DATA-SPACE INVERSION FOR EFFICIENT GEOTHERMAL RESERVOIR MODEL PREDICTIONS AND UNCERTAINTY QUANTIFICATION

Andrew Power, Daniel Wong, Ken Dekkers, Michael Gravatt, Oliver J. Maclaren, John P. O'Sullivan, Michael J. O'Sullivan, Ruanui Nicholson

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

ruanui.nicholson@auckland.ac.nz

Keywords: *Data-space inversion, reservoir modelling, geothermal model calibration, uncertainty quantification, data-worth analysis*

ABSTRACT

The ability to make accurate predictions with quantified uncertainty provides a crucial foundation for successfully managing a geothermal reservoir. The standard, state-of-the-art approach to delivering accurate predictions for a geothermal model requires two steps. First, the underlying uncertain parameters (such as subsurface permeability, strength, and location of deep upflows, etc.) are estimated (i.e., calibrated), as well as the associated posterior uncertainty based on measured field data. Second, this uncertainty is propagated from the parameters to the predictions using linear uncertainty analysis or full nonlinear model runs. In most cases, calibrating the unknown parameters accounts for most of the computational costs and time. However, these parameters, and their associated uncertainty, are not always of direct interest. Instead, we are often only interested in the predictions themselves.

The so-called data-space inversion (DSI) methodology provides a solution to this problem. It effectively bypasses the need to calibrate the unknown parameters and directly provides approximate posterior predictions, i.e., model predictions conditioned on the measured field data. Here we review the DSI framework and demonstrate its effectiveness and robustness on several geothermal problems. We find that the DSI approach can lead to significant savings in both computational resources and time, providing an attractive alternative to the more conventional calibration-based approach.

1. INTRODUCTION

Computational models are often used in geothermal resource management to aid in decision making. One of the key features of these models is the ability to make predictions with quantified uncertainty. Generating accurate predictions and uncertainty estimates generally requires calibration of the models used, i.e., fitting model parameters such as subsurface permeability and deep upflows. In many situations, these parameters are not particularly of interest; instead, the predictions are of primary importance. However, the calibration process is typically the most computationally and time demanding step to getting predictions. Here we consider a method to make predictions based on measured data, which avoids the need to calibrate, though still takes into account the measured data. The approach uses the so-called data-space inversion (DSI) methodology (Sun and Durlofsky, 2017; Jiang et al., 2019), which provides a computationally cheap, though approximative, alternative to estimate predictions and the associated uncertainty.

The DSI approach has been used successfully in the subsurface hydrology community; see, for example, Park and Caers (2020), Sun and Durlofsky (2017), Sun et al. (2017), Jiang et al. (2019), among many. However, the application to computational geothermal modelling is novel and potentially more challenging as the governing equations are generally more nonlinear in the geothermal context. Therefore, to test the applicability of the DSI methodology to geothermal applications, we consider two synthetic examples (outlined in Section 3), one based on a simplified zero-dimensional reservoir and one based on a three-dimensional test example outlined in an accompanying paper by Renaud et al. (2021).

Before presenting the DSI framework, we briefly recall the key concepts and steps involved in a typical (Bayesian) statistical approach to geothermal model calibration and prediction and the associated uncertainty quantification.

2. METHODOLOGY

Here we provide a brief review of a standard approach to quantifying uncertainty in geothermal model predictions, as well as the DSI methodology. Uncertainty quantification (UQ) for geothermal reservoir models often uses a Bayesian framework (Kaipio and Somersalo, 2006), which naturally allows for incorporation and quantification of various sources and types of uncertainty. The first step requires calibrating the model parameters, such as subsurface permeability and deep upflows, based on field measurements such as well temperature profiles. The next step is to approximately quantify the remaining posterior uncertainty in the parameters. Finally, the posterior parameter uncertainty is then propagated forward through to the predictions or quantities of interest. We now outline each of these steps.

Calibration of the geothermal reservoir model is typically based on assuming the parameters and data are linked by

$$y_{\text{obs}} = f(k) + e,$$

where $y_{\text{obs}} \in \mathbb{R}^d$ is the data, $k \in \mathbb{R}^n$ denotes the unknown parameters, f represents the forward model, and $e \in \mathbb{R}^d$ the measurement and, possibly, model error. Assuming the errors are independent of the parameters, it is well known, see e.g., Kaipio and Somersalo (2006), that the parameter posterior is given by

$$\pi(k|y_{\text{obs}}) \propto \pi_e\big(y_{\text{obs}} - f(k)\big)\pi(k).$$

The parameter posterior $\pi(k|y_{\text{obs}})$ is the conditional distribution of the parameters given the data, which describes our state of information concerning (or 'beliefs' about) the parameter after taking the data into account. On the other hand, the prior distribution $\pi(k)$ describes our state of information concerning the parameters before (prior to) considering the data. The distribution $\pi_e(y_{\text{obs}} - f(k))$ is called the likelihood, and encodes the new information that the data provides about the parameter. It's worth noting that the likelihood distribution is generally simply $\pi(y_{\text{obs}}|k)$; however, assuming the link between the data and parameters is via the additive error model, $y_{\text{obs}} = f(k) + e$, we have $\pi(y_{\text{obs}}|k) = \pi_e(y_{\text{obs}} - f(k))$ (see e.g. Kaipio and Somersalo 2006).

Assuming *a priori* that the parameters and errors are normally distributed, i.e., $e \sim \mathcal{N}(0, C_{\text{obs}})$ and $k \sim \mathcal{N}(k_0, C_k)$, the posterior is given by

$$\pi(p|y) \propto \exp\left(-\frac{1}{2}\left(\|y_{\text{obs}} - f(k)\|^2_{C_{\text{obs}}^{-1}} + \|k - k_0\|^2_{C_k^{-1}}\right)\right).$$

Fully characterizing the posterior generally requires sampling-based approaches such as Markov chain Monte Carlo (MCMC) (Maclaren et al., 2020; Cui et al. 2011), which for most geothermal models are computationally infeasible. As a computationally cheaper alternative, the *maximum a priori* (MAP) estimate, i.e., the point in parameter space which maximises the posterior density, can be found, after which the posterior uncertainty is approximated using a local Gaussian approximation. Specifically, the MAP estimate is given by

$$k_{\text{MAP}} = \arg\min_{k \in \mathbb{R}^n} \|y - f(k)\|^2_{C_{\text{obs}}^{-1}} + \|k - k_0\|^2_{C_k^{-1}},$$

while the local Gaussian approximation to the posterior is given by $\pi(k|y) \approx \mathcal{N}(k_{\text{MAP}}, C_{\text{post}})$, where

$$C_{\text{post}} = \left(J^T C_{\text{obs}}^{-1} J + C_k^{-1}\right)^{-1},$$

with J denoting the sensitivity (Jacobian) matrix of model with respect to parameters, i.e., $J_{ij} = \partial f_i / \partial k_j$ for $i = 1,2,\dots,d$ and $j = 1,2,\dots,n$.

Carrying out UQ for the predictions relies on a having a further model relating the parameters to the predictions, i.e.

$$y_{\text{pred}} = f_{\text{pred}}(k),$$

where $y_{\text{pred}} \in \mathbb{R}^m$, f_{pred} denotes the predictive model. Note that in various settings, the forward model f, and the predictive model f_{pred}, may in fact be the same models but simply evaluated at different locations in space or time (see Section 3). In any case, to propagate the uncertainty from the parameters through to the predictions, either samples from the parameter posterior are generated and run through f_{pred} to give prediction samples, i.e., $f_{\text{pred}}(k^{(i)})$ where $k^{(i)} \sim \mathcal{N}(k_{\text{MAP}}, C_{\text{post}})$, or an approximation to this process is used. For example, it may in some cases be sufficient to carry out linear uncertainty propagation, resulting in a Gaussian approximation of the prediction distribution, i.e., $y_{\text{pred}} \sim \mathcal{N}(\bar{y}_{\text{pred}}, C_{\text{pred}})$, where

$$\bar{y}_{\text{pred}} = f_{\text{pred}}(k_{\text{MAP}}), \qquad C_{\text{pred}} = J_{\text{pred}} C_{\text{post}} J_{\text{pred}}^T,$$

where J_{pred} denotes the sensitivity (Jacobian) matrix of prediction model with respect to parameters, i.e., $J_{ij} = \partial f_{\text{pred}_i} / \partial k_j$ for $i = 1,2,\dots,m$ and $j = 1,2,\dots,n$.

The main computational bottleneck in the standard procedure is computation of p_{MAP} as it requires the solution of a high dimensional, possibly non-convex, optimisation problem. The DSI framework essentially marginalizes over the uncertain parameters p, as these are of little interest, to focus on the prior and posterior predictive distributions. This is natural when the prediction and associated uncertainties are of primary interest and avoids the need to solve for p_{MAP} altogether. The DSI method is implemented using our *ccandu* Python package (MUD, 2021), and is outlined below.

Initially an ensemble of $E \in \mathbb{N}$ samples from the prior are drawn, $k^{(i)}$ for $i = 1,2,\dots,E$, and are then run through both the forward model and prediction model, giving a new ensemble $\left(k^{(i)}, y_{\text{obs}}^{(i)}, y_{\text{pred}}^{(i)}\right)$, where $y_{\text{obs}}^{(i)} = f(k^{(i)}) + e^{(i)}$ and $y_{\text{pred}}^{(i)} = f_{\text{pred}}(k^{(i)})$. From this ensemble we can compute the sample covariance matrix,

$$\tilde{C}_F = \frac{1}{E-1} \mathcal{C}\mathcal{C}^T = \begin{bmatrix} \tilde{C}_{y_{\text{obs}}} & \tilde{C}_{y_{\text{obs}} y_{\text{pred}}} \\ \tilde{C}_{y_{\text{pred}} y_{\text{obs}}} & \tilde{C}_{y_{\text{pred}}} \end{bmatrix},$$

where $\mathcal{C}(:, j) = \left(y_{\text{obs}}^{(i)}, y_{\text{pred}}^{(i)}\right) - \left(\tilde{y}_{\text{obs}}, \tilde{y}_{\text{pred}}\right)$, with the sample means given by

$$\tilde{y}_{\text{obs}} = \frac{1}{E} \sum_{k=1}^{E} y_{\text{obs}}^{(k)}, \qquad \tilde{y}_{\text{pred}} = \frac{1}{E} \sum_{k=1}^{E} y_{\text{pred}}^{(k)}.$$

ISSN 2703-4275

The conditional mean $y_{\text{pred}}|y_{\text{obs}}$ and conditional covariance $C_{y_{\text{pred}}|y_{\text{obs}}}$, i.e., the expectation and covariance of the predictions conditioned on the measured data can then be computed as

$$\mathbb{E}[y_{\text{pred}}|y_{\text{obs}}] = \tilde{y}_{\text{pred}} + \tilde{C}_{y_{\text{pred}}y_{\text{obs}}}\tilde{C}_{y_{\text{obs}}}^{-1}(y_{\text{obs}} - \tilde{y}_{\text{obs}}), \qquad C_{y_{\text{pred}}|y_{\text{obs}}} = \tilde{C}_{y_{\text{pred}}} - \tilde{C}_{y_{\text{pred}}y_{\text{obs}}}\tilde{C}_{y_{\text{obs}}}^{-1}\tilde{C}_{y_{\text{obs}}y_{\text{pred}}}.$$

Generating samples from the conditional posterior predictive distribution is then a simple problem of linear algebra. That is, taking $r \sim \mathcal{N}(0, I)$, and $LL^T = C_{y_{\text{pred}}|y_{\text{obs}}}$, a sample y_{pred}^s can be simulated by computing

$$y_{\text{pred}}^s = \mathbb{E}[y_{\text{pred}}|y_{\text{obs}}] + Lr.$$

3. NUMERICAL EXAMPLES

We next show the applicability of the DSI approach through two synthetic examples, the details of which are outlined below.

3.1 Lumped Parameter Model

Before applying DSI to a large-scale geothermal model, we test out the methodology on a simple zero-dimensional lumped parameter model (LPM) based loosely on the Ngatamariki geothermal reservoir (Taupo Volcanic Zone, New Zealand). Specifically, we consider the so-called slow drainage model relating the pressure and mass of the geothermal system, which is based on various approximations, see Fradkin et al. (1981) and the references therein, and Figure 1 for a schematic diagram of the model.

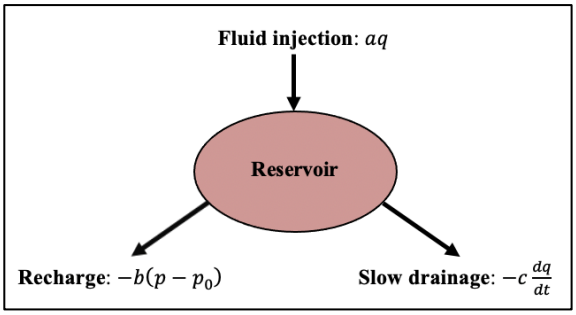

Figure 1: Schematic diagram of the zero-dimensional, slow drainage model.

The resulting model is of the form:

$$\frac{dp}{dt} = aq - b(p - p_0) - c\frac{dq}{dt},$$

where p is the pressure which has initial condition p_0, q is the mass recharge rate, a is the injection coefficient, b the recharge coefficient, and c the slow drainage coefficient. The parameters a, b, and c are all taken as unknown random variables. For the current study we assume the initial pressure and all the parameters are each Gaussian distributed and statistically independent. As such, the prior over the parameters is fully characterized by setting the marginal priors which are taken as

$$\pi(p_0) = \mathcal{N}(2 \times 10^{-2}, 10^{-4}), \qquad \pi(a) = \mathcal{N}(9 \times 10^{-4}, 10^{-6}), \qquad \pi(b) = \mathcal{N}(7 \times 10^{-1}, 10^{-3}), \qquad \pi(c) = \mathcal{N}(9 \times 10^{-5}, 10^{-6}).$$

To construct the required covariance matrices for the DSI we ran the model on 100 samples of the parameters from their associated prior densities. The model was run from $t = 0$ until $t = 35$ days, with two measurements of pressure taken per day. The first 45 data points were used as the observation period, while the last 26 data points were treated as the predictive quantity of interest, see Figure 2. The true data, i.e., the data used to condition the predictions on, is also shown in Figure 2 along with the corresponding true pressure over the predictive time interval, and 100 samples from the predictive posterior. The DSI appears to give fairly good results, with the true pressure over the predictive interval generally lying well within the distribution of the predictive posterior. We also see that there is a substantial decrease in the uncertainty from the priors, and that in line with intuition, the uncertainty in the predictions increases with time. It is also worth pointing out that since the uncertainty in the parameters is so large to begin with that many of the model runs give rise to unphysical negative pressures. On the other hand, after conditioning on the data almost all the predicted pressures at all the predictive times of interest are positive, i.e., physically reasonable.

ISSN 2703-4275

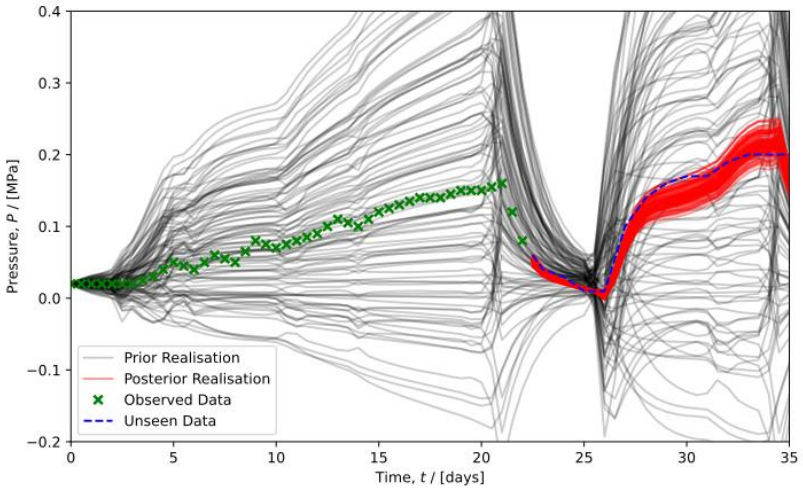

Figure 2: Pressure of the reservoir. Prior samples (grey lines), data used (green crosses), DSI posterior predictions (red lines), and true pressure over predictive interval (blue dashed line).

3.2 Geothermal Reservoir Model

The geothermal reservoir model used here is based on the synthetic model used for teaching and demonstration purposes, described in a submitted accompanying paper (Renaud et al., 2021). The model was generated by GNS Science and is based loosely on various existing New Zealand geothermal fields. The model contains 103 rock types with 3 fault structures, see Figure 3. The equations of state (EoS) used are pure water and energy, while the computational domain is discretised using 13864 blocks which are more finely discretised towards the centre of the domain, see Figure 3.

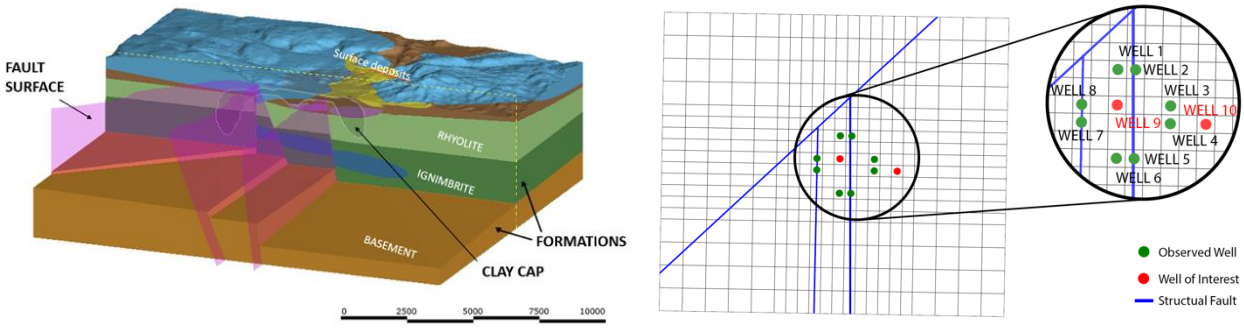

Figure 3: The synthetic geothermal model. The conceptual model (left) and the placement of wells from above (right).

The parameters for the geothermal model are the subsurface (x-, y- and z-) permeabilities of each rock type, as well as the mass flow rates of 10 deep upflows. As such, there are a total of 319 unknown parameters. Observations consist of approximately 25 natural state temperature measurements down each of eight wells (Wells 1-8) which are located in a circular manner, as illustrated in Figure 3 (green wells), near the centre of the computational domain. On the other hand, predications of the pressure are made at approximately 25 points down each of two different wells also at the natural state of the model, one of which (Well 9) is located within the circle of observation wells and the other (Well 10) is located outside of the circle of observation wells, see Figure 3 (red wells).

For each of the parameters (permeabilities and deep upflows) we assign Gaussian priors. For each of the rock-type permeabilities we encode some correlation between the horizontal permeabilities and the vertical permeability. Specifically, we take the correlation between each of the horizontal (i.e., x- and y-) permeabilities to have a Pearson's correlation coefficient of 0.5, while the correlation between the permeability in the x and y direction is 0.8. The permeability of each rock type and strength of each deep upflow are taken to be *a priori* independent of each other. From this prior distribution 1000 samples were generated and run through to natural state using the Waiwera geothermal simulator (Croucher, 2020). Of the 1000 parameter samples, 922 managed to converge. The model temperature output at the observation wells from these converged runs are shown in Figure 4 along with the observed data used to make the predictions at the unobserved wells, i.e., the model run using the true permeabilities and deep upflow strengths.

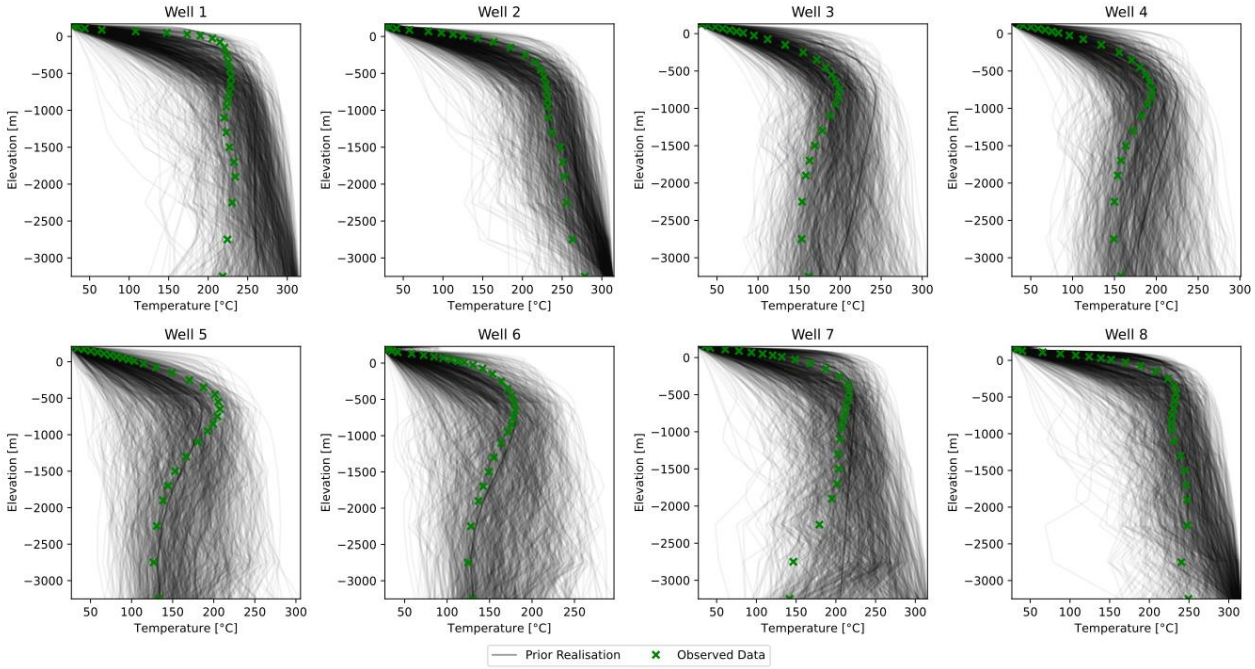

Figure 4: Measurement well temperature. Prior samples (grey lines) and data used (green crosses) down Wells 1-8.

In Figure 5 we show the model predictions of the temperature for the 922 converged runs down Well 9 and Well 10. Also shown is the true, but unobserved, prediction, i.e., the temperature output down the prediction wells from the model run with the true parameters, as well as 100 samples from the posterior found using the DSI methodology. Somewhat expectedly, the DSI predictions are generally better in Well 9 than in Well 10, as Well 9 is located inside the circle of observation wells. The posterior predictions down Well 9 contain the truth for the most part, though the posterior predictions seem to overestimate the shallow temperatures, particularly near the temperature inversion (see Figure 5). The predictive uncertainty down both wells is significantly reduced from the prior uncertainty in the predictions, though more so down Well 9. The posterior prediction down Well 10 follows the truth well, though seems to predict cooler temperatures, generally, than the true temperature down the well.

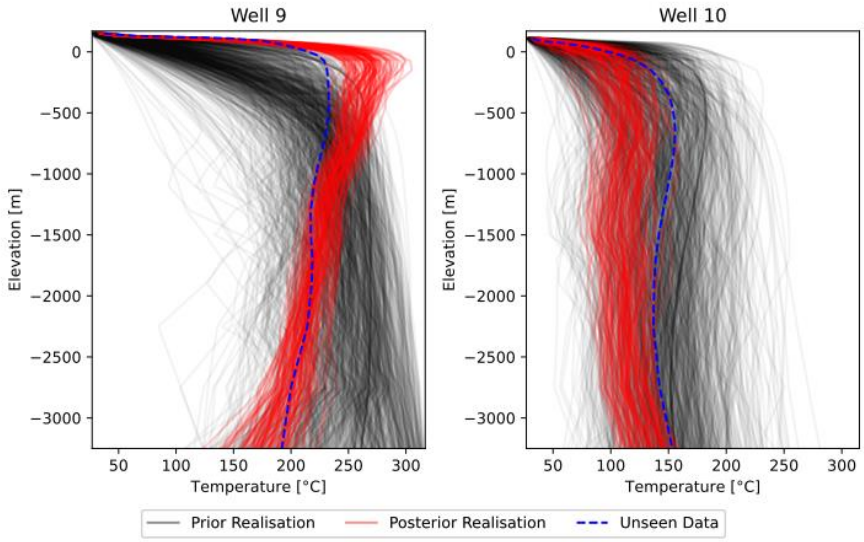

Figure 5: Prediction well temperatures. Prior samples (grey lines), DSI posterior predictions (red lines), and true temperatures (blue dashed line).

4. CONCLUSIONS AND OUTLOOK

In this paper we have considered application of the DSI methodology as a means to efficiently compute geothermal model predictions with quantified uncertainty conditioned on observed data. Before considering a synthetic geothermal model, we tested the applicability of the method to the relatively simple lumped parameter slow-drainage model. In both cases considered here, the DSI approach performed reasonably well, particularly for the simpler slow drainage model.

Future work will likely be pursued in several directions. Firstly, the results found using the DSI method should be compared to results attained using the more conventional approach of first calibrating the model parameters, such as subsurface permeabilities and strength and location of deep upflows, and then propagating these and their associated uncertainties through to the predictions of interest. Secondly, it will be critical to consider the applicability of the DSI approach to making predictions into the future, i.e., predictions in time rather than predictions in space such as those considered in this study, as these are often what decision makers require. Thirdly, it will be important to assess the applicability of the methodology to geothermal reservoirs in which different equations of state are used, for example water-air-energy models. Fourthly, the use of different data types such as pressure and/or the presence (or absence) of surface features should also be investigated. Finally, it will also be crucial to test the applicability of the methods on real world cases, rather than synthetic examples.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL https://www.nesi.org.nz. The authors would like to thank the MBIE research programme "Empowering Geothermal" which has in part funded this research. We would also like to thank Jeremy O'Brien and Sequent for support and use of Leapfrog Geothermal, this is instrumental in model creation of geothermal reservoir model and also Samantha Alcaraz from GNS for the original geological model used in the geothermal reservoir model.

REFERENCES

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

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).

Fradkin, L. J., Sorey, M. L., and McNabb, A. On identification and validation of some geothermal models. *Water Resources Research*, *17*(4), 929-936. (1981).

Jiang, S., Sun, W., and Durlofsky, L. J. A data-space inversion procedure for well control optimization and closed-loop reservoir management. *Computational Geosciences*, 1-19. (2019).

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

Maclaren, O. J., Nicholson, R., Bjarkason, E. K., O'Sullivan, J. P., and 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), e2018WR024240. (2020).

Modeling, Uncertainty, and Data (MUD) Group, The University of Auckland. *ccandu*: conditional composition and uncertainty (alpha-version) (2021). Available on request: oliver.maclaren@auckland.ac.nz.

Park, J., and Caers, J. Direct forecasting of global and spatial model parameters from dynamic data. *Computers & Geosciences*, *143*, 104567. (2020).

Renaud, T., Popineau, J., Riffault, J., Gravatt, M., Yeh, A, and O'Sullivan, M.J.: Practical workflow for training geothermal reservoir modeling. *Proc. 43rd New Zealand Geothermal Workshop*, Wellington, New Zealand. (2020).

Sun, W., & Durlofsky, L. J. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Mathematical Geosciences*, *49*(6), 679-715. (2017).

Sun, W., Hui, M. H., and Durlofsky, L. J. Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure. *Computational Geosciences*, *21*(5), 1443-1458. (2017).