sampled upflows in the blocks on the bottom layer of the model.**

The same sets of upflow samples were used with both the geologically consistent and naïve permeability samples. This ensured that any differences in the associated sets of models were the result of the techniques used to sample the permeabilities. Figure 6 shows two examples of sampled upflow sets.

### 2.3 Computation

We sampled 2000 sets of parameters from both the geologically consistent and naïve priors. The model was run to a steady state with each of these sets of parameters using the Waiwera geothermal simulator (Croucher et al., 2019). The New Zealand eScience Infrastructure (NeSI) was used to carry out all required computation. We observed minimal difference in the convergence rates when the model was run with parameters from each type of prior; the model converged to a steady-state solution when using 1784 (89.2%) of the geologically consistent sets of parameters, and 1773 (88.7%) of the naïve sets of parameters.

### 2.4 Analysis of Results

Ideally, parameters sampled from an effective prior distribution should produce plausible geothermal systems. To determine the effectiveness of each type of prior, we carried out several comparisons using summary statistics designed to evaluate the degree to which a set of model outputs resembles a typical geothermal system. Here, we give a high-level overview of each statistic; for a detailed discussion, the reader is referred to Power et al. (2023). The statistics used were:

- The sum of squared differences between the modeled temperatures outside the reservoir, and temperatures predicted using a temperature gradient of 30°C/km, under the assumption that the modeled surface temperatures were accurate (ignoring any instances where the modeled temperature is less than the predicted temperature). One would expect to observe a temperature gradient similar to this if the geothermal system were not present (Fridleifsson et al., 2008; Earle, 2019).
- The sum of absolute differences between the modeled temperatures immediately beneath the bottom of the clay cap, and 190°C. 190°C is the approximate temperature at which the clay cap of a geothermal system begins to form (Gunderson et al., 2000;

Maza et al., 2018); one would therefore expect the modeled temperatures immediately beneath the bottom of the clay cap to be close to this temperature.

A lower value of each summary statistic is indicative of a more realistic geothermal system. The quality of both sets of samples were evaluated using each statistic separately, then the effect of using both statistics to identify the highest-quality samples of each set was investigated.

Subsequently, we considered the match between modeled temperatures produced when running the model using parameters sampled from each type of prior, and downhole temperature profiles.

Finally, we investigated the shapes of the convective plumes produced when running the model using the highest-quality sets of parameters sampled from each type of prior, to determine whether there were any qualitative differences.

## 3. RESULTS AND DISCUSSION

The following sections assess the differences observed in each of the aforementioned comparisons.

### 3.1 Modeled Temperatures Outside Reservoir

Figure 7 shows the distribution of the sum of squared differences between the modeled temperatures outside the reservoir, and those predicted using a temperature gradient of 30ºC/km, for the sets of parameters sampled from each prior. It appears that the distribution associated with the geologically consistent prior has a slightly higher variance than that associated with the naïve prior. The best sets of parameters from each type of prior appear to produce similarly realistic temperature profiles outside the reservoir; however, the worst sets of parameters sampled from the naïve prior appear to produce, in general, slightly more realistic temperature profiles than those sampled from the geologically consistent prior.

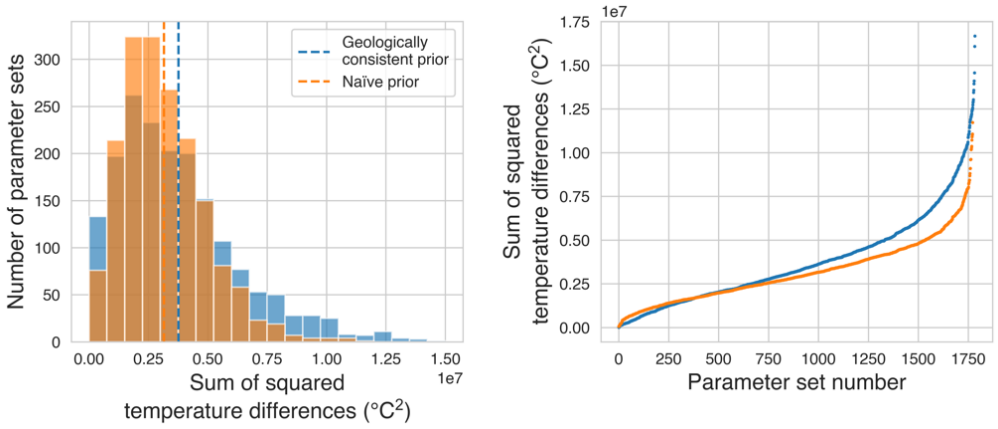


**Figure 7: The sum of squared differences between the modeled temperatures outside the reservoir, and those predicted using a temperature gradient of 30ºC/km, associated with sets of parameters sampled from the geologically consistent (blue) and naïve (orange) priors. The left-hand plot shows the distributions of these differences. The right-hand plot shows each set of parameters ranked from best to worst.**

### 3.2 Modeled Temperatures Immediately Beneath Clay Cap

Figure 8 shows the distribution of the mean modeled temperature immediately beneath the clay cap produced by running the model with parameter sets sampled from both types of prior. It appears that, for both types of prior, the mean temperature is slightly below the ideal value of 190ºC; however, in both cases there exists a relatively high concentration of model outputs in which the mean temperature ranges between 180ºC and 200ºC.

Figure 9 shows the distribution of the sums of absolute differences between the modeled temperatures immediately beneath the clay cap and 190ºC, for parameter sets sampled from both priors. These distributions have similar characteristics to those presented in Figure 7. The distribution associated with parameters sampled from the geologically consistent prior has a slightly greater variance than the distribution associated with parameters sampled from the naïve prior. The best parameter sets sampled from each type of prior appear to produce similarly realistic temperature profiles directly beneath the bottom of the clay cap. However, the worst parameter sets sampled from the naïve prior appear to produce, in general, slightly more realistic temperature profiles.

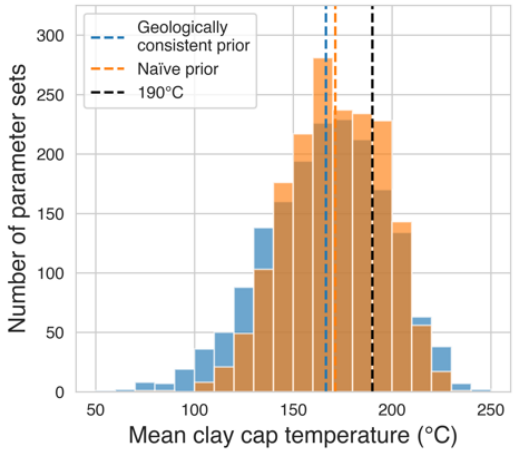


**Figure 8: Distributions of the mean temperature immediately beneath the clay cap, associated with sets of parameters sampled from the geologically consistent (blue) and naïve (orange) priors.**

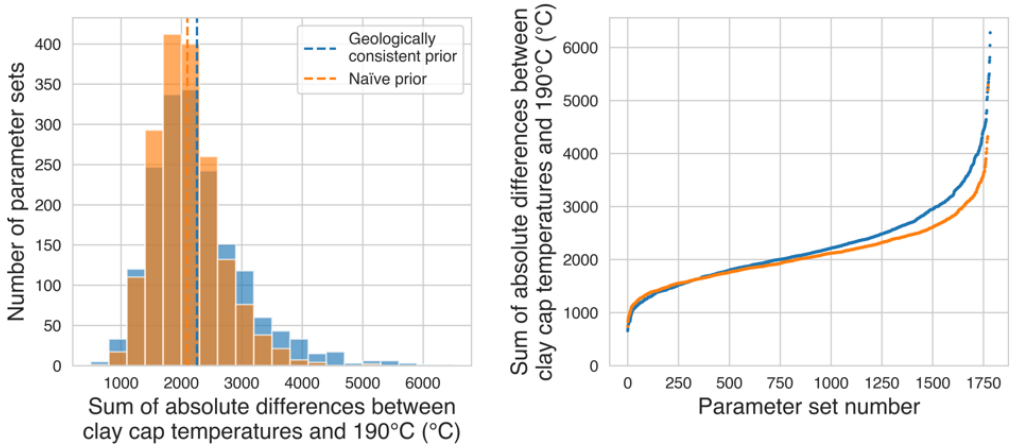


**Figure 9: The sum of the absolute differences between the modeled temperatures inside the clay cap and 190ºC, associated with sets of parameters sampled from the geologically consistent (blue) and naïve (orange) priors. The left-hand plot shows the distributions of these differences. The right-hand plot shows each set of parameters ranked from best to worst.**

### 3.3 Combined Evaluation

In some instances, the modeler is only interested in the degree to which the best-fitting model outputs fit the data. For instance, in its simplest form, approximate Bayesian computation (ABC) (see e.g. Beaumont, 2019) proceeds by sampling sets of parameters from the prior, then running the model with these sets of parameters and evaluating the fit of the outputs to the data using a suitable distance metric. If the distance between a particular sets of model outputs and the data is less than a specified tolerance, the corresponding set of parameters is accepted; otherwise, it is rejected. After the model has been run with an appropriate number of sets of parameters, the accepted sets are used to characterize the posterior. Often, the tolerance is selected such that it corresponds to a desired quantile of the empirical distribution of the distances between the model outputs and observed data, to ensure that a desired proportion of models are accepted. In this scenario, it is clear that while the degree to which the outputs of the accepted sets of parameters fit the data will impact the accuracy with which the posterior is characterized, the degree to which those of the rejected sets of parameters fit the data will not.

To evaluate the characteristics of the best sets of parameters sampled from each prior, we used both statistics in a sequential manner to reject the sets of lowest quality, resulting in those that produced the most realistic temperatures at the bottom of the clay cap and outside the reservoir being accepted. For both sets of parameter samples, the best 25% were selected, based on the modeled temperatures outside the reservoir. Of these, the best 25% were selected, based on the modeled temperatures immediately beneath the clay cap.

Figure 10 shows the distributions of both statistics for both groups of accepted parameter sets. In both cases, these are fairly similar. On average, the modeled temperatures outside the reservoir corresponding to parameter sets sampled from the geologically consistent prior appear to be slightly more realistic, while the modeled temperatures at the bottom of the clay cap in the corresponding parameter sets sampled from the naïve prior appear to be slightly more realistic. However, neither of these differences appear to be very large.

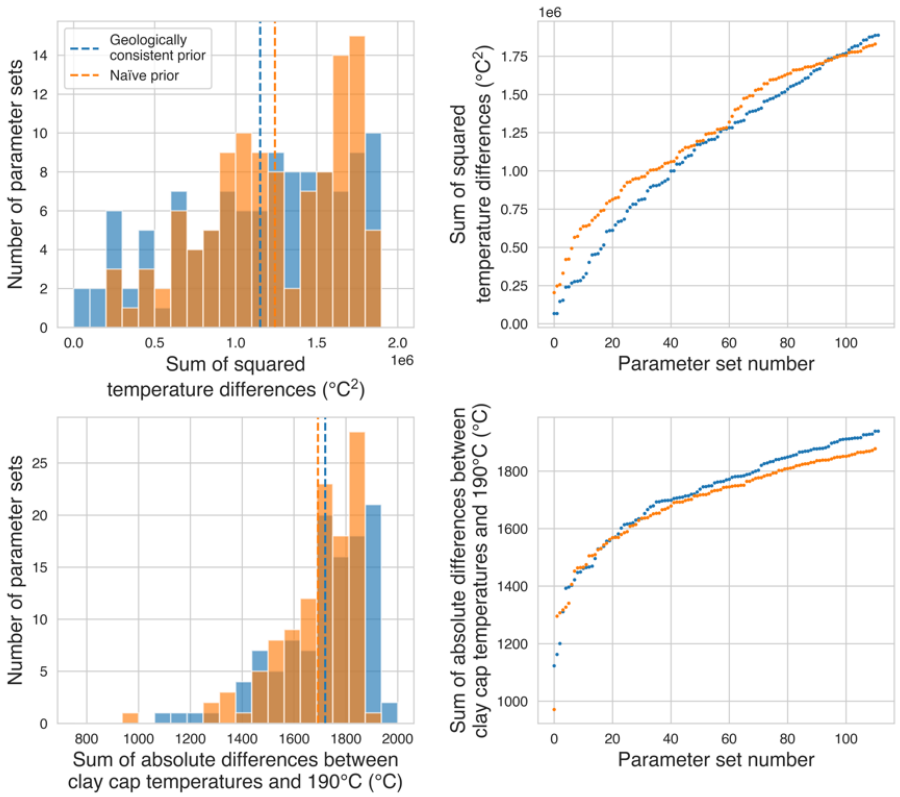


**Figure 10: The sum of squared temperature differences between the temperatures outside of the reservoir and those predicted using a temperature gradient of 30°C/km (top), and the sum of absolute differences between the temperature immediately beneath the clay cap and 190°C (bottom), for sets of parameters from the geologically consistent (blue) and naïve (orange) priors, filtered using these statistics. Distributions of each statistic are shown in the left-hand plots. The right-hand plots show each set of parameters ranked from best to worst.**

### 3.4 Natural State Downhole Temperature Profiles

Next, we investigated the match between the modeled temperatures associated with each prior, and natural state downhole temperature data. The location of the wells within the system are shown in Figure 11. Each was associated with a set of synthetic downhole data. The 10% of samples that matched this data the best, based on the sum of squared distances between the modeled temperatures and measured data, were filtered for both the naïve and geologically consistent priors.

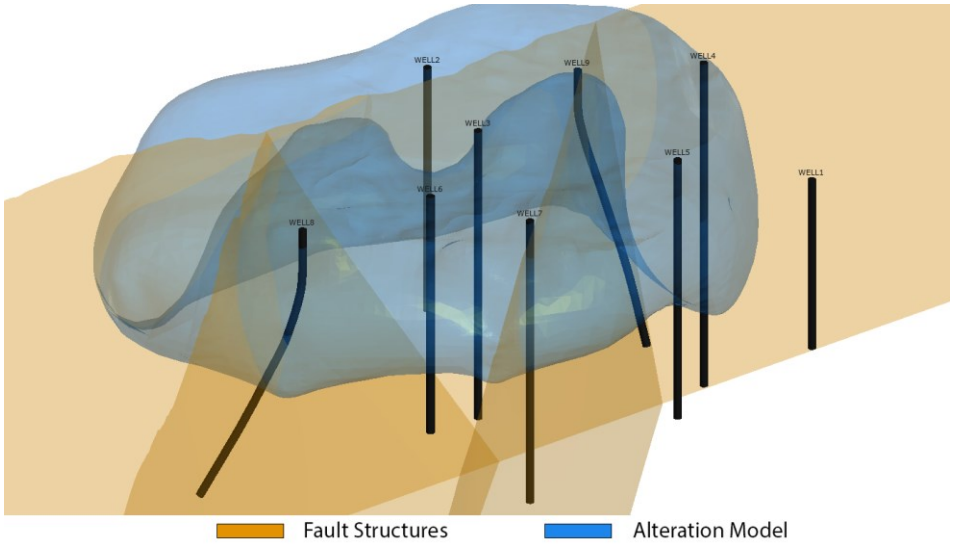


**Figure 11: Locations of wells used for conditioning on natural state downhole temperatures.**

Figure 12 shows the modeled downhole temperatures associated with all samples from each prior, the filtered samples, and the data used. These are shown for WELL-3, WELL-5 and WELL-8, which are representative of a well in a fault, a well in the formation and a well on the outer edge of the reservoir respectively. We see that the fit between the data and the modeled temperature profiles is comparable between the two sets of priors. This is consistent with the discussion in Sections 3.1-3. It appears possible that samples from the geologically consistent prior tend to produce temperature profiles that respect the data in zones where there is high upflow (WELL-3) or no upflow (WELL-8) and hence show a slightly better fit to the data. However, more investigation is required as to the underlying cause of this result.

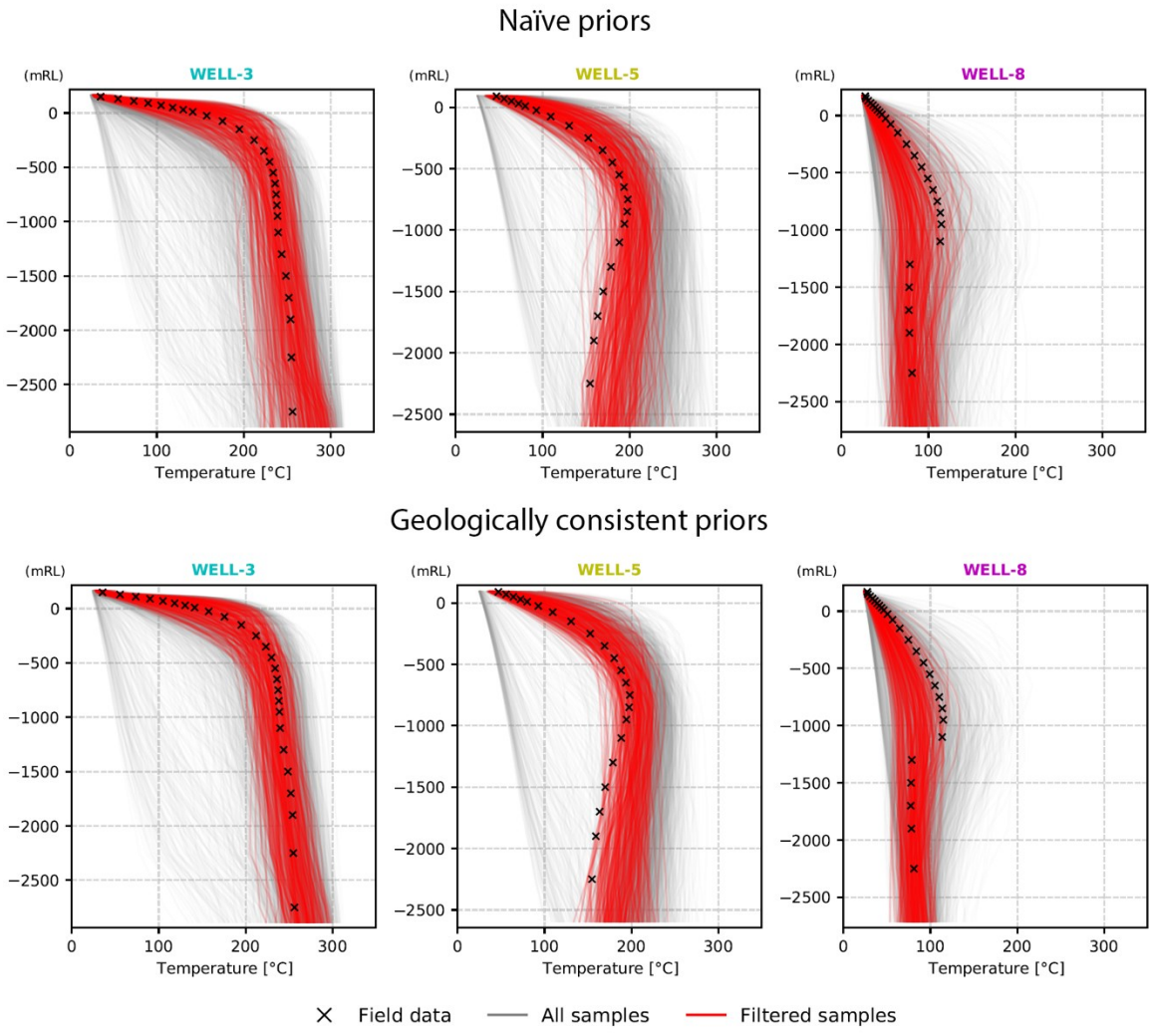


**Figure 12: Modeled temperatures produced by naïve (top) and geologically consistent (bottom) prior samples conditioned on natural state downhole temperature data. Three wells are shown (the locations of each are indicated in Figure 11). Rejected temperature profiles are shown in gray, filtered profiles are shown in red, and field data is marked with crosses.**

### 3.5 Plume Comparisons

Finally, we investigated the convective plumes produced by the best samples from each type of system, to determine whether there were any qualitative differences.

Figure 13 shows the plumes produced by the three best sets of parameters based on the clay cap condition, after the filtering process described in Section 3.3, from both the geologically consistent prior and the naïve prior. There are some obvious differences between each set of plumes; those from the geologically consistent prior appear to be well-contained within the clay cap and follow the fault structures well. By contrast, the plumes from the naïve prior shift laterally as they move upwards, moving across one of the faults and extending beyond the edges of the clay cap. It appears that when the parameter sets sampled from the naïve prior are used, the fault acts as less of a barrier to impede the movement of fluid across it. The plumes associated with the samples from the geologically consistent prior follow the fault structures better, showing greater alignment with how we would expect the system to behave in reality.

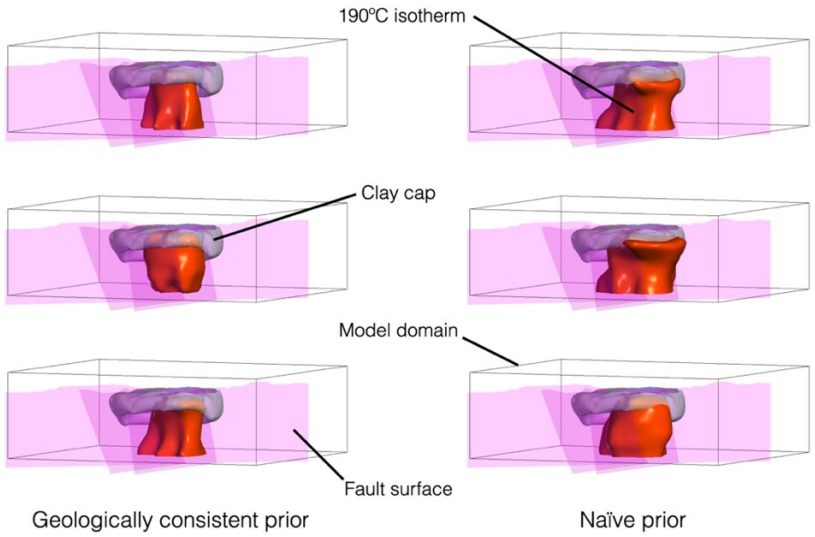


**Figure 13: Comparisons of the plumes produced using parameters sampled from geologically consistent (left) and naïve (right) priors. The gray lines indicate the boundary of the model domain, and each purple slice denotes a fault. 190°C isotherms are shown in orange and the clay cap is shown in gray.**

## 4. CONCLUSIONS

We have developed a method to characterize a prior distribution for the parameters of a reservoir model that adheres to simple geological principles. We have also investigated several methods of comparing a geologically consistent prior with a naïve prior which treats different rock types within the same formation as independent. We have drawn several conclusions from this work.

Based on the summary statistics we have used to quantify whether each prior produces model outputs that are representative of a typical geothermal system, it appears difficult to determine whether the model outputs generated by running the model with parameter sets sampled from a geologically consistent prior are better than those generated using a naïve prior. The same is true for the agreement between the modeled temperatures associated with samples from each prior, and downhole temperature data. On one hand, this suggests that perhaps the outputs generated using a naïve prior are not as poor as one might intuitively expect. On the other hand, there are certainly some differences in the distributions of each summary statistic, as well as the filtered downhole temperature profiles, when the model is run using parameters sampled from each type of prior. This suggests that adhering to geological principles has at least some degree of influence on the characteristics of the resulting model outputs. The summary statistics we have used condense large amounts of information contained within the model outputs into a single dimension; therefore, while useful, they have limitations in the amount of information they are able to convey. It is of interest to investigate alternative ways of quantifying the characteristics of reservoir model outputs, with the aim of being able to more effectively explain some of the differences we have observed between the outputs associated with each type of prior.

The convective plumes associated with the best sets of parameters sampled from the geologically consistent prior appear more realistic than those associated with sets of parameters sampled from the naïve prior, as they follow the fault structures of the model in a way that is more consistent with how we would expect this type of system to operate in reality. However, this conclusion has been drawn based on a small sample size, and there is no guarantee that the trends we have observed here will hold outside the best sets of parameters. It is of interest to compare a greater number of plumes produced by parameters generated from each type of prior, including models that the statistics we have used as part of this study suggest are poorly representative of a typical geothermal system.

In addition to the aforementioned improvements to the methods used in this study, we have identified some further areas in which work could be conducted to help in determining the effectiveness of a geologically consistent prior relative to a naïve prior.

This paper has only considered natural-state simulations. There may, however, be differences in how parameters sampled from a geologically consistent prior result in transient phenomena being modeled. It is of interest to simulate the production history of the model using sets of parameters sampled from each type of prior; this will allow for a comparison of how phenomena such as pressure drops within the reservoir as a result of fluid extraction are modeled when each type of prior is used.

Also of interest is to use this sampling technique for the parameters of a more complex model. The model we used in this study is small and fairly robust to changes in its parameters; the non-convergence rate during uncertainty quantification studies is often a lot higher than what was observed here. It is conceivable that when using a larger reservoir model, we would observe a higher failure rate or longer run times for models generated using parameters sampled from a naïve prior, as a result of inconsistent parameter choices causing complex reservoir dynamics.

Overall, we believe that geologically consistent priors are likely to be an effective way to represent one's initial beliefs during reservoir model calibration within the Bayesian framework, and should be investigated further.

**REFERENCES**

Aster, R. C., Borchers, B., & Thurber, C. H.: Parameter Estimation and Inverse Problems, Third Edition. *Elsevier* (2018).

Beamont, M. A.: Approximate Bayesian Computation. *Annual Review of Statistics and Its Application*, 6, (2019), 379-403.

Croucher, A., O'Sullivan, M., O'Sullivan, J., Yeh, A., Burnell, J., & Kissling, W.: Waiwera: A parallel open-source geothermal flow simulator. *Computers and Geosciences*, 141, (2020), 104529.

Cui, T., Fox, C., & O'Sullivan, M. J.: Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis-Hastings algorithm. *Water Resources Research*, 47(10), (2011), W10521.

Dekkers, K., Gravatt, M., Maclaren, O. J., Nicholson, R., Nugraha, R., O'Sullivan, M., Popineau, J., Riffault, J., & O'Sullivan, J.: Resource Assessment: Estimating the Potential of a Geothermal Reservoir. *Proceedings 47th Workshop on Geothermal Reservoir Engineering*, Stanford, CA (2022).

Earle, S.: Physical Geology, Second Edition. *BCcampus* (2019).

Fridleifsson, I.B., Bertani, R., Huenges, E., Lund, J.W., Ragnarsson, A., & Rybach, L.: The possible role and contribution of geothermal energy to the mitigation of climate change. *Proceedings IPCC Scoping Meeting on Renewable Energy Sources*, Luebeck, Germany (2008).

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A.: Visualization in Bayesian workflow. *Journal of the Royal Statistical Society A*, 182(2), (2019), 389-441.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A.: Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association,* 100(470), (2005), 680-701.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P., & Modrák, M. Bayesian workflow. *arXiv preprint*, (2020).

Gunderson, R., Cumming, W., Astra, D., & Harvey, C.: Analysis of smectite clays in geothermal drill cuttings by the methylene blue method: for well site geothermometry and resistivity sounding correlation. *Proceedings World Geothermal Congress*, Kyushu - Tohoku, Japan (2000).

Kaipio, J., & Somersalo, E.: Statistical and Computational Inverse Problems. *Springer Science and Business Media* (2006).

Maclaren, O. J., Nicholson, R., Bjarkason, E. K., O'Sullivan, J. P., & O'Sullivan, M. J.: Incorporating Posterior-Informed Approximation Errors Into a Hierarchical Framework to Facilitate Out-of-the-Box MCMC Sampling for Geothermal Inverse Problems and Uncertainty Quantification. *Water Resources Research*, 56(1), (2020), e2019WR024240.

Maza, S.N., Collo, G., Morata, D., Lizana, C., Camus, E., Taussi, M., Renzulli, A., Mattioli, M., Godoy, B., Alvera, B., Pizarro, M., Ramírez, C., & Rivera, G.: Clay mineral associations in the clay cap from the Cerro Pabellón blind geothermal system, Andean Cordillera, Northern Chile. *Clay Minerals*, 53(2), (2018), 117-141.

Nicholson, R., Maclaren, O. J., O'Sullivan, J. P., O'Sullivan, M. J., Suzuki, A., & Bjarkason, E. K.: Representation of Unknown Parameters in Geothermal Model Calibration. *Proceedings 42nd New Zealand Geothermal Workshop*, Waitangi, New Zealand (2020).

O'Sullivan, M. J., Pruess, K., & Lippmann, M. J.: State of the art of geothermal reservoir simulation. *Geothermics*, 30(40, (2001), 395-429.

O'Hagan, A.: Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, 73(S1), (2019), 69-81.

Omagbon, J., Doherty, J., Yeh, A., Colina, R., O'Sullivan, J., McDowell, J., Nicholson, R., Maclaren, O. J., & O'Sullivan, M.: Case studies of predictive uncertainty quantification for geothermal models. *Geothermics*, 97, (2021), 102263.

Power, A., Dekkers, K., Gravatt, M., Maclaren, O., Nicholson, R., O'Sullivan, J., de Beer, & A., O'Sullivan, M.: Improved Filtering for a new Resource Assessment Method. *Proceedings 48th Workshop on Geothermal Reservoir Engineering*, Stanford, CA (2023).

Renaud, T., Popineau, J., Riffault, J., O'Sullivan, J., Gravatt, M., Yeh, A., Croucher, A., & O'Sullivan, M.: Practical Workflow for Training in Geothermal Reservoir Modelling. *Proceedings 43rd New Zealand Geothermal Workshop*, Wellington, New Zealand (2021).

Scott, S. W., O'Sullivan, J. P., Maclaren, O. J., Nicholson, R., Covell, C., Newson, J., & Guðjónsdóttir, M. S.: Bayesian Calibration of a Natural State Geothermal Reservoir, Krafla, North Iceland. *Water Resources Research*, 58(2), (2022), e2021WR031254.