

Declaration of Contributions

All of the results presented in this report are the collective work of the author of this paper, Andrew Power, and my project partner, Daniel Wong with the exception of the following contributions:

- The Python package *ccandu* was developed by our supervisors in the Modelling Uncertainty and Data Group,
- The Waiwera geothermal simulations were performed by our supervisors in the Geothermal Institute,
- The code to analyse the effect of normalising the data on prediction accuracy was written by Daniel Wong,
- The modelling of the geothermal reservoir in the transient case, and the comparison of data-space inversion to approximate Bayesian computation were done by myself.

Additionally, the conference article prepared for the New Zealand Geothermal Workshop referenced in this report is the collaborative effort of both Daniel Wong and myself, as well as our supervisors.

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF ENGINEERING SCIENCE

DATA-SPACE INVERSION FOR COMPUTATIONALLY
EXPENSIVE GEOTHERMAL MODEL PREDICTIONS AND
UNCERTAINTY QUANTIFICATION

PART IV PROJECT FINAL REPORT

Author:
Andrew POWER

Supervisors:
Ruanui NICHOLSON
Oliver MACLAREN
Michael GRAVATT



ENGINEERING
DEPARTMENT OF ENGINEERING SCIENCE

October 15, 2021

Abstract

The use of simulation modelling to make predictions of certain quantities of interest, such as the temperature and pressure, in a geothermal system is necessary for their effective operation. Traditional methods of making predictions are time-intensive and computationally expensive due to the serial nature of model simulations during calibration. First, parameters are estimated by calibrating to measured data, before these parameters are propagated through to posterior predictions. However, these parameters are often not of direct interest; the predictions themselves provide much more value. As such, data-space inversion provides a more efficient method to making predictions with quantifiable uncertainty without the need to explicitly estimate the parameter values. The approach is implemented on three models of varying complexity: a simple polynomial model, a zero-dimensional lumped parameter model, and a three-dimensional geothermal model. This has shown that data-space inversion has the ability to make accurate predictions in both space and time, in a variety of scenarios. However, factors such as the quantity of observed data relative to the number of predictions to be made, and the location and arrangement of wells in a geothermal field can have a significant impact on the quality of results. Data-space inversion is also compared to an alternative calibration-free approach of approximating the posterior predictive distribution. Approximate Bayesian computation is less sensitive to the above factors making more accurate predictions than data-space inversion, yet this comes at the cost of a posterior predictive distribution with greater uncertainty. The effect of normalising the measured data is inconsistent, with predictions both increasing and decreasing in accuracy while others remain relatively unchanged. Implementing data-space inversion is more efficient than existing approaches. This has implications on the ability to define geothermal systems with higher resolution leading to more realistic, accurate predictions. It is statistically well-founded, and allows the use of simulation modelling to be accessible to a wider range of applications on which further research can be performed.

Acknowledgements

I would like to offer special thanks to Ru, Oliver, and Michael, the supervisors of this project, who have provided seemingly infinite wisdom and time to ensure that this project went smoothly. Thanks must also go to Ken for his involvement in performing the many geothermal simulations.

Finally, I would like to acknowledge the support of my family and friends throughout the year. As always, this is greatly appreciated.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Literature Review | 2 |
| 2.1 | Existing Literature | 2 |
| 2.1.1 | Geothermal Reservoir Modelling | 2 |
| 2.1.2 | Traditional Forward and Inverse Problems | 2 |
| 2.1.3 | Data-Space Inversion | 3 |
| 2.1.4 | Approximate Bayesian Computation | 3 |
| 2.2 | Current Research | 4 |
| 2.3 | Research Objectives | 4 |
| 3 | Methods | 5 |
| 3.1 | Traditional Bayesian Problems | 5 |
| 3.2 | Data-Space Inversion | 6 |
| 3.3 | Geothermal Models | 6 |
| 3.3.1 | Lumped Parameter Model | 7 |
| 3.3.2 | 3D Synthetic Model | 7 |
| 3.3.3 | Waiwera | 8 |
| 3.4 | <i>ccandu</i> | 8 |
| 3.5 | Approximate Bayesian Computation | 9 |
| 4 | Results | 10 |
| 4.1 | Toy Polynomial Model | 10 |
| 4.2 | Geothermal Lumped Parameter Model | 11 |
| 4.3 | Realistic Geothermal Model | 13 |
| 4.3.1 | Natural State | 14 |
| 4.3.2 | Production | 15 |
| 4.4 | Approximate Bayesian Computation | 16 |
| 4.5 | Normality | 18 |
| 5 | Discussion | 22 |
| 5.1 | Findings | 22 |
| 5.2 | Related Research | 23 |
| 5.3 | Limitations and Future Work | 23 |
| 5.3.1 | Modelling Approaches | 24 |
| 5.3.2 | Non-Invasive Modelling | 24 |
| 6 | Conclusions | 26 |

List of Figures

| | | |
|----|---|----|
| 1 | Simplified Conceptual Model of the Geothermal Lumped Parameter Model | 7 |
| 2 | Conceptual Model of the 3D Geothermal System [4] | 8 |
| 3 | DSI on a First-Order Polynomial without Noise (Left) and with Noise (Right) . . . | 10 |
| 4 | DSI on a Second-Order Polynomial (Left) and a Third-Order Polynomial (Right) . . | 11 |
| 5 | DSI on a Geothermal Lumped Parameter Model | 12 |
| 6 | The Effect on Accuracy and Variability of Fewer Observations | 13 |
| 7 | Visualisation of the Observation and Prediction Wells in the Spatial Domain . . . | 14 |
| 8 | Prior Ensemble of Model Realisations and Observed Data for the Observation Wells | 15 |
| 9 | Spatial Posterior Distribution of Temperature Predictions for an Interpolated Well (Left) and Extrapolated Well (Right) in the Natural State | 15 |
| 10 | Visualisation of the Feedzone for Well 2 in the Spatial Domain | 16 |
| 11 | Transient Posterior Distribution of Temperature Predictions at the Feedzone of Well 2 | 17 |
| 12 | Temperature Predictions at the Feedzone of an Injection Well | 17 |
| 13 | Temperature Predictions at the Feedzone of a Monitor Well | 17 |
| 14 | ABC on a Geothermal Lumped Parameter Model | 18 |
| 15 | DSI on the Same Model for Comparison | 19 |
| 16 | ABC on a Geothermal Lumped Parameter Model | 19 |
| 17 | DSI on the Same Model for Comparison | 20 |
| 18 | Visualisation of a Feature Before (Left) and After (Right) Transforming to Normality | 20 |
| 19 | Posterior Distribution of Temperature Predictions after Transforming for Normality | 21 |

1 Introduction

In the field of geothermal reservoir modelling, computational models that solve mass and energy conservation over a spatial and temporal domain are often used to inform decision making involving the operation of a geothermal system. This modelling could be used to understand more about the geothermal reservoir, or to make predictions about physical properties of the system, for example, the temperature and pressure of subsurface fluid. Factors such as the duration of economic exploitation of the reservoir, how it will respond to varying levels of production, and the optimal re-injection of fluid can all be predicted through effective geothermal modelling [1].

The traditional method of generating predictions requires two steps. First, uncertain parameters, such as subsurface permeabilities of different rock types and strength of deep upflows, are calibrated to measured data, along with the associated posterior uncertainty. Second, these uncertainties from the parameters are propagated through to predictions using model runs. Calibration is the process of tuning the model parameters to fit the measured data. This can be done computationally, although, in geothermal modelling, calibration is regularly performed manually and can take extensive periods of time. However, the model parameters are often not of direct interest, rather, the predictions about the reservoir contain more useful information. The calibration of model parameters typically requires the most time in the modelling process, so a new method is proposed that will remove the need for calibration while still considering the measured data.

Data-space inversion (DSI) is a novel approach, presenting an alternative to the traditional method by providing approximative estimate predictions and their associated uncertainty. This should, in theory, reduce the time taken by allowing model runs to be performed in parallel rather than in series as would be required from calibration techniques while still giving reasonably accurate predictions with quantifiable uncertainty. Research into the performance of DSI has focused on subsurface hydrology [2, 3], so an analysis of the approach in the context of geothermal reservoir modelling is unique.

This research will examine DSI through its performance on a set of models of varying complexity. A polynomial model demonstrates the feasibility of the approach, before the geothermal context is introduced with a simplified zero-dimensional reservoir model and a synthetic three-dimensional geothermal reservoir [4] which is on a more realistic scale. The application of DSI to the full geothermal model is further examined in both natural state and production scenarios, providing insight into how accurately the properties of a reservoir can be predicted in both space and time.

Finally, some critical analysis of DSI through its comparison to another calibration-free approach is presented, highlighting the benefits and shortcomings of both methods for predicting the properties of geothermal reservoirs. Approximate Bayesian computation (ABC) is a simple approach that gives approximate posterior predictions with quantifiable uncertainty, and is performed on the same zero-dimensional model as DSI for an unbiased comparison. The effect of normalising data is also explored to justify whether a transformation is beneficial before DSI is implemented.

2 Literature Review

The computational modelling of geophysical systems, which are often complex and non-linear due to the governing equations, is critical for the effective operation of geothermal reservoirs, yet calibration of the model parameters can be time consuming. Before an alternative approach to mitigate this limitation is explored, an overview of what is currently the norm in geothermal reservoir modelling and solving inverse problems using a standard Bayesian statistical approach is presented.

2.1 Existing Literature

2.1.1 Geothermal Reservoir Modelling

Initial attempts to model geothermal reservoirs were received with skepticism, but were widely accepted in 1980 with the advent of the Code Comparison Project when a set of test problems was designed for the evaluation and comparison of reservoir simulators [5]. Fradkin et al. [6] presented a modelling framework for the Wairakei geothermal field which concluded that a slow drainage lumped parameter model (LPM) had good forecasting power, and supplied good estimates of model parameters such as porosity and permeability. The LPM in this study forms the basis of the zero-dimensional geothermal model considered in Section 4.2. In 1985, Bodvarsson et al. [7] exposed a significant shortcoming in geothermal reservoir modelling stating that no universal modelling strategy was applicable to all the hydrological, thermal, chemical, and mechanical processes occurring. Furthermore, no techniques had yet been devised to simulate the flow of fluid and heat in porous media in one-, two-, or three-dimensions [7]. Reservoir simulators with the ability to model fluid and heat flow through porous media were presented in 1990 [8], however, they still lacked the capacity to represent chemical and mechanical processes. The efficiency of numerical computations was a key limitation to solving complex geothermal problems, particularly to three-dimensional models which offered greatly improved geological realism [8].

Advances in computing and modelling power have since led to the ability to define models with increased definition and solve them reasonably efficiently, although geothermal system models with 50,000 to 100,000 blocks may still take weeks to run [9]. It is not uncommon to discretise three-dimensional models to contain a number of blocks of the order $\mathcal{O}(10^3 - 10^4)$ [9, 10], however, these blocks are not of an inconsequential size, often being at least 200 metres square horizontally and 100 metres in depth [10]. While existing geothermal modelling tools are capable of incorporating most of the relevant processes of a reservoir as either input or calibration data [9], Burnell et al. [9] suggests that the coupling of different processes, such as the fluid flow, geophysics, mechanics, and chemistry, will provide confidence in the development and validation of such models. This added complexity requires improved numerical computation for it to be efficient, or an alternative approach to modelling and making predictions about geothermal systems.

2.1.2 Traditional Forward and Inverse Problems

Formulating predictions for the properties and behaviour of a geothermal reservoir requires the solving of both the inverse problem and the forward problem. The inverse problem attempts to reconstruct the model, in particular the model parameters, from measured data, and can be done in multiple ways. Typically, manual calibration adjusts the parameter values using knowledge of the geothermal system, and comparing the resulting model simulation to measured data. It can also be performed computationally by estimating the *maximum a posteriori* (MAP) estimate

which maximises the posterior probability density [11, 12]. While estimating these parameter values is sensitive to the measured data due to the non-uniqueness of the inverse problem [13, 14], the calibrated parameter values, and their uncertainty, can be mapped to predictions by running the forward model [12]. Calibration comes at the greatest computational expense as calculating the MAP estimate often includes solving a high-dimensional, possibly non-convex, optimisation problem [15].

Inverse problems are typically solved in the Bayesian framework [11, 14, 16], the derivation of which is discussed in Section 3.1. A significant benefit of approaching such a problem from a Bayesian perspective is that it allows for the quantification of uncertainty which naturally permits various sources and types of uncertainty to be incorporated into the model [15]. This uncertainty quantification is the result of linear uncertainty propagation which gives a Gaussian approximation to the posterior predictive distribution [15]. Multiple posterior geological realisations of the model are generated from this predictive distribution as a means to quantify the uncertainty of predictions, however, it can be difficult to ensure geological realism resulting in unphysical predictions, for example, negative pressures in the reservoir [2].

2.1.3 Data-Space Inversion

Subsurface hydrology research communities have been amongst the first to adopt a DSIs approach for reservoir forecasting [2, 3]. This initial exploration of the method has shown that direct forecasts can be made without explicit model inversion [3], and the predictions made are consistent with those produced using traditional methods [2, 17]. Moreover, in one example, this traditional approach required $\mathcal{O}(10^5 - 10^6)$ flow simulations, while DSIs required only 500 [2] leading to substantial computational savings. An ensemble of prior model realisations simulated from prior geothermal knowledge of parameter values is central to the application of DSIs, however, when measured data lies outside the bounds of these prior models, the forecasts may be unreliable. Sun and Durlofsky [2] recommend that wider priors are used, so some thought should be given to the prior distribution of each model parameter to ensure geological realism and reliable results. Further, it is important to note that DSIs also allows for the quantification of uncertainty [17] through the generation of a conditional posterior predictive distribution [15] which is discussed in more detail in Section 3.2.

2.1.4 Approximate Bayesian Computation

ABC is another method for approximating a posterior distribution that relies on Bayesian statistics [18]. As the evaluation of the likelihood function can be computationally expensive or even impossible, ABC approximates the likelihood by simulating model realisations using parameter values sampled from the prior distribution and running them through the forward model [19].

The tolerance parameter ε that determines whether a simulation is accepted is important in modelling the model error. This is because Wilkinson [19] has shown that the tolerance used can be related to the noise level under the assumption that the noise is uniformly distributed. Since it is not often the case that model errors are uniformly distributed, adapting the tolerance to be an acceptance density, in which the probability of a simulation being accepted varies with its distance from the observed data, allows the uniform error to be replaced by a general distribution [19]. This offers significant flexibility to modelling realistic scenarios.

2.2 Current Research

The DSI approach is relatively new resulting in little literature currently being published. It has, however, been used on occasion by the subsurface hydrology community [2, 3], and, to a lesser extent, in a reservoir modelling context [17, 20].

The use of DSI to make spatial predictions with quantifiable uncertainty in a geothermal context is presented in a conference article prepared by the author and their supervisors of this project for the New Zealand Geothermal Workshop [15]. It is written as a companion paper to this project. However, a transient case, i.e., making predictions in time, using a production history reservoir model is not considered, so this current study builds on the findings to develop the ideas presented.

Lima et al. [21] has demonstrated that using an iterative ensemble smoother to define the posterior distribution is computationally more efficient and robust than the standard implementation of DSI. Furthermore, this new DSI method also produces predictions of a comparable accuracy to those made using a traditional DSI approach [21].

2.3 Research Objectives

This research aims to build on current literature by considering the application of DSI to geothermal reservoir modelling which is potentially more challenging due to the non-linearity of many of its governing equations. Showing that the use of DSI is viable to make accurate predictions, both in space across the geometry of the geothermal reservoir and in time, could be highly advantageous to the reservoir modelling community. Gains to the efficiency with which reliable predictions can be made would allow for more complex geometries and processes to be modelled without the concern of simulations taking excessively long to run. Finally, this research aims to compare DSI to another calibration-free approach to making posterior predictions in an attempt to discover whether particular methods give more accurate results in different situations.

3 Methods

Before these modelling approaches are applied to each model, an overview of the mathematical foundation of each is presented. Derivations of the Bayesian approach and DSI are shown, and the conceptual models for the LPM and simplified three-dimensional geothermal reservoir are described to aid in an understanding of the results. A brief explanation of the underlying principle of ABC is also discussed. For more details see Wilkinson [19].

3.1 Traditional Bayesian Problems

The Bayesian framework for solving inverse problems relies heavily on statistics, and is outlined in Kaipio and Somersalo [11] and Power et al. [15]. A typical model is of the form

$$y_{obs} = f(\theta) + \varepsilon, \quad (1)$$

where $y_{obs} \in \mathbb{R}^d$ are the observed data, f is the model that maps from $\mathbb{R}^n \rightarrow \mathbb{R}^d$, $\theta \in \mathbb{R}^n$ are the model parameters that are being estimated, and $\varepsilon \in \mathbb{R}^d$ encapsulates any error in the modelling process [11]. To calibrate the parameters to the data, the estimate that maximises the parameter posterior distribution is calculated. This posterior distribution of Θ is the conditional distribution of Θ given $Y = y_{obs}$

$$\pi(\theta|y_{obs}) = \frac{\pi(\theta, y_{obs})}{\pi(y_{obs})} = \frac{\pi_{pr}(\theta)\pi(y_{obs}|\theta)}{\pi(y_{obs})}, \quad (2)$$

where $\pi_{pr}(\theta)$ is the prior probability density of θ that contains the beliefs or domain knowledge of the parameters.

If it is assumed that the random variable for the parameters Θ is independent of the random variable for the error in the model E , the probability density of the error remains unchanged after conditioning on the parameters $\Theta = \theta$ [11]. Therefore, Y conditioned on $\Theta = \theta$ must be distributed in the same way as E allowing the likelihood function for y_{obs} to be expressed as

$$\pi(y_{obs}|\theta) = \pi_e(y_{obs} - f(\theta)), \quad (3)$$

where π_e is the probability density of the model error. From Equations (2) and (3), the parameter posterior distribution can be expressed as

$$\pi(\theta|y_{obs}) \propto \pi_{pr}(\theta)\pi_e(y_{obs} - f(\theta)). \quad (4)$$

In the Bayesian framework, it is common to assume *a priori* that the parameters and errors are normally distributed, i.e., $\Theta \sim \mathcal{N}(\theta_0, C_{pr})$ and $E \sim \mathcal{N}(0, C_e)$ [15] where C_{pr} and C_e are the covariances of the prior distribution and the error distribution respectively. The density functions for these distributions are well-known and are expressed in Equations (5) and (6),

$$\pi_{pr}(\theta) \propto \exp\left(-\frac{1}{2}(\theta - \theta_0)^\top C_{pr}^{-1}(\theta - \theta_0)\right), \quad (5)$$

$$\pi_e(y_{obs} - f(\theta)) \propto \exp\left(-\frac{1}{2}(y_{obs} - f(\theta))^\top C_e^{-1}(y_{obs} - f(\theta))\right). \quad (6)$$

Substituting Equations (5) and (6) into Equation (4) and using the definition of a weighted norm results in the parameter posterior distribution

$$\pi(\theta|y_{obs}) \propto \exp\left(-\frac{1}{2}(\|y_{obs} - f(\theta)\|_{C_e^{-1}}^2 + \|\theta - \theta_0\|_{C_{pr}^{-1}}^2)\right). \quad (7)$$

The point in parameter space that maximises the posterior density function is known as the *maximum a posteriori* (MAP) estimate, and is given by

$$\theta_{MAP} = \arg \min_{\theta \in \mathbb{R}^n} \{ \|y_{obs} - f(\theta)\|_{C_e^{-1}}^2 + \|\theta - \theta_0\|_{C_{pr}^{-1}}^2 \}, \quad (8)$$

as outlined in Power et al. [15]. This is then used as the calibrated parameter vector when the forward model is run.

Uncertainty quantification is often performed by sampling from the parameter posterior distribution and running these samples through the forward and predictive models.

3.2 Data-Space Inversion

As previously mentioned, calibration consumes the most computational resources and time due to a complex optimisation problem needing to be solved. It is often the case that the parameters themselves are not of direct interest, so DSI marginalises over the parameters focusing on the prior and posterior predictive densities.

Following Power et al. [15], $i = 1, 2, \dots, N$ prior parameter realisations are simulated $\theta^{(i)}$ which are run through both the forward model f and the predictive model f_{pred} . This gives an ensemble of parameters and model runs $(\theta^{(i)}, y_{obs}^{(i)}, y_{pred}^{(i)})$ where $y_{obs}^{(i)} = f(\theta^{(i)}) + \varepsilon^{(i)}$ and $y_{pred}^{(i)} = f_{pred}(\theta^{(i)})$.

Ultimately, DSI attempts to define a conditional posterior predictive distribution for the predicted data given the measurement data that have been observed, i.e., $y_{pred}|y_{obs}$. By marginalising over the parameters, a mapping from observations to predictions without the need for the explicit computation of model parameters can be generated. The sample means for both the observations and the predictions are

$$\bar{y}_{obs} = \frac{1}{N} \sum_{i=1}^N y_{obs}^{(i)}, \quad \bar{y}_{pred} = \frac{1}{N} \sum_{i=1}^N y_{pred}^{(i)}, \quad (9)$$

and the sample covariance matrix is given by

$$\bar{C} = \frac{1}{N-1} C C^\top = \begin{bmatrix} \bar{C}_{y_{obs} y_{obs}} & \bar{C}_{y_{obs} y_{pred}} \\ \bar{C}_{y_{pred} y_{obs}} & \bar{C}_{y_{pred} y_{pred}} \end{bmatrix}, \quad (10)$$

where $C = [\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_k]$ and $\mathbf{c}_j = [y_{obs}^{(j)} \ y_{pred}^{(j)}]^\top - [\bar{y}_{obs} \ \bar{y}_{pred}]^\top$ [15, 21, 22].

The expectation and covariance of the distribution for the predicted data conditioned on the observed data are shown in Equation (11) [15, 22],

$$\mathbb{E}[y_{pred}|y_{obs}] = \bar{y}_{pred} + C_{y_{pred} y_{obs}} C_{y_{obs} y_{obs}}^{-1} (y_{obs} - \bar{y}_{obs}), \quad C_{y_{pred}|y_{obs}} = C_{y_{pred}} - C_{y_{pred} y_{obs}} C_{y_{obs} y_{obs}}^{-1} C_{y_{obs} y_{pred}}. \quad (11)$$

Samples can then be randomly generated from this posterior distribution by computing

$$y_{pred}^s = \mathbb{E}[y_{pred}|y_{obs}] + Lr, \quad (12)$$

where $L L^\top = C_{y_{pred}|y_{obs}}$ and r is sampled from a Gaussian distribution $r \sim \mathcal{N}(0, 1)$ [15, 21].

3.3 Geothermal Models

The conceptual models and modelling methodology are discussed for both of the geothermal models under consideration.

3.3.1 Lumped Parameter Model

The LPM is based on the Ngatamariki geothermal reservoir located in the Taupō Volcanic Zone as outlined by Fradkin et al. [6], the conceptual model of which can be seen in Figure 1. Modelling the reservoir using a LPM allows the governing equations of state to be simplified resulting in an ordinary differential equation (ODE) that describes the pressure in the reservoir over time given data about the rate of injection of fluid into the system. The most significant assumption is that the reservoir is treated as one combined mass whose pressure changes only in response to the injection or extraction of fluid to the system. The ODE used to model the pressure is

$$\frac{dP}{dt} = aq - b(P - P_0) - c\frac{dq}{dt}, \quad (13)$$

where a , b , and c are the model parameters corresponding to the injection coefficient, recharge coefficient, and slow drainage coefficient respectively. Further, P is the pressure in the reservoir, q is the rate of injection of fluid into the reservoir, P_0 is the initial pressure of the system, and t is the time, the domain of which is 35 days.

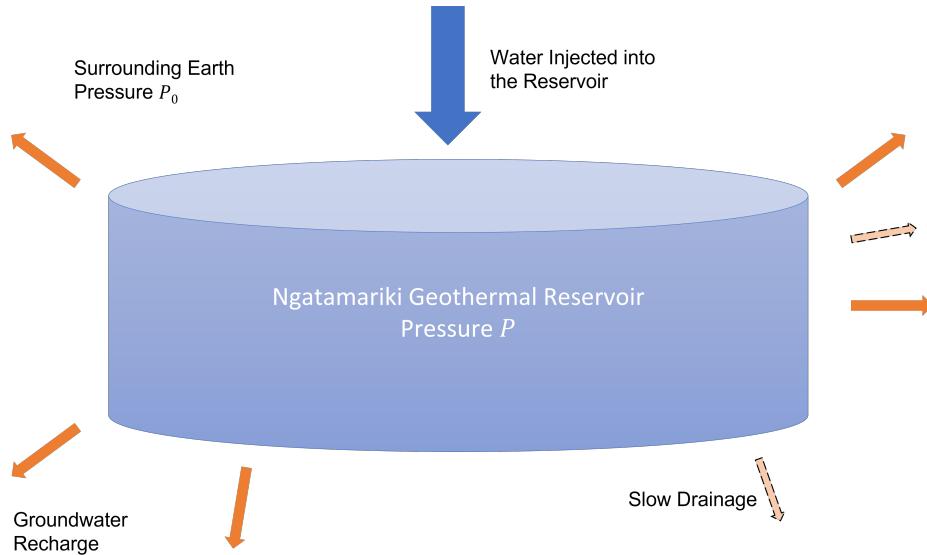


Figure 1: Simplified Conceptual Model of the Geothermal Lumped Parameter Model

Solving this ODE requires a numerical solver as there is no analytic solution without making additional simplifications or assumptions. Although the rate of injection into the reservoir is known through measured data, it evolves over time making it difficult for standard ODE solvers. Furthermore, the times at which data for the rate of injection were measured are not the same as the time points at which the pressure is solved, thus requiring interpolation adding complexity. Therefore, a custom implementation of the classic Runge-Kutta (RK4) method is used to solve this ODE.

3.3.2 3D Synthetic Model

To model a realistic geothermal reservoir, a synthetic three-dimensional model outlined in detail by Renaud et al. [4] was used. The conceptual model for the system can be seen in Figure 2. The domain of this model contains 13,864 blocks, with these being more fine towards the centre of the geometry. There are 103 different rock types each with their own x-, y-, and z-permeabilities,

and 10 deep upflows of varying strengths resulting in 319 model parameters. The entire domain represents an area of approximately 13 by 15 kilometres [4].

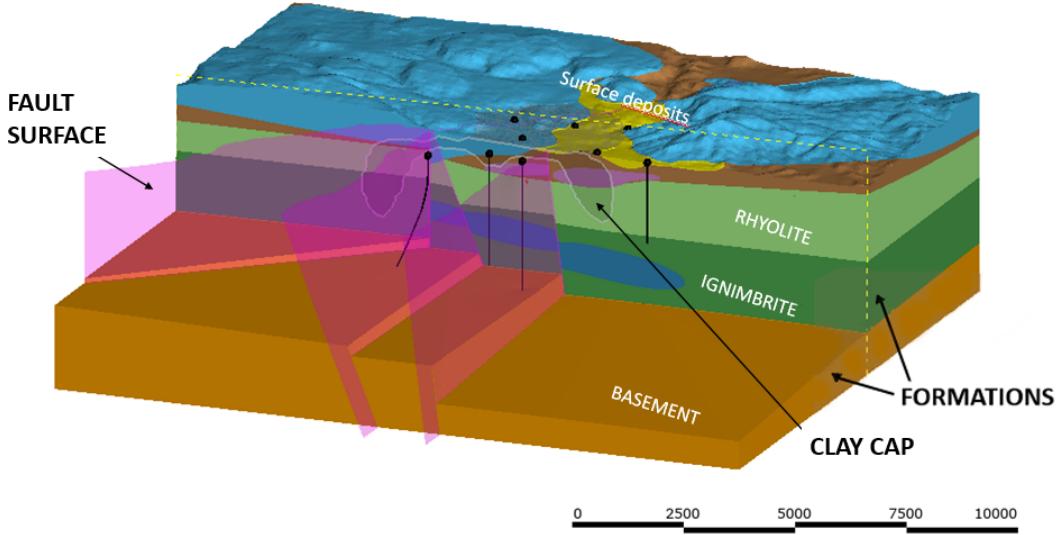


Figure 2: Conceptual Model of the 3D Geothermal System [4]

As this model is synthetic, it is trivial to compute known data values. Therefore, two model scenarios were run: the first sampled 1,000 parameter values randomly from their prior distributions and ran the model the same number of times to create the ensemble of model realisations, and the second ran the model using known calibrated parameter values to provide the observed data on which the data is conditioned. Of the 1,000 prior simulations, 905 converged resulting in the ensemble used for conditioning containing the same number of observations.

3.3.3 Waiwera

These model simulations were performed using Waiwera [23], a parallel, open-source geothermal flow simulator. Waiwera is a numerical simulator developed at the University of Auckland's Geothermal Institute which has benefit over existing geothermal modelling software in that it exploits parallel processing making running the simulations more efficient. The governing equations used for the geothermal reservoir model discussed in Section 4.3 involve time-dependent equations for the conservation of mass and energy flow over the simulation domain [15, 23].

3.4 *ccandu*

To perform the linear algebra required for DSI, a Python package *ccandu* was used [24]. This software is currently being developed by the Modelling Uncertainty and Data Group at the University of Auckland for the analysis of conditional random variables and their uncertainty. The underlying implementation of *ccandu* involves the Gaussian processes detailed in Section 3.2, and the use of conditional random variables (CRV) provides support for conditioning, joining, and marginalising [24], all of which are required in the methodology of DSI.

For the purposes of applying DSI to the set of models, one class method was particularly useful in defining the conditional random variable, i.e., the posterior predictive distribution for y_{pred} given y_{obs} : `linear_sample_approximation()`. Slicing the prior ensemble into those indices which had

been observed and those which were to be predicted, `linear_sample_approximation()` takes as input these two arrays and conditions the predicted data measurements with the observed data. The resulting object is a CRV corresponding to the conditional posterior predictive distribution.

3.5 Approximate Bayesian Computation

ABC is an algorithm that uses sampling from the prior model to approximate a posterior distribution [19]. Multiple parameters are sampled from the prior parameter distribution which are run through the forward model to give data \hat{D} [19]. One benefit of performing DSI beforehand is that this ensemble has already been simulated, so there is no need to recompute it, reducing computational expense. Conditioning involves comparing the simulations with the observed data D and accepting those that fall within a defined tolerance under a certain distance measurement. If d is the distance measure for comparing two data sets and ϵ is the tolerance, the simulations that will be accepted are those where

$$d(\hat{D}, D) \leq \epsilon. \quad (14)$$

An alternate method of conditioning the data defines the tolerance in a slightly different manner. Instead of considering a specified distance at which model realisations will be rejected, the tolerance refers to the proportion of simulations that are accepted. For example, if $\epsilon = 0.05$, 5% of the ensemble realisations will be accepted and form the posterior distribution.

This second definition has been used in this paper. Furthermore, the choice of distance measure employed is the sum of squared distances between the simulations and the observed data yielding

$$\|\hat{D} - D\|_2^2, \quad (15)$$

as the computed distance, of which ϵN simulations will form the posterior predictive distribution.

While ABC is traditionally used to do standard Bayesian inference, a new method called ABCDSI can combine the ideas of both ABC and DSI. Starting from the prior ensemble of realisations that is common to both methods, this ensemble can be conditioned on the data using the same methods as traditional ABC, and marginalising out the parameters as in DSI. This then gives the ABCDSI posterior predictive distribution with its own quantifiable uncertainty, from which inferences can be made. This is a novel approach that has not yet currently been considered in the literature.

4 Results

The implementation of DSI on three different models is outlined along with the resulting predictions and their uncertainty to access the the viability of the approach on models of varying complexity. These models are a polynomial function, a geothermal LPM, and a three-dimensional synthetic geothermal model based on existing geothermal fields in New Zealand. A variation of DSI is explored in normalising the data before generating the conditional posterior distribution, and a comparison to ABC is presented to examine the advantages of both approaches.

4.1 Toy Polynomial Model

Initially, a polynomial model is used to introduce the concept of DSI and to ensure that the approach gives the expected results. A first-order polynomial is considered

$$y(x; \theta) = a + bx + \varepsilon, \quad (16)$$

where $\theta = [a \ b]$, $a, b \sim \mathcal{N}(1, 1)$, and $\varepsilon \sim \mathcal{N}(0, 0.2)$ in the case where observation error, or noise, is explicitly added. 50 prior realisations are simulated by sampling randomly from these prior parameter distributions and running them through the forward model, and these are conditioned on the observed data using *ccandu*. From the computed posterior predictive distribution, a further 50 posterior predictions are simulated. The results of implementing DSI on this first-order polynomial can be seen in Figure 3. Two cases are shown, one without observation error and one with

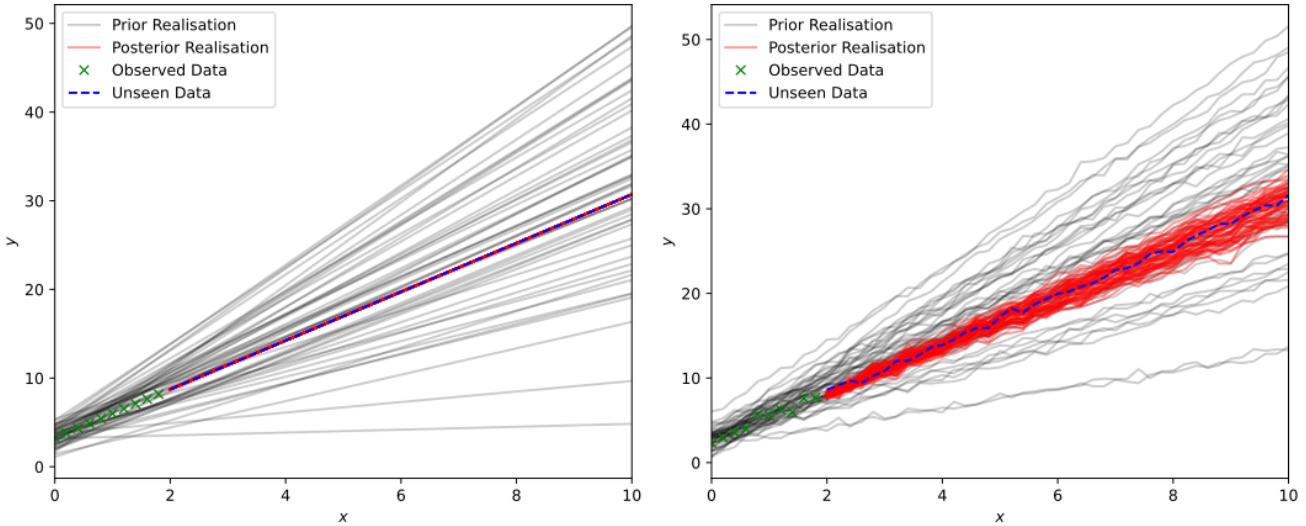


Figure 3: DSI on a First-Order Polynomial without Noise (Left) and with Noise (Right)

observation error added explicitly. Without the introduction of observation error, the posterior samples are able to exactly predict the true model as seen by the 50 posterior predictions having no variance and being identical in value to the unseen, true measurements. Adding observation error, or noise, to the ensemble adds variance to the predictions, however, the uncertainty of the posterior realisations is greatly reduced from that of the prior ensemble. The uncertainty in the posterior predictions also captures the unseen data in its entirety suggesting that the predictions made from the posterior distribution are accurate.

The polynomial model can be generalised and the results of implementing DSI on a second- and third-order polynomial are shown in Figure 4. In both of these cases, the observation error has a Gaussian distribution $\varepsilon \sim \mathcal{N}(0, 0.2)$, the same as that in Figure 3. However, it is not as noticeable due to the differing scales, the observation error being of order $\mathcal{O}(10^{-1})$ whereas the dependent variable y is of order $\mathcal{O}(10^2 - 10^3)$. Similarly, the posterior distribution accurately recaptures the unseen data, and the variance of posterior predictions is reduced from the prior ensemble.

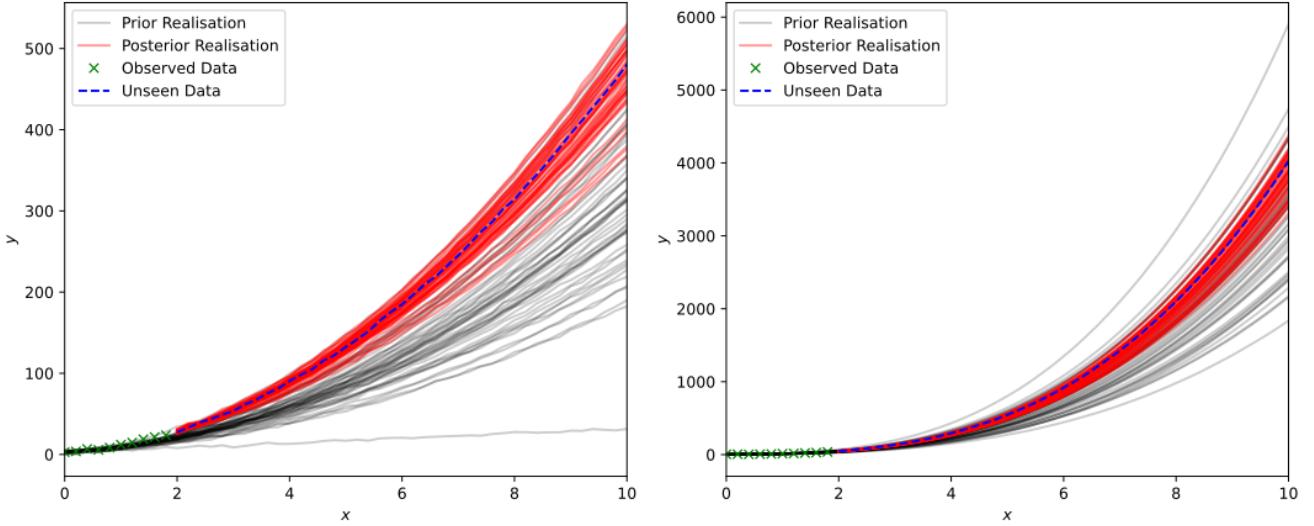


Figure 4: DSI on a Second-Order Polynomial (Left) and a Third-Order Polynomial (Right)

4.2 Geothermal Lumped Parameter Model

As an intermediary between the simple polynomial model and a realistic three-dimensional geothermal model, DSI is implemented on a geothermal LPM based on the Ngatamariki geothermal reservoir. As outlined in Section 3.3.1, the ODE model describing the system is

$$\frac{dP}{dt} = aq - b(P - P_0) - c\frac{dq}{dt}. \quad (17)$$

The initial pressure of the reservoir P_0 and the three coefficients a , b , and c are all taken as unknown random variables, and assumed to be statistically independent. P_0 is included as a random variable as there is no guarantee that $P(t = 0)$ is a known data point, although it is in this model. The parameters have the Gaussian prior distributions

$$\begin{aligned} \pi(a) &= \mathcal{N}(9 \times 10^{-4}, 10^{-6}), \pi(b) = \mathcal{N}(7 \times 10^{-1}, 10^{-3}), \\ \pi(c) &= \mathcal{N}(9 \times 10^{-5}, 10^{-6}), \pi(P_0) = \mathcal{N}(2 \times 10^{-2}, 10^{-4}), \end{aligned} \quad (18)$$

where the mean of each parameter is the best-fitting value after calibrating the model to the measured data. As one of the key motivations for using DSI is to avoid calibration due to its large computational expense, this method is not recommended and may add some bias to the posterior predictions. Generally, the prior distributions are generated using domain knowledge of the system, however, since this is not available for the reservoir under consideration, calibration provides a reasonable alternative for a model with relatively few parameters. The variance of

each parameter is selected to be approximately two orders of magnitude below that of the mean. Exploration into the behaviour of the injection coefficient a shows that this parameter has the largest influence on the output of the model, often causing non-physical results such as negative pressure. For this reason, the variance of a is approximately three orders of magnitude smaller than its mean to reduce the number of unreasonable prior simulations. A small amount of observation error is also added to the solution for pressure, i.e., $P_{obs} = P(t; \theta) + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, 10^{-4})$.

To create the prior ensemble, 100 realisations are simulated by sampling parameters randomly from the prior distributions given in Equation (18), and solving the forward model for pressure using the custom implementation of a RK4 numerical solver. This ensemble is then conditioned on the observed data in *ccandu* to compute the posterior distribution from which 100 predictions are made.

The results of implementing DSI on this geothermal LPM can be seen in Figure 5. As expected, random sampling of the parameters from their prior distributions produces some non-physical simulations where the pressure becomes negative. However, after conditioning the prior ensemble on the observed data, the posterior predictive distribution ensures that virtually all of the pressure values are physically reasonable. A few posterior realisations give negative pressures at $t = 26$ which could warrant further investigation.

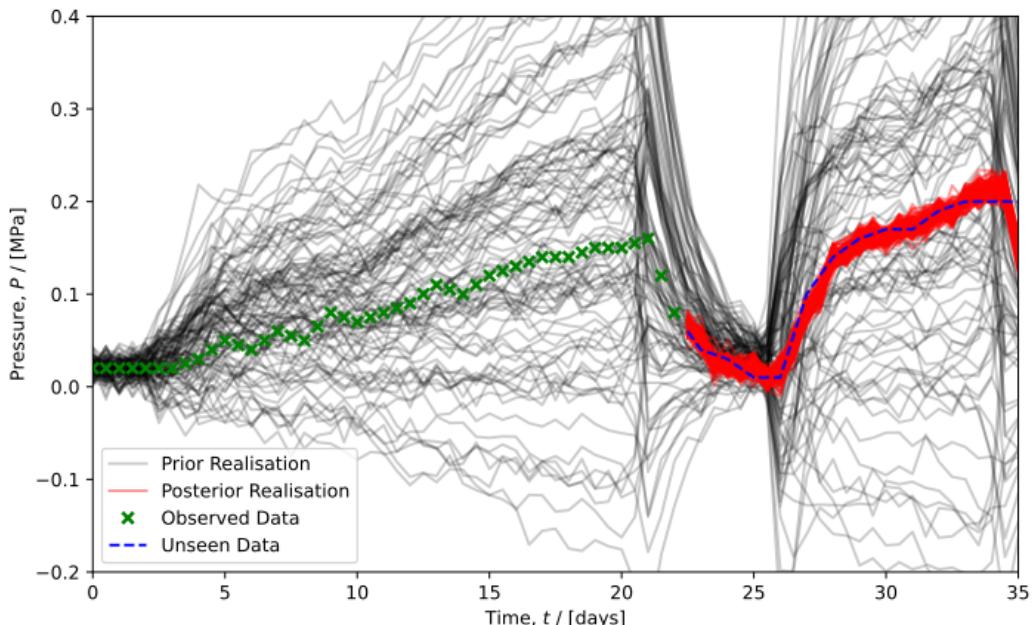


Figure 5: DSI on a Geothermal Lumped Parameter Model

The pressure at the end of the time domain exhibits strange behaviour with both the prior simulations and posterior predictions deviating from the unseen pressure measurements, but this is most likely due to the boundary conditions in the RK4 solver and the need to interpolate data adding some error to the solution of pressure in the reservoir.

Due to the increased complexity and sensitivity of the model to changes in the data, particularly the rate of injection, the accuracy and variability of results depend significantly on the quantity of observed data on which the prior ensemble is conditioned. As seen in Figure 6, conditioning on less than half of the data that is used in Figure 5 has a severe impact on the accuracy and variability of the results. The posterior predictive distribution fails to capture, or barely captures,

the true data at several points in the time domain, and the uncertainty is much larger as shown by increased variation in the posterior predictions. Clearly, there is a dependence on the quantity of observed data which affects the accuracy and uncertainty of predictions made. Further research could focus on this relationship, exploring the requirements for DSI to make adequate predictions and constrain the posterior distribution.

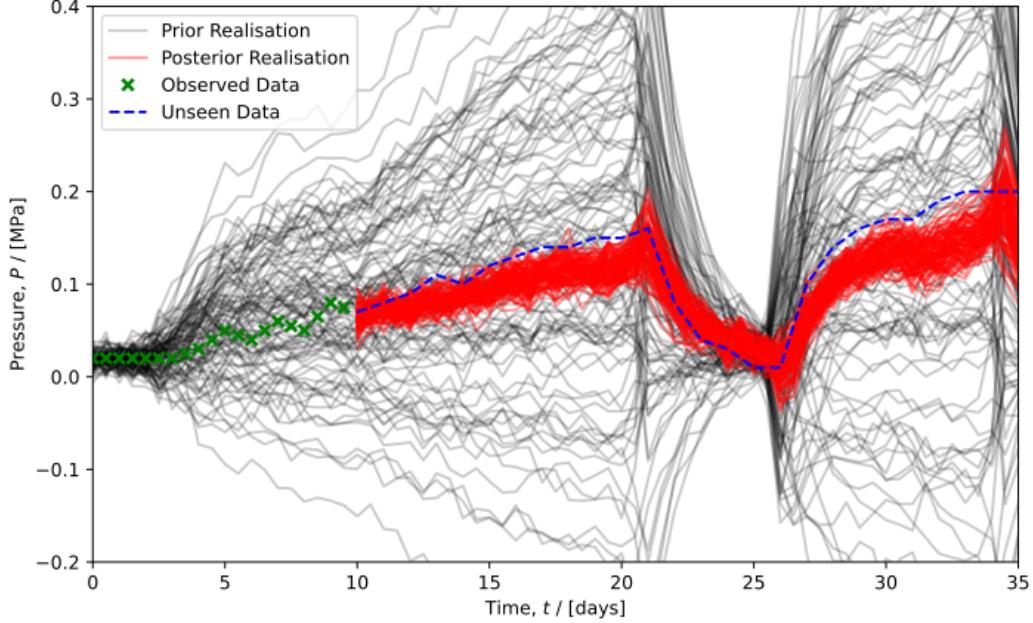


Figure 6: The Effect on Accuracy and Variability of Fewer Observations

4.3 Realistic Geothermal Model

The use of DSI on the synthetic geothermal model is presented in two scenarios. The system in its natural state is used to make spatial predictions about the temperature down two wells, and, in the transient case, predictions about the temperature at the feedzone of wells are made over time.

For both scenarios, the model parameters are consistent with the same prior distributions used in the the natural state and production analysis. As mentioned in Section 3.3.2, there are 319 model parameters corresponding the to the x-, y-, and z-subsurface permeabilities of each of the 103 rock types, and the strengths of 10 deep upflows. All of these are given a Gaussian distribution, although not all of the priors are independent of each other. Following Power et al. [15], the horizontal permeabilities, i.e., the x- and y-permeabilities, are assigned a Pearson's correlation coefficient of 0.5 to the vertical permeability, i.e., the z-permeability. Similarly, a Pearson's correlation coefficient of 0.8 is assigned to the correlation between the x- and y-permeabilities. The strengths of the 10 deep upflows are given Gaussian distributions with means equal to the strengths of the true, calibrated values. Their variances are defined to be 10% of these mean values, i.e., $s_i \sim \mathcal{N}(\mu_i, 0.1\mu_i)$, where s_i is the strength of upflow i with a mean strength μ_i . The subsurface permeabilities of each rock type are independent of each other, as are each of the 10 deep upflows.

4.3.1 Natural State

The natural state refers to the geothermal reservoir before any production has occurred [1]. Here, the quantities of interest are independent of time allowing for the DSI approach to be implemented to make predictions in space without the temperature in the reservoir evolving over time.

Eight wells spaced in a circular arrangement near the centre of the spatial domain are selected to be the wells at which the temperature is observed (Wells 1-8). To explore the capabilities of DSI, predictions are made at two well locations, one inside the circle of known wells (Well 9) and one outside (Well 10), as shown in Figure 7. By examining the predictions made in both cases,

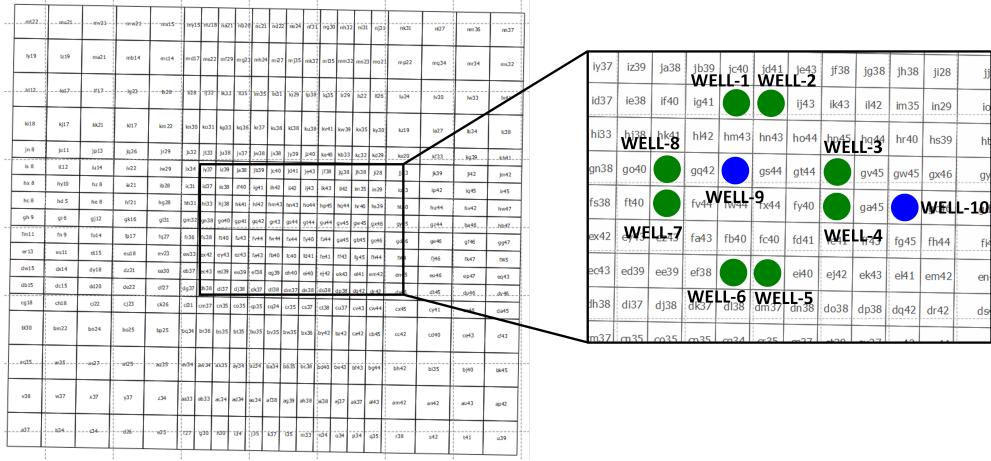


Figure 7: Visualisation of the Observation and Prediction Wells in the Spatial Domain

an illustration of the extent to which DSI can be used to extrapolate outside the observed data is provided. Due to the topology of the geothermal system, not all wells are of the same depth, however, this has no effect on the ability of DSI to make predictions.

The prior ensemble of temperature profiles down each of the eight observation wells can be seen in Figure 8 along with the true data on which this ensemble is conditioned. There is significant variability in the distribution of prior simulations, but, in the case of each well, the prior distribution contains the observed temperature profile.

The 100 posterior predictions for the temperature profile down each of the two prediction wells are shown in Figure 9. For the well on the interior of the observation wells, the temperature is predicted reasonably accurately at very shallow depths and from 1,000 metres below the surface with the posterior distribution containing the true temperatures. At shallow depths approximately 200 to 1,000 metres below the surface, the posterior distribution overestimates the temperature by between 20 and 50°C. As expected, the temperature predictions for the well in which data are extrapolated are not as accurate. The posterior distribution consistently underestimates the temperature at all depths below 100 metres suggesting that DSI struggles to extrapolate data outside of the observed measurements. However, the method does capture the correct characteristic shape of the temperature profile despite the fact that it predicts cooler than observed temperatures. Since DSI has only been implemented on one set of wells, analysis into the number and location of observation wells, and how that affects the accuracy of the posterior predictions, would provide insight into the applicability of the method in various scenarios.

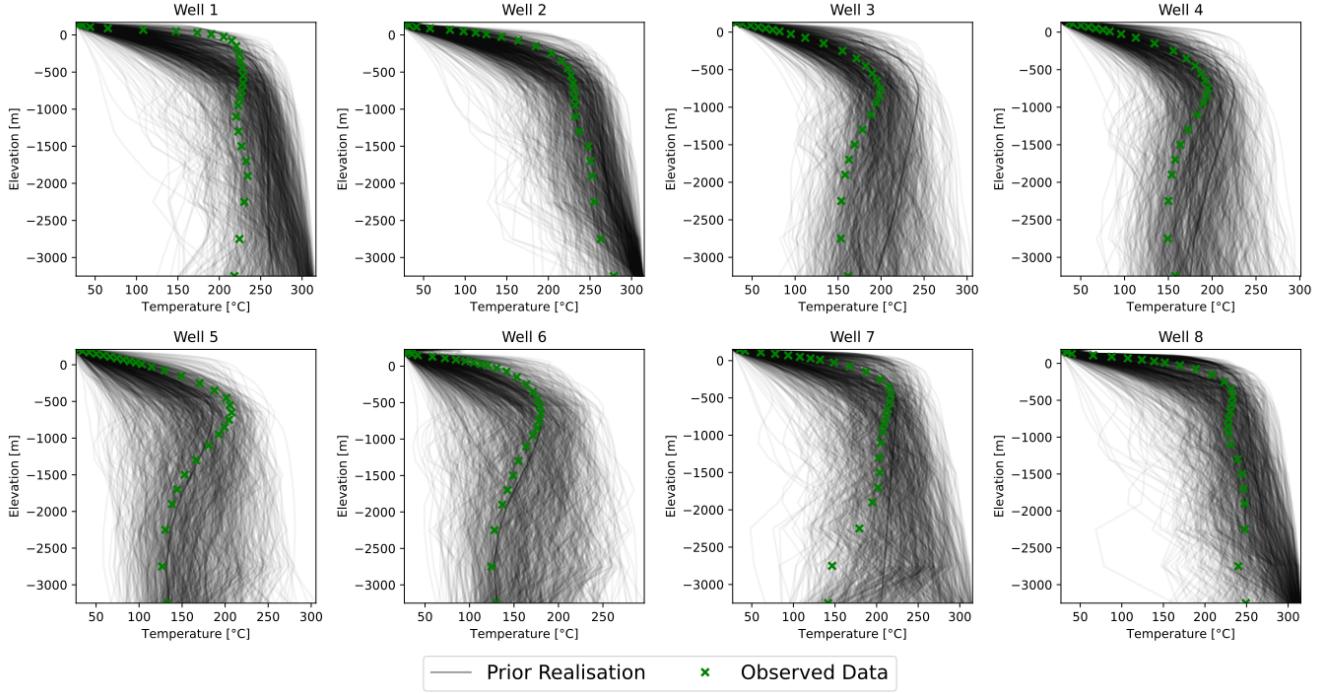


Figure 8: Prior Ensemble of Model Realisations and Observed Data for the Observation Wells

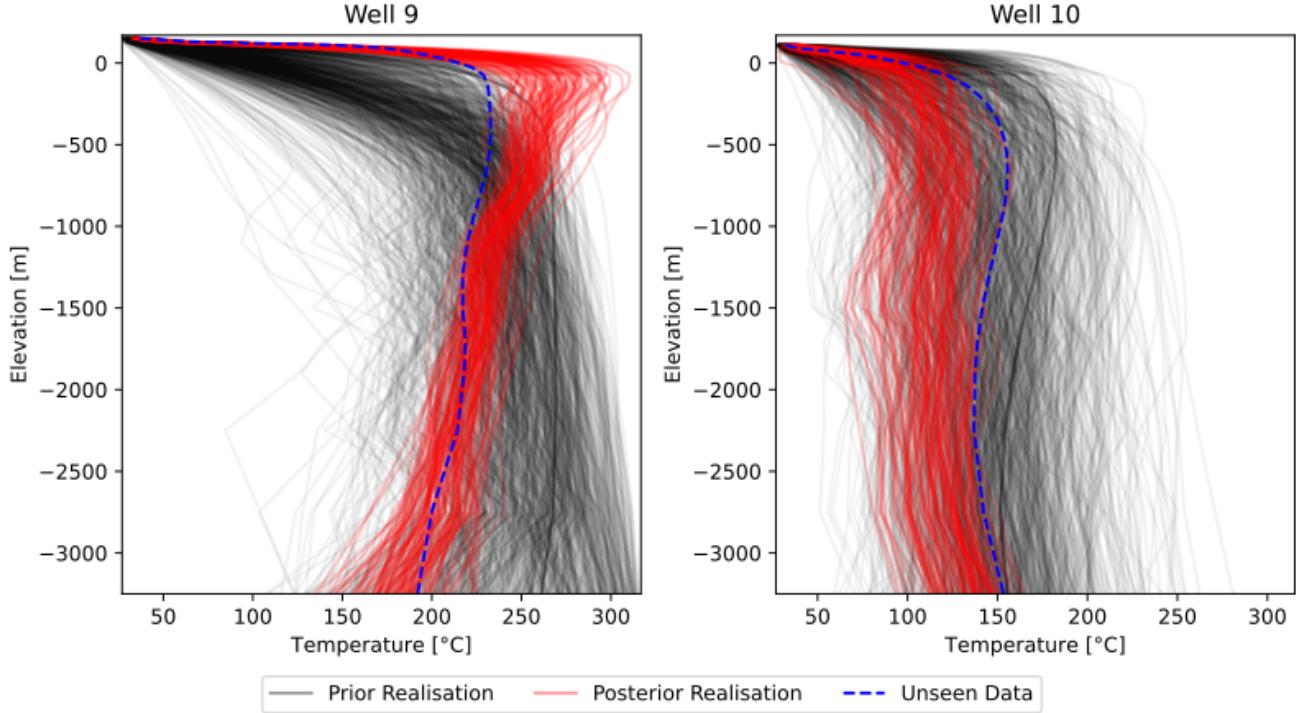


Figure 9: Spatial Posterior Distribution of Temperature Predictions for an Interpolated Well (Left) and Extrapolated Well (Right) in the Natural State

4.3.2 Production

To take energy from a geothermal system, production wells extract hot fluid from which the steam turns large turbines to make electricity. Predicting the temperature down production wells is vital

for the efficient operation of a geothermal power plant, so DSI is used to predict the temperature at the feedzones of a production well over a period of three years. The feedzone consists of blocks in the model whose location is constant over time resulting in an independence of position.

Well 2, shown in Figure 10, is a production well near the centre of the spatial domain with a feedzone comprising three blocks in the numerical model. Three years of production in the geothermal reservoir, between 2017 and 2020, are taken as observed temperature measurements, and predictions of the temperature are made for these three blocks of the next three years to 2023.

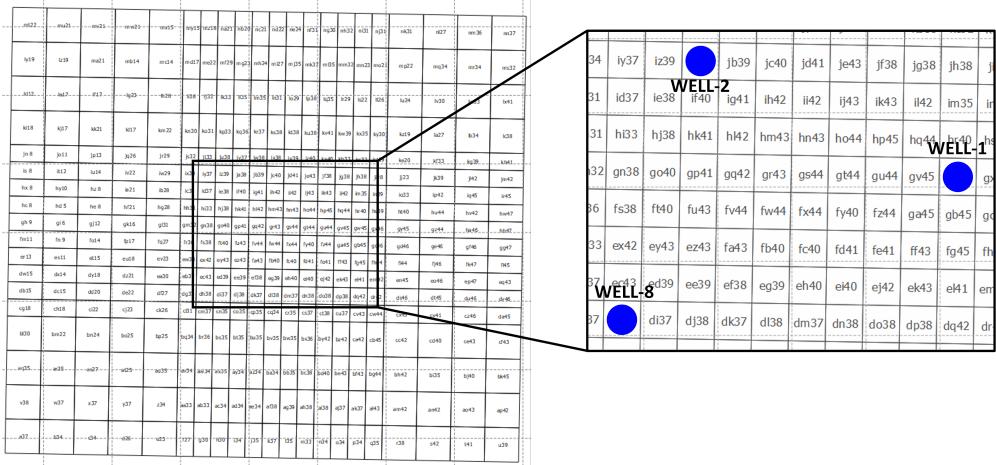


Figure 10: Visualisation of the Feedzone for Well 2 in the Spatial Domain

The prior ensemble of temperature measurements over the six year time domain can be visualised in Figure 11, along with the resulting posterior predictive distribution from 100 predictions after implementing DSI. In each of the three wells, the predictive distribution contains the true, unobserved temperatures, and the variance of these posterior distributions is greatly reduced from that of the prior ensemble. As the characteristic shape of the prior ensemble gets less complex, the predictions are more accurate and are made with less variability. This allows for predictions to be made with greater certainty, however, in all cases, DSI is able to accurately forecast the temperature in the feedzone of a production well over the time domain of interest.

For contrast, Wells 1 and 8, as shown in Figure 10, are injection and monitor wells respectively. Injection wells in a geothermal reservoir are wells in which cool fluid is reinjected into the system. They may be unused production wells that had been drilled but were subsequently not hot enough, yet still have good permeability in the rock structure. Monitor wells are used for the purpose of measuring properties of the geothermal reservoir. In both cases, the temperature of the feedzone to the wells is not expected to vary significantly over time, so DSI should be able to make accurate predictions with little uncertainty in the posterior distribution. The results of implementing DSI to make predictions of the temperature in the feedzones of Wells 1 and 8 for the next three years can be seen in Figures 12 and 13 respectively. The temperature in the injection well barely evolves over time, and the temperature in the monitor well does not change over the six year period. This leads to DSI making highly accurate predictions with very little posterior uncertainty.

4.4 Approximate Bayesian Computation

To critically analyse the performance of DSI, ABC is implemented on the geothermal LPM introduced in Section 3.3.1 providing a comparison between the two methods. ABC does not require

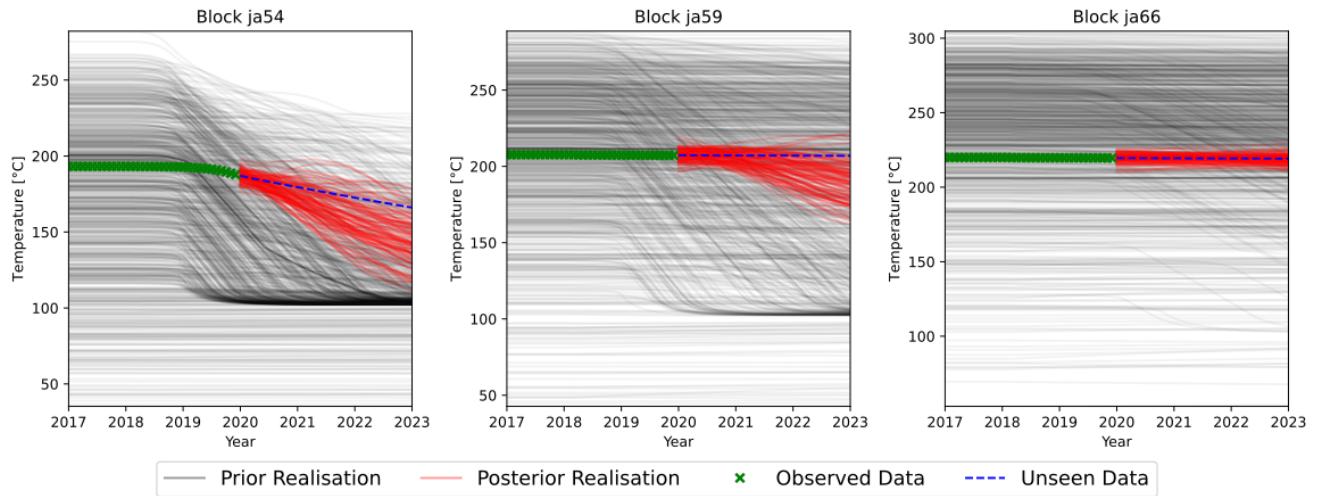


Figure 11: Transient Posterior Distribution of Temperature Predictions at the Feedzone of Well 2

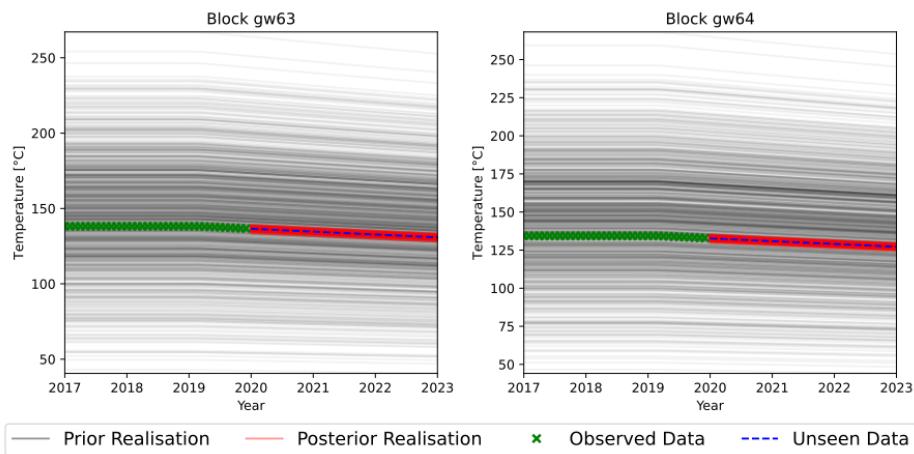


Figure 12: Temperature Predictions at the Feedzone of an Injection Well

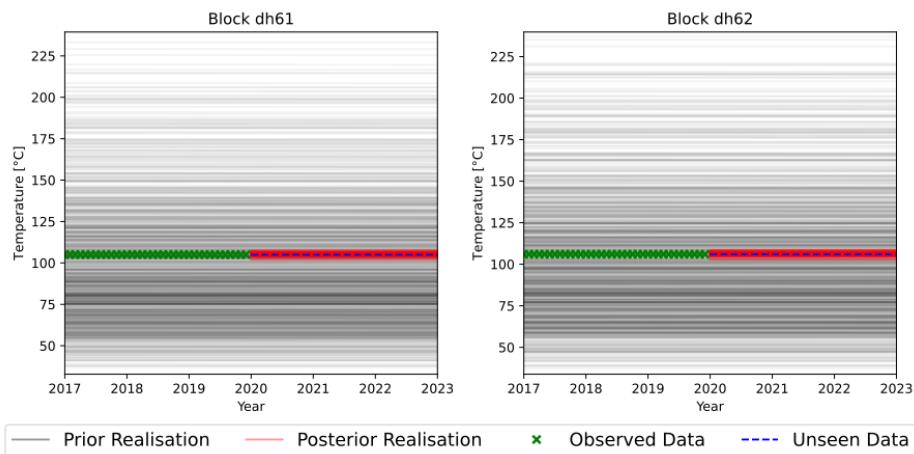


Figure 13: Temperature Predictions at the Feedzone of a Monitor Well

the calibration of model parameters and uses the same prior ensemble of model realisations as DSI which makes for a fair comparison in the accuracy of posterior predictions and their uncertainty.

The results of implementing ABC with a tolerance of $\epsilon = 0.01$ are shown in Figure 14 and DSI on the same prior ensemble can be seen in Figure 15 for comparison. 10,000 prior model realisations are simulated, of which 100 form the posterior distribution. One of the most significant differences is that ABC approximates the observed data whereas DSI treats it as known. The posterior predictive distribution accurately captures both the observed data and the true, unseen data, however, the predictions of the unseen data are more variable when ABC is used.

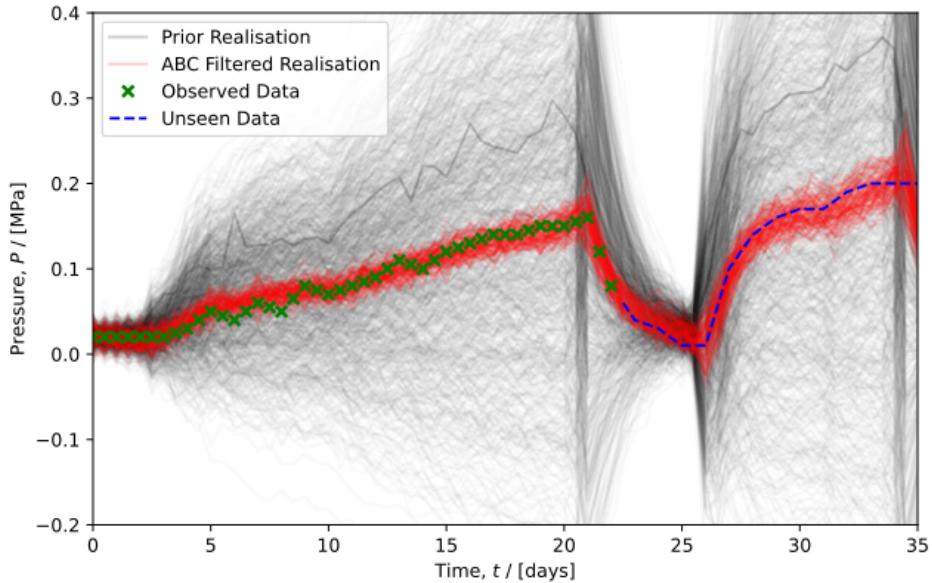


Figure 14: ABC on a Geothermal Lumped Parameter Model

As in Section 4.2, analysis into the dependence on the quantity of observed data is performed to see how well ABC can make predictions when fewer data are known. The results are shown in Figure 16, again with a comparison to DSI on the same prior ensemble in Figure 17. ABC captures the true data better than DSI in this case, albeit only a slight advantage in accuracy. There is larger uncertainty in the posterior prediction when conditioning of fewer observation measurements, but, as in the previous analysis, the predictions made by DSI are more certain as seen in the smaller amount of variation. This suggests that in certain scenarios, there is a trade-off between accuracy and uncertainty in the predictions made. DSI gives predictions with more certainty, but the posterior predictive distribution can fail to capture the true data accurately in some cases. ABC appears to make consistently accurate predictions at the cost of a less certain posterior distribution. However, as both DSI and ABC use the same prior ensemble of model realisations, it comes at very little additional computational expense to implement both approaches where the benefits of each can be exploited.

4.5 Normality

An important assumption of DSI is that the prior and posterior predictive distributions need to be well-approximated by a Gaussian distribution. All of the prior distributions throughout this

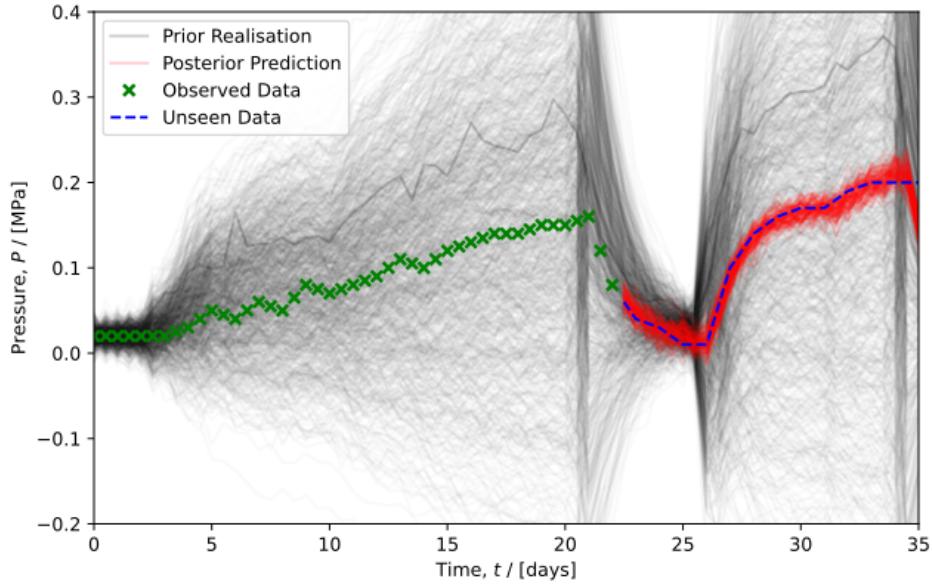


Figure 15: DSI on the Same Model for Comparison

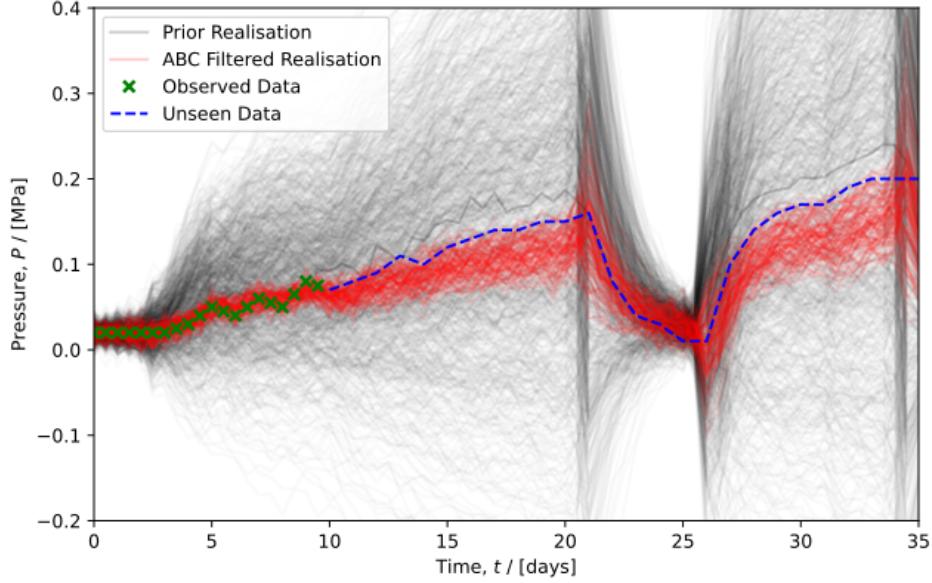


Figure 16: ABC on a Geothermal Lumped Parameter Model

analysis are Gaussian to respect this assumption, however, the effect of normalising the prior ensemble of model realisations is now explored. As the natural state spatial predictions discussed in Section 4.3.1 appear to have the worst performance, this is the model that is used.

The prior ensemble of model simulations that DSI uses to generate a conditional posterior distribution is a matrix $\hat{D} \in \mathbb{R}^{N \times m}$ which can be thought of as containing N observations and m features. The interpretation of these features varies depending on the scenario, however, in this model, there is a feature for each block in each well being observed and predicted. A Yeo-Johnson

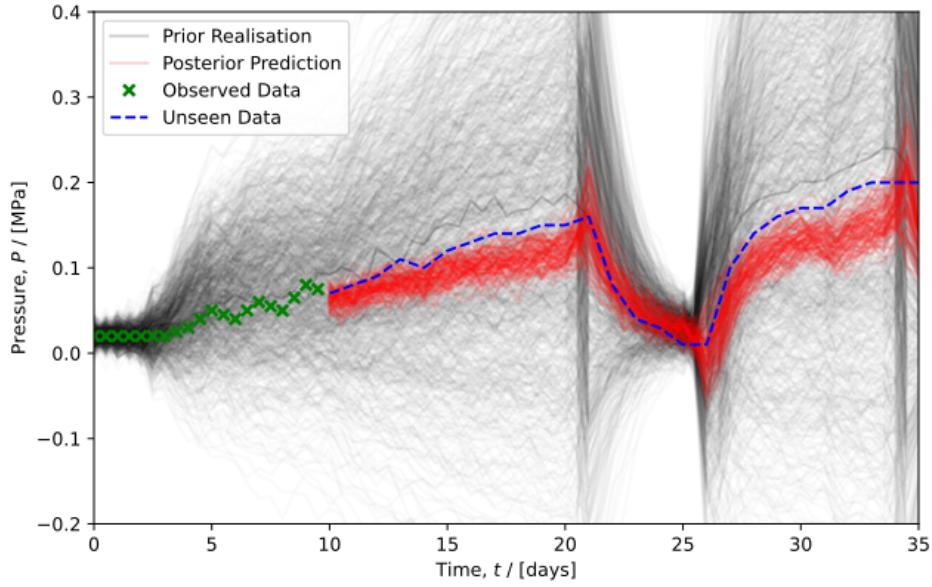


Figure 17: DSI on the Same Model for Comparison

transformation to normality is applied to each feature, a representative demonstration of which can be seen in Figure 18. Before the transformation, the distribution of temperatures for this

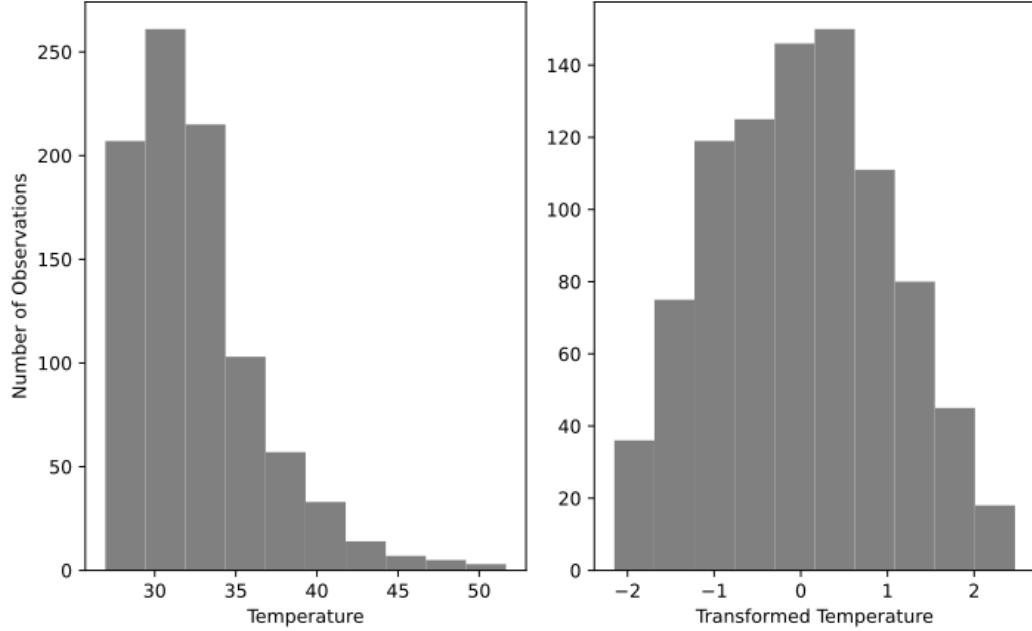


Figure 18: Visualisation of a Feature Before (Left) and After (Right) Transforming to Normality

feature is noticeably left-skewed. Applying a Yeo-Johnson transformation has caused the features of the prior ensemble to much more closely match a normal distribution as desired.

The results of implementing DSI on the transformed prior ensemble to generate a conditional posterior distribution are shown in Figure 19. Normalising the features of the prior ensemble has

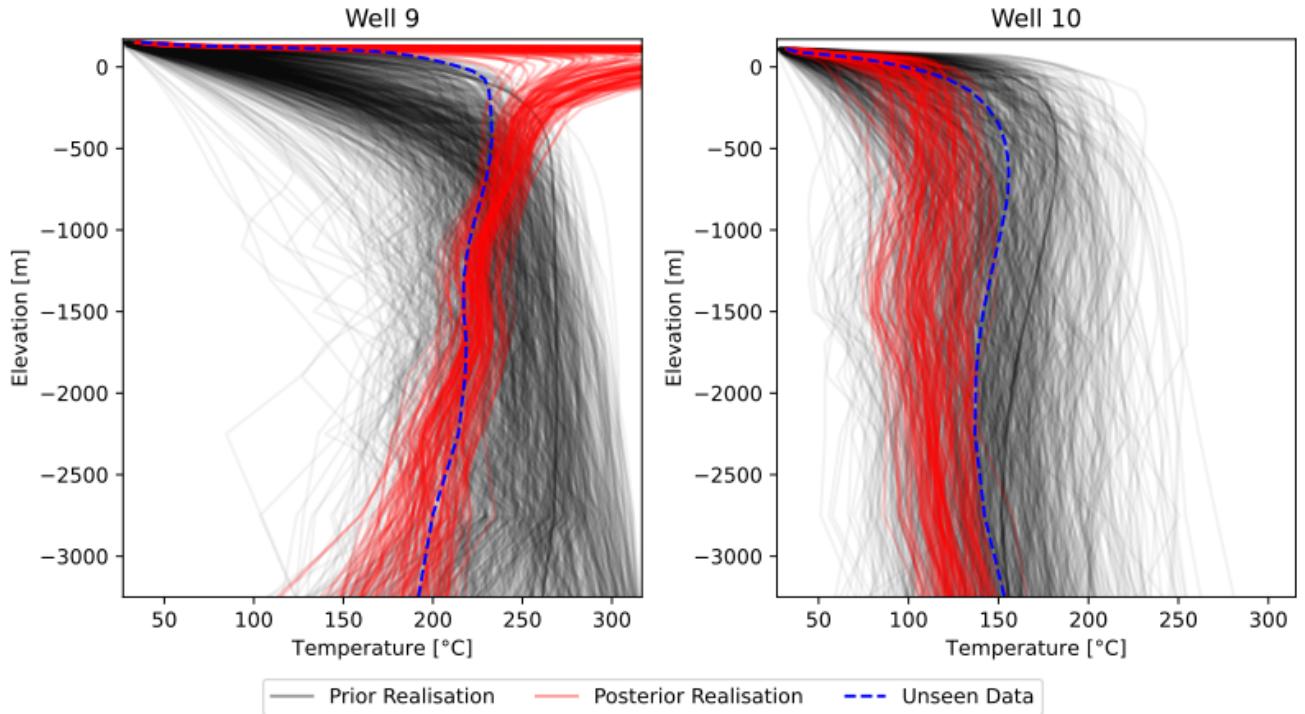


Figure 19: Posterior Distribution of Temperature Predictions after Transforming for Normality

a significant effect on the posterior predictive distribution. For Well 9, the distribution contains the true, unseen measurements for a larger proportion of the temperature profile, particularly between 500 and 1,000 metres in depth. However, from 0 to 500 metres, the inaccuracies previously established in Figure 9 are exaggerated resulting in significantly less accurate predictions in this section of the temperature profile. For the well at which data had to be extrapolated, Well 10, the effect of normalising the prior ensemble is not as evident. The posterior predictions still follow the same shape as the true data, although the distribution consistently underestimates the temperature by between 10 to 60°C. In both cases, there is no discernible difference in the variability in predictions with both wells exhibiting similar uncertainty to that in Figure 9. As this transformation has both benefits and consequences to its use, further research into when it provides increased accuracy of posterior predictions would be valuable.

5 Discussion

The implications of this research for the field of geothermal reservoir modelling is now discussed. A summary of the findings is presented, along with an evaluation of how the results fit into the current body of existing literature. How well the original research objectives have been met is also presented, and the limitations of DSI and how these could be used to inform future work into this area of research are critically evaluated.

5.1 Findings

Applying DSI to three models of varying complexity shows that it is a feasible alternative to traditional calibration-based approaches of modelling and making predictions about geothermal reservoirs. In most cases, the predictions are accurate, and with an uncertainty that is easily quantifiable and significantly reduced from that of the prior distribution.

As can be expected, the complexity of the model has an impact on the quality of predictions made by DSI. This complexity comes in the form of the dimensionality of the model, i.e., the number of model parameters, as well as the number of measurements made. The polynomial model discussed in Section 4.1 has two parameters and pressure measurements are made at 71 times throughout the temporal domain. In contrast, the three-dimensional geothermal model discussed in Section 4.3 contains 319 parameters, and temperature measurements are made in each of the 13,864 blocks at 73 times over a six year period. In the polynomial model, highly accurate predictions are made with little uncertainty, whereas the three-dimensional geothermal model exhibits some inaccuracies, particularly when making predictions in the spatial domain. These predictions are also less certain as shown by increased variability in the posterior predictive distributions. However, the results suggest that DSI can be used in certain scenarios to make accurate forecasts of the temperature and pressure in geothermal systems.

Additionally, the models are increasingly more non-linear in terms of the governing equations and model parameters which requires that approximations are made to estimate the posterior predictive distribution. A key assumption to DSI is that the prior and posterior predictive distributions are well-approximated by Gaussian distributions. While non-linear models might lead to approximately Gaussian data-space distributions, there is no guarantee that this will be the case. Therefore, since a large number of realistic models are non-linear, these results show that DSI can still make reasonable predictions despite the approximations necessary to estimate the posterior predictive distribution. Section 4.5 shows that a transformation of the prior ensemble of model realisations to be approximately Gaussian distributed does not always result in the predictions being more accurate. In fact, this transformation can cause the inaccuracies to be exaggerated.

It is found that ABC is a suitable alternative to DSI in that it makes accurate predictions, albeit with higher uncertainty. However, the main drawback of implementing ABC is that it requires a large number of prior model simulations to be run for accurate predictions to be made. The tolerance parameter used to accept simulations is analogous to the noise level under the assumption that the noise is uniformly distributed. Therefore, as the tolerance converges towards $\epsilon = 0$, the predictions made by ABC approach the exact solution rather than an approximation. For an appropriate number of posterior realisations to comprise the predictive distribution, the number of prior realisations needs to be large if the tolerance is to approach $\epsilon = 0$. Since a key objective of this research is to reduce the computational expense and time spent running models, ABC might not be a viable alternative to DSI, although it has been shown to give reasonable results.

5.2 Related Research

The findings of this research are consistent with those presented by Sun and Durlofsky [2] and Sun et al. [17] in that DSIs provides a more efficient method to generate posterior predictions. The ability to perform the model runs in parallel rather than serially, as would be required by a calibration-based approach, reduces the time necessary to make reasonable predictions. However, for realistic models this time is still not negligible.

The novelty of this research is the application of DSIs to a geothermal context, for which there is very little existing literature. Ideally, with continuing exploration of the approach in this context, DSIs could be implemented for the academic and commercial analysis of geothermal systems to ensure that the resource is effectively managed for an extended period of time.

Furthermore, this study introduces a new approach to DSIs based on ABC called ABCDSI. There is currently no existing literature on this novel method, providing the opportunity for further research to be performed which considers the benefits of ABCDSI relative to what has already been presented.

5.3 Limitations and Future Work

Throughout the analysis on the geothermal reservoir, a synthetic model designed to represent existing New Zealand geothermal fields and typical geophysical features has been used. As such, the data has not been measured directly from a real geothermal system. This has limitations on the accuracy and reliability of the results presented. Although the data should be representative of a true geothermal system, future research could take pressure and temperature measurements from a real reservoir to validate the conclusions drawn from this study.

While these results provide useful results for the purpose of making qualitative conclusions, there is currently no quantifiable comparison between the accuracy of the models or their uncertainty. A distance measure between the predictions and the true measurements would be reasonably simple to implement, and would allow different models, predictive approaches, and transformations to be impartially evaluated. Currently, this is limiting the quantifiability and comparability of the results, particularly if further research were to be undertaken into the implementation of DSIs on a geothermal reservoir.

A significant limitation is that, while DSIs might be less computationally and time intensive than traditional methods, the time taken to perform all of the model runs necessary to generate the prior ensemble is not insignificant. The 1,000 geothermal simulations analysed in this research took multiple days to run. As a larger number of prior realisations will generally result in more accurate predictions with less uncertainty, there is a trade-off between the quality of predictions and the resources required to make reasonable predictions. The prediction quality cannot increase indefinitely, so some exploration of the number of prior model simulations required to make predictions within a specified tolerance would be beneficial.

While numerically solving a geothermal model, the system is discretised into blocks of a substantial size. A small block could still be 50 metres in length, over which the measurements made could vary significantly. This research assumes that the temperature profiles of wells pass exactly through the centre of each block. In reality, wells will not pass through the centre of every block which limits the accuracy of the observed data and, therefore, the predictions made. This effect is more consequential when considering deviated wells which are not vertical and may run through many blocks diagonally. The use of interpolative methods to more precisely calculate the temperature or pressure at a particular location in the reservoir would mitigate this limitation, the effect

of which could be further studied.

From a geothermal perspective, there are certain constraints on the data that could be exploited by the predictive process. An example of this is that pressure measurements in the reservoir must be strictly non-negative. It would be of interest to see if the predictive ability of DSi is strengthened by imposing these additional conditions on the data. By filtering the prior ensemble of realisations or the posterior predictive distribution to disregard samples which contain non-physical results, the accuracy of the predictions could be evaluated to explore whether filtering is advantageous.

Finally, it is important to note that, while the analysis of the three-dimensional geothermal model only considered the temperature in the reservoir, this research could be repeated using pressure or enthalpy measurements. Depending on the behaviour of these quantities in the system, DSi might be able to make more accurate predictions than have been presented when considering the temperature of the reservoir. As previously mentioned, physical constraints on the pressure in the geothermal reservoir would allow for the filtering of unreasonable samples to potentially give better forecasts.

5.3.1 Modelling Approaches

There are three modelling factors that have been shown to have a significant effect on the accuracy and uncertainty of the posterior predictive distribution, each of which merits their own study.

The number of data points on which the prior distribution is conditioned can impact the quality of predictions, particularly when the underlying model is complex. In the context of geothermal reservoir modelling, the number of temporal observations relative to how distant the predictions are to be made or the number of observation wells will affect the ability to make good predictions both in space and in time. Analysis into the number of observations necessary to make predictions with a specified accuracy could provide limits on the applicability of DSi to ensure that any forecasts are as close to the truth as possible.

Similarly, the location and arrangement of observation wells could have an effect on the predictive ability of DSi. It has been shown that predictions in which the data has been extrapolated outside the observed spatial region can be less accurate than when the data is interpolated. Studying the impact of different arrangements of observations wells and where they are located relative to the prediction wells would be beneficial for forecasting.

As has already been discussed, the tolerance used to accept samples in ABC is closely related to the noise level of posterior predictions. The comparison of DSi to ABC defines the tolerance to be the proportion of prior realisations that are accepted which is different to the definition required for it to be considered the noise level. By considering the tolerance to be the difference between the data and the prediction, quantification of the uncertainty would be more rigorous. The effect of varying the tolerance on the accuracy and uncertainty of predictions could also be useful to understand.

5.3.2 Non-Invasive Modelling

Pre-drilling data can be used to determine where the best location for a production well might be. As this does not require test wells to be drilled, it is a relatively inexpensive process to gather this data. One way in which this pre-drilling data could be used to make predictions about the properties of a geothermal system is through geophysical features which provide information, specifically boundary conditions on the temperature in the reservoir. The subsurface clay cap is an example of a geophysical feature that is prominent in geothermal systems. Using magnetotelluric surveying, a

non-invasive exploration technique, the bottom of the clay cap can be approximated, at which the temperature is known to be roughly 200°C. This information can be treated as the observed data on which the prior ensemble of model realisations are conditioned by implementing DSI leading to an alternative method of making predictions in a reservoir without the need to perform expensive drilling. If this non-invasive technique were to work, it could have extensive implications on the management of geothermal systems, particularly to reducing costs when determining a site for a new production well. Research into the feasibility and accuracy of implementing DSI to make predictions without requiring observations from drilling would be advantageous, and could provide significant results for the future of geothermal reservoir management.

6 Conclusions

The main objective of this research was to provide an efficient alternative approach to making accurate predictions with quantifiable uncertainty. As traditional methods of prediction are time intensive due to the many serial model runs required during calibration, DSI is presented as a viable option. There is little existing literature focusing on the implementation of DSI in a geothermal context, so the research also extends the current body of literature in a way that is novel.

In most scenarios, DSI has the potential to make accurate predictions, both in space and in time, of the temperature in a geothermal reservoir with quantifiable uncertainty. However, there are situations in which the predictions made by a DSI approach are not as accurate as desired which poses the opportunity for further research into the field. The ability to include additional information in the modelling approach, such as a constraint on the pressure to ensure that physically feasible predictions are made, is a promising extension to this analysis.

Future work might include applying the methods discussed to predict the pressure in the three-dimensional geothermal reservoir rather than just the temperature, and exploring the use of DSI to make predictions using geophysical features of the system such as the clay cap. An understanding of the sensitivity of the method to the number of observations on which the data is conditioned, the location and arrangement of observation wells, and the transformation of prior data to an approximately Gaussian distribution all present potential further studies to develop a thorough understanding of the approach.

DSI has achieved its objective of providing more efficient predictions of a geothermal reservoir, and they are of a reasonable accuracy. By allowing the prior ensemble of model simulations to be run in parallel, the time and computational expense is significantly reduced from what would be required of traditional approaches. Alternatively, DSI allows for a larger number of model simulations to be run in the same time, or simulations of a more complex model. This has important implications for the future of geothermal reservoir modelling. The ability to define a model with greater resolution, and be able to make predictions within an appropriate time frame, presents an opportunity to more precisely model a geothermal system allowing more accurate predictions to be made. This has impacts on the management of geothermal systems as they rely on the forecasting of reservoir properties, making their operation more efficient. Models with increased definition would ordinarily require too much time and computational expense for them to be viable. But, DSI provides a feasible alternative that is statistically well-founded, reliable, and expands the breadth of current problems that can be explored efficiently using simulation modelling.

References

- [1] John O'Sullivan. *Geothermal Reservoir Modelling: Uses and Limitations*. Oct. 2021.
- [2] Wenyue Sun and Louis J Durlofsky. “A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems”. In: *Mathematical Geosciences* 49.6 (2017), pp. 679–715.
- [3] Jihoon Park and Jef Caers. “Direct forecasting of global and spatial model parameters from dynamic data”. In: *Computers & Geosciences* 143 (2020), p. 104567.
- [4] Theo Renaud et al. “Practical workflow for training geothermal reservoir modeling”. In: *Proceedings 43rd New Zealand Geothermal Workshop*. Wellington, Nov. 2021.
- [5] Stanford Geothermal Program. “Geothermal Reservoir Engineering Code Comparison Project”. In: *Proceedings Special Panel on Geothermal Model Intercomparison Study*. Stanford, California, Dec. 1980.
- [6] Larissa Ju Fradkin, Michael L Sorey, and Alex McNabb. “On identification and validation of some geothermal models”. In: *Water Resources Research* 17.4 (1981), pp. 929–936.
- [7] Gudmundur S. Bodvarsson, Karsten Pruess, and Marcelo J. Lippmann. “Modeling of geothermal systems”. In: *SPE 1985 California Regional Meeting*. Bakersfield, California, Mar. 1985.
- [8] Karsten Pruess. “Modeling of geothermal reservoirs: fundamental processes, computer simulation and field applications”. In: *Geothermics* 19.1 (1990), pp. 3–15.
- [9] John Burnell et al. “Geothermal supermodels: the next generation of intergrated geophysical, chemical and flow simulation modelling tools”. In: *Proceedings World Geothermal Congress 2015*. Melbourne, Apr. 2015.
- [10] Michael J. O'Sullivan, Karsten Pruess, and Marcelo J. Lippmann. “Geothermal reservoir simulation: the state-of-practice and emerging trends”. In: *Proceedings World Geothermal Congress 2000*. Kyushu, May 2000.
- [11] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Vol. 160. Springer Science & Business Media, 2006.
- [12] R. Snieder and J. Trampert. “Inverse Problems in Geophysics”. In: *Wavefield Inversion*. Ed. by Armand Wirgin. Vienna: Springer Vienna, 1999, pp. 119–190.
- [13] Dennis Wokiyi. “Non-linear inverse geothermal problems”. PhD thesis. Linköping University, 2017.
- [14] A. M. Stuart. “Inverse problems: a Bayesian perspective”. In: *Acta Numerica* 19 (2010), pp. 451–559.
- [15] Andrew Power et al. “Data-space inversion for efficient geothermal reservoir model predictions and uncertainty quantification”. In: *Proceedings 43rd New Zealand Geothermal Workshop*. Wellington, Nov. 2021.
- [16] Jérôme Idier. *Bayesian approach to inverse problems*. ISTE Ltd and John Wiley & Sons, Inc., 2008.
- [17] Wenyue Sun, Mun-Hong Hui, and Louis J Durlofsky. “Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure”. In: *Computational Geosciences* 21.5 (2017), pp. 1443–1458.

- [18] Mark A. Beaumont. “Approximate Bayesian computation in evolution and ecology”. In: *Annual Review of Ecology, Evolution, and Systematics* 41.1 (2010), pp. 379–406.
- [19] Richard David Wilkinson. “Approximate Bayesian computation (ABC) gives exact results under the assumption of model error”. In: *Statistical applications in genetics and molecular biology* 12.2 (2013), pp. 129–141.
- [20] Su Jiang, Wenyue Sun, and Louis J Durlofsky. “A data-space inversion procedure for well control optimization and closed-loop reservoir management”. In: *Computational Geosciences* 24.2 (2020), pp. 361–379.
- [21] Mateus M Lima, Alexandre A Emerick, and Carlos EP Ortiz. “Data-space inversion with ensemble smoother”. In: *Computational Geosciences* 24.3 (2020), pp. 1179–1200.
- [22] Morris L. Eaton. “Multivariate statistics: a vector space approach”. In: New York: Wiley, 2006. Chap. 3.
- [23] Adrian Croucher. *Waiwera user guide*. Auckland, Sept. 2021.
- [24] Modelling Uncertainty and Data Group UoA. *ccandu: conditional composition and uncertainty (alpha version)*. 2021.