



جامعة الملك عبد الله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Optimization and uncertainty quantification model for time-continuous geothermal energy extraction undergoing re-injection

Item Type	Preprint
Authors	Hoteit, Hussein; He, Xupeng; Yan, Bicheng; Vahrenkamp, Volker
Eprint version	Pre-print
Publisher	arXiv
Rights	Archived with thanks to arXiv
Download date	03/10/2022 08:10:21
Link to Item	http://hdl.handle.net/10754/674025

Optimization and uncertainty quantification model for time-continuous geothermal energy extraction undergoing re-injection

Hussein Hoteit*, Xupeng He, Bicheng Yan, Volker Vahrenkam

Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

December 13, 2021

Abstract

Geothermal field modeling is often associated with uncertainties related to the subsurface static properties and the dynamics of fluid flow and heat transfer. Uncertainty quantification using simulations is a useful tool to design optimum field-development alternatives and to guide decision-making. The optimization process includes assessments of multiple time-dependent flow mechanisms, which are functions of operational parameters subject to subsurface uncertainties. This process requires careful determination of the parameter ranges, dependencies, and their probabilistic distribution functions. This study presents a new approach to assess time-dependent predictions of thermal recovery and produced-enthalpy rates, including uncertainty quantification and optimization. We use time-continuous and multi-objective uncertainty quantification for geothermal recovery, undergoing a water re-injection scheme. The ranges of operational and uncertainty parameters are determined from a collected database, including 135 geothermal fields worldwide. The uncertainty calculation is conducted non-intrusively, based on a workflow that couples low-fidelity models

*Corresponding author: hussein.hoteit@kaust.edu.sa

with Monte Carlo analysis. Full-physics reservoir simulations are used to construct and verify the low-fidelity models. The sampling process is performed with Design of Experiments, enhanced with space-filling, and combined with analysis of covariance to capture parameter dependencies. The predicted thermal recovery and produced-enthalpy rates are then evaluated as functions of the significant uncertainty parameters based on dimensionless groups. The workflow is applied for various geothermal fields to assess their optimum well-spacing in their well configuration. This approach offers an efficient and robust workflow for time-continuous uncertainty quantification and global sensitivity analysis applied for geothermal field modeling and optimization.

Keywords: Geothermal reservoirs, Uncertainty quantification , Design of Experiments, time-dependent Optimization , Proxy modeling.

1 Introduction

The increasing global demand for clean, sustainable, and environmentally-friendly energy is driving accelerated growth for the renewable energy market worldwide. The contribution of geothermal energy was about 4% of renewable energy in 2018 [1], whose role is to complement the other sources of energy for electricity generation, water desalination, cooling, greenhouse farming, among others [2].

Geothermal resources manifest in the subsurface under different temperature conditions ranging from low to high enthalpy. The temperature ranges determine the applicability for electricity generation or other direct use. Resources with temperatures between 150 °C - 300 °C are suitable for electricity generation, while the ones with temperatures below 100 °C are suitable for direct use [3]. Geothermal resources are typically classified into three types of reservoir rocks: volcanic, sedimentary, metamorphic, and mix of them [4]. The rock hydraulic and thermal properties govern the flow behavior and the efficiency of thermal extraction. One of the main challenges of geothermal energy is related to its low efficiency that is typically below 16% [5]. This low efficiency makes geothermal energy less competitive than other renewable energy sources, such as solar and wind. The efficiency of geothermal recovery is a function of complex parameters, including subsurface static reservoir properties (such as permeability, porosity, thickness, and depth), thermal properties (thermal conductivity and temperature), and energy extraction method

(production/injection rate and well spacing) [6, 5, 7, 8]. Other technologies proposed improving efficiency and economics by including the circulation of carbon dioxide or combination with natural gas production [9, 10, 11].

Subsurface properties are often poorly characterized and therefore associated with significant uncertainties, leading to inaccurate predictions and optimization [12, 13, 14, 15, 16]. Therefore, optimizing field recovery performance under various time-dependent uncertainties is crucial to improve efficiency. [6] used tracer and flow-rate data to maximize the net present value of production from fractured geothermal reservoirs. Various multi-objective-optimization approaches have been proposed to address uncertainties in the subsurface and in operational conditions [17, 18, 19]. [20] applied multi-objective particle swarm optimization with experimental-design proxy models to optimize the performance of low enthalpy geothermal reservoirs, taking into account the uncertainty in the reservoir heterogeneity. [21] used an analytical approach to optimize geothermal production from fractured reservoirs, considering the reservoir structural uncertainty. [22] proposed an optimization approach based on dynamic mesh optimization to address the challenges in fixed simulation grids. This issue of poor characterization of subsurface properties has also been reported for geothermal resources in hot dry rock (HDR), where rigorous analytic and numerical modeling tools are crucial to capture uncertainties and optimize the development strategies [23, 24].

There are three essential elements in performing uncertainty quantification and global sensitivity analysis: 1) accurate interpolation of time-dependent responses of thermal recovery and produced-enthalpy rates, 2) efficient evaluation and propagation of multivariable uncertainties with broad ranges of possibilities, and 3) identify the significant parameters and determine their interdependencies. Performing full-physics simulations to capture all uncertainties is computationally impractical. A non-intrusive approach, based on low-fidelity models, is developed to efficiently perform uncertainty quantification [25]. Low-fidelity modeling is a reduced-order process used to project the physical space onto a low-dimensional manifold. This model can be generated using different approaches, such as neural network, multivariate polynomial, k-nearest-neighbors, and decision tree [26, 27, 28, 29, 30]. The construction of a low-fidelity model follows the "training and validation" process in machine learning, which treats the high-fidelity model as a black-box [31, 32, 33, 34, 35]. With rigorous training and validation, the low-fidelity model is expected to capture the governing physics as in the high-fidelity

model but with significantly less computational effort. The training and validation process usually involves the use of Design of Experiments (DoE) for sampling [36, 37, 38, 14, 39, 40, 41]. When the output is time-dependent, the response is often built by several models constructed at discrete time intervals [42]. This approach is frequently utilized in petroleum engineering applications to perform resource assessment and decision-making in field development [36, 38]. In geothermal applications, [14] used it for uncertainty quantification for thermal recovery factor of the injection scheme in a tilted geothermal reservoir. [43] conducted uncertainty quantification for power capacity in the Ngatamariki geothermal field in New Zealand. [44] presented an uncertainty quantification for resource assessment in Atadei geothermal field, Indonesia. However, this approach is computationally intensive due to the construction of a separate model for each time point. The selection of the optimum number and positions of the time points is not systematic as it depends on the smoothness of the response surface in time and space.

Another challenge in developing a global sensitivity analysis process is to identify the significant parameters and their dependencies. Within a geothermal system, there are various complex interacting parameters that can contribute to the performance of the recovery [45, 46, 44]. [47] deployed Sobol-sensitivity indices, based on variance decomposition, to perform global sensitivity analyses for a geothermal heat exchanger. However, when dealing with time-continuous or nested functions, the method becomes inaccurate with the presence of correlated inputs [48].

In this work, we propose a new workflow to efficiently and accurately perform uncertainty quantification and global sensitivity analysis applicable for time-continuous and multi-objective functions. The proposed approach captures the time-continuous profile of the response surface. The concept is to first capture the overall time response with a parameterized empirical model. The reduced-order model is then used to predict the parameters of the empirical model. To reduce the number of runs of the high-fidelity model, we utilize DoE to sample the inputs and conduct feature selection to exclude insignificant inputs. The feature selection is based on a regression-derived sensitivity measure. Secondly, the uncertainty propagation is then performed for the nested functions using Monte Carlo. Finally, we use ANCOVA (Analysis of Covariance) [48] to perform global sensitivity analysis for the time-continuous low-fidelity models. To our knowledge, this workflow is applied in geothermal applications for the first time.

The workflow produces time-continuous uncertainty response to hinder

inaccurate interpolations that occur when the traditional time-independent models are used to capture the overall time response. With this approach, the number of high-fidelity simulations needed to construct the time-continuous low-fidelity models can be reduced. The global sensitivity analysis provides accurate identification of significant parameters (heavy-hitters) that can be strongly correlated. We apply the workflow for a geothermal reservoir undergoing re-injection with a doublet arrangement. The workflow provides predictions of time-dependent thermal recovery and produced-enthalpy rate under subsurface uncertainties, including permeability, porosity, rate, heterogeneity, temperature, well spacing, and thermal conductivity. The optimization model is then applied to 25 geothermal fields to estimate the optimum well spacing. The results are compared to their actual well configurations.

The outline of this paper is the following: In section 2, we discuss the ranges of subsurface uncertainties based on a database including 135 geothermal fields. In section 3, the problem of geothermal production undergoing re-injection is reviewed, followed by the proposed workflow and the time-continuous parametrization in section 4. In section 5, the reservoir simulation model, the physical phenomena, and the approximation functions are presented. In section 6, we present the uncertainty quantification results, followed by sensitivity analysis to show the effect of different designs and sample numbers towards the training accuracy, predictability, feature selection results, uncertainty propagation, and global sensitivity analysis. Finally, a discussion and conclusions are outlined.

2 Variability in subsurface uncertainties

Rock formation, depositional setting, and diagenetic and tectonic overprints create wide variability in subsurface geothermal reservoirs, resulting in substantial variability in hydraulic and thermal properties [4]. We compiled data from 135 hydrothermal fields worldwide, covering various types of reservoir rocks, including sedimentary, volcanic, and metamorphic. The distribution of the geothermal fields, as appears in Figure 1, shows that most of the fields are located around the Ring of Fire (Indonesia, Japan, Philippines, Russia, Tibet, New Zealand, western part of USA, and some Latin American countries), Turkey, western Europe, and some African countries. The high-temperature fields are mostly around the Ring of Fire, while the low-temperatures are mostly in Europe. The temperatures of the geothermal

fields vary between 80 and 242 °C, are utilized for various applications, including electricity generation and heating. Six critical parameters are iden-

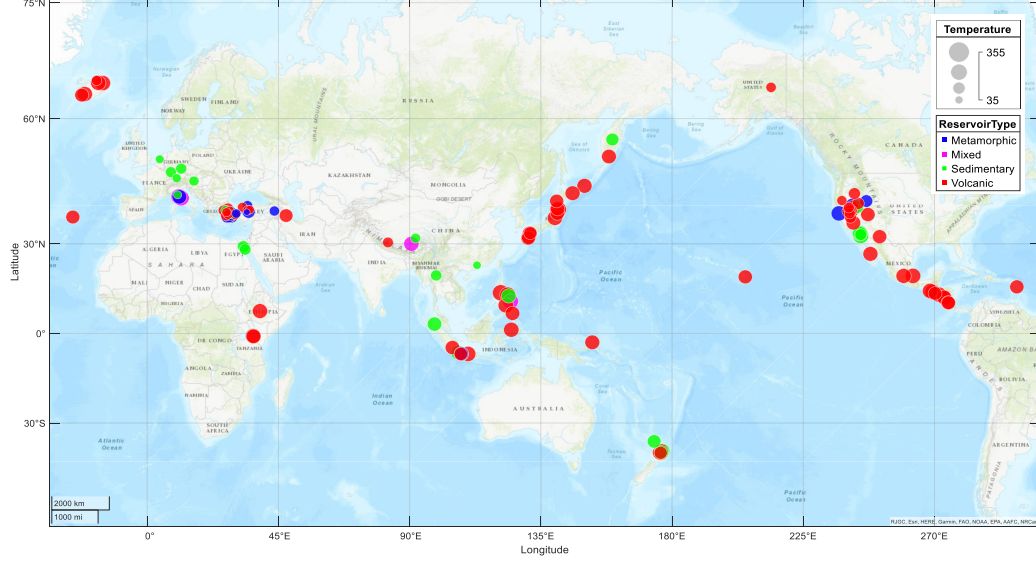


Figure 1: Map showing the locations of 135 geothermal fields. The size of the bubbles reflects the temperature range, while the color indicates the reservoir rock type.

tified from the database to be addressed in the uncertainty quantification study, including temperature, permeability, porosity, production rate, well spacing, and formation thermal conductivity. The distributions of the six parameters are depicted in Figure 2, which show that the permeability from various fields varies between 20 and 1600 mD , reflecting a wide range of variability due to the inherent heterogeneity of geothermal reservoirs. The porosity varies between 0.02 and 0.16. The low-porosity ranges correspond to volcanic and metamorphic reservoirs, while high-porosity ranges correspond to sedimentary reservoirs. The thermal conductivity is within the range of 1.6 - 3 $W/(m \cdot K)$, reflecting the variations among basaltic, sandstone, and metamorphic rocks. The 10th (P10) and 90th (P90) percentiles of the average production rate per well varies within 9.2 - 77.6 l/s . The operating rate depends on the permeability and the amount of thermal energy of the reservoirs. Most fields are developed under doublet arrangement with injection and production wells with a well spacing between 302 and 1514 m .

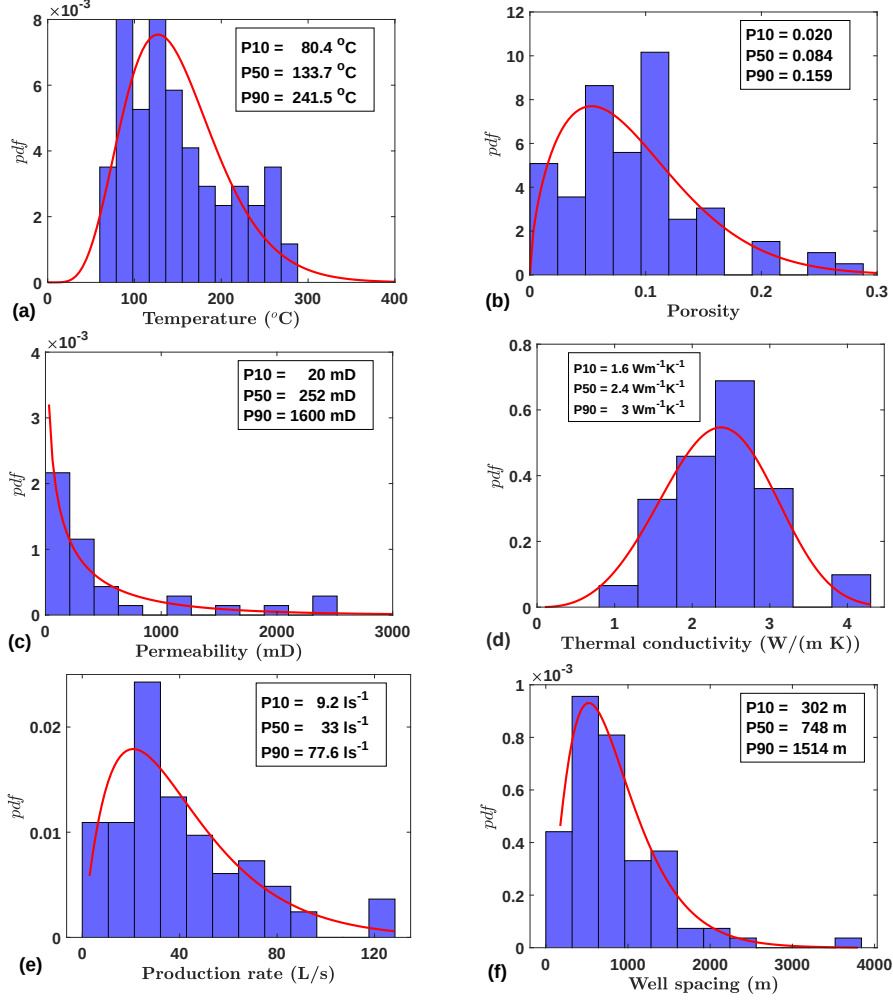


Figure 2: Analog data gathered from 135 geothermal fields reflecting the ranges of several parameters, including (a) temperature, (b) porosity, (c) permeability, (d) thermal conductivity, (e) production rate per well, and (f) well spacing. The data is adopted from [7].

3 Problem formulation

The heat extraction from a geothermal reservoir with water re-injection scheme is governed by three fundamental equations: mass, momentum, and energy conservation equations [49, 50]. The mass conservation in 3D space

is expressed by:

$$\frac{\partial}{\partial t} (\phi \rho_\alpha S_\alpha) + \nabla \cdot (\rho_\alpha u_\alpha) = Q_\alpha \quad \alpha = l, v, \quad (1)$$

where t is time, ϕ is the porosity of the formation, ρ is the phase density, S is the fluid phase saturation in the formation, u is the phase superficial velocity, and Q is the phase sink/source term. The subscription α denotes the phase where l is for liquid, and v is for vapor. The number of mass conservation equations follows the number of phases, N .

The superficial velocity for each phase is modeled using the extended Darcy's law, that is,

$$u_\alpha = -\frac{\mathbf{k} k_{r\alpha}}{\mu_\alpha} \nabla \Phi_\alpha, \quad (2)$$

where $\mathbf{k}$ is the absolute permeability of the reservoir, k_r is the phase relative permeability, μ is the phase dynamic viscosity, and Φ is the phase potential, which is given by:

$$\Phi_\alpha = P_\alpha - \rho_\alpha g z, \quad (3)$$

where P_α is the phase pressure, g is the gravitational acceleration, and z is the phase hydrostatic height.

The energy equation describes the dynamics of the temperature in the reservoir, that is,

$$\begin{aligned} \frac{\partial}{\partial t} \left(\phi \sum_{i(\alpha)}^N \rho_i S_i U_i + (1 - \phi) \rho_r U_r \right) + \nabla \cdot \left(\sum_{i(\alpha)}^N \rho_i u_i H_i \right) \\ - \nabla \cdot \left(\left(\phi \sum_{i(\alpha)}^N S_i \sigma_i + (1 - \phi) \sigma_r \right) \nabla T \right) = W, \end{aligned} \quad (4)$$

$$H_\alpha = U_\alpha + P_\alpha \bar{V}_\alpha, \quad (5)$$

$$U_\alpha = C_\alpha (T - T_{ref}), \quad (6)$$

$$U_r = C_r (T - T_{ref}), \quad (7)$$

where U is the internal energy of a phase or rock, H is the enthalpy of a phase or rock, $\bar{V}$ is the volume per unit mass of a phase, C is specific heat capacity per unit mass of a phase or rock, T is the temperature of the reservoir system, T_{ref} is the reference temperature, typically corresponds to the fluid injection temperature, σ is the thermal conductivity of a phase or rock, and W is the energy source/sink term. The subscription r denotes the rock. The density of the fluid is a function of pressure and temperature, which can be calculated with an Equation of State [50]. The unknowns correspond to the saturation, pressure, and temperature/energy. In this work, the thermal recovery equations are solved implicitly using CMG-STARs simulator [51]. The boundary conditions, corresponding to the heat loss/gain to the overburden and underburden, are also considered.

4 Proposed workflow

4.1 Dimensionless variables

The main objective of the study is to evaluate the thermal recovery and produced-enthalpy rate under uncertainties. To be physically scalable for different static and dynamic conditions of a reservoir, several dimensionless groups are considered. These dimensionless groups correspond to the operational time, thermal recovery factor, and enthalpy production factor, defined by:

$$t_D = \frac{Q_\alpha t}{V_p}, \quad (8)$$

where V_p is the reservoir's pore-volume. The dimensionless time represents the cumulative volume of injected water into the reservoir at reservoir conditions relative to the reservoir's pore-volume. We use the thermal recovery factor and enthalpy production factor to quantify the energy production efficiency.

The thermal recovery factor (F_{TR}) reflects the change in reservoir energy over time due to the extraction of energy and interaction with the re-cycled cold water relative to the original energy in place, that is,

$$F_{TR}(t_D) = \frac{U(t_D = 0) - U(t_D)}{U(t_D = 0)}. \quad (9)$$

The enthalpy production factor (F_{EP}) measures the decline rate (i.e., sustainability) of heat production. It is defined as the ratio of the produced

enthalpy at a given time and the initial enthalpy, that is,

$$F_{EP}(t_D) = \frac{\dot{H}(t_D)}{\dot{H}(t_D = 0)}, \quad (10)$$

where $\dot{H}$ is the enthalpy production rate during the extraction process. Initially, this indicator will be unity, and then it declines as a result of reservoir cooling and cold front breakthrough. The performance indicators F_{EP} and F_{TR} are used as the objective functions to maximize in our uncertainty quantification and optimization.

4.2 Non-intrusive uncertainty quantification

A non-intrusive uncertainty quantification consists of two essential steps: 1) constructing low-fidelity (surrogate) model to represent the high-fidelity (physics-based) model, and 2) propagating uncertainty with Monte Carlo using the low-fidelity model. The low-fidelity model is expressed in spectral expansion as follows:

$$f(\xi) = \sum_{\alpha \in A} f_{\alpha} \Psi_{\alpha}(\xi), \quad (11)$$

where f represents the quantity of interest (QoI), $\Psi_{\alpha}(\xi)$ is an orthogonal basis function, and ξ is the variable of interest. The symbol α denotes a multi-index scheme, defined by:

$$A = (\alpha^1, \alpha^2, \dots, \alpha^{M+1}), \quad (12)$$

where, $(M + 1)$ denotes the number of basis functions, including the zeroth-order. For multiple inputs, the indexing for α becomes α_i^k , where i is the index for the variables, k is the index for the basis function, and α returns the polynomial order used for a certain variable and certain basis number. The truncation of the basis functions follows the total order scheme [37], expressed as:

$$|\alpha| = \sum_{i=1}^d \alpha_i^k \leq p, \quad (13)$$

where d is the total number of input parameters, and p is the highest order within the polynomial basis. The size of α , corresponding to the number of basis functions within this scheme, is given by:

$$(M + 1) = \frac{(d + p)!}{d!p!}. \quad (14)$$

For instance, a case with two input variables and a polynomial of order 2 has 6 basis functions, including the zeroth-order. The expansion is written as follows:

$$\begin{aligned} f(\xi) = & f_0 \Psi_0(\xi_1) \Psi_0(\xi_2) + f_1 \Psi_1(\xi_1) + f_2 \Psi_1(\xi_2) \\ & + f_3 \Psi_2(\xi_1) + f_4 \Psi_2(\xi_2) + f_5 \Psi_1(\xi_1) \Psi_1(\xi_2) \end{aligned} \quad (15)$$

The unknowns are the coefficients, f_k , of the basis functions. The coefficients in Eq. 15 can be obtained through projection using Gauss quadrature or Least Square minimization [37]. In this study, we use the Least Square minimization method, which seeks to minimize the residual error, defined as:

$$\mathbf{f} = \arg \min_{f_k \in \mathbb{R}} \frac{1}{N_s} \sum_{i=1}^{N_s} \left(f_i^{true} - \sum_{\alpha \in A} f_{\alpha} \Psi_{\alpha}(\xi) \right)^2, \quad (16)$$

where N_s is number of observations. Superscription *true* denotes the observation data.

For uncertainty propagation, Monte Carlo is applied to perform sampling from the input ranges. The outputs are then calculated through the low-fidelity model. These outputs form a distribution from which the P10 (10% probability), P50 (50% probability), and P90 (90% probability) are deduced.

4.3 Modeling of time-continuous response

We propose a new method to model the time-continuous responses of thermal recovery and enthalpy production, based on a nested function approach. With this approach, a time-dependent objective function, y , is represented by an empirical function as follows:

$$y = g(w_1(\xi), \dots, w_{N_w}(\xi), t), \quad (17)$$

where g is a given regression function, designed to fit y , and $w_{i,i=1,\dots,N_w}$ are the corresponding fitting parameters to be determined. The number of coefficients, N_w , is preferred to be small but high enough to capture the desired accuracy of approximation. We apply the spectral polynomial expansion to approximate the time-independent coefficients w_i , rather than the time-dependent objective function, y , which is commonly used in the literature. In this work, the thermal recovery factor is described with a two-parameter

function, expressed by:

$$F_{TR}(t_D, \xi) = \frac{w_1(\xi) t_D}{1 + w_2(\xi) t_D}. \quad (18)$$

While the enthalpy production function is described with a three-parameter function:

$$F_{EP}(t_D, \xi) = w_3(\xi) + \frac{1 - w_3(\xi)}{1 + \left(\frac{t_D}{w_4(\xi)}\right)^{w_5(\xi)}}. \quad (19)$$

The regression functions in Eqs. 18 and 19 were selected based on their suitability to capture the profiles of the physical functions with the least number of parameters. The parameters $w_{i,i=1,\dots,N_w}$ are determined from the low fidelity model given in Eq. 11 .

4.4 Uncertainty quantification and optimization workflow

The proposed workflow to develop time-continuous, regression-based uncertainty quantification consists of five steps, as illustrated in Figure 3. The workflow consists of selection and identification of the significant uncertainty parameters, building the proxy models, and perform sensitivities, which are described in more detail, as follows:

Step 1 - Identification of uncertainty parameters: uncertainty parameters are identified based on the impact of their range of variability (upper and lower bounds) on the response functions, corresponding to the thermal recovery and enthalpy production factors. Therefore, the parameter ranges affect the universality of the reduced-order model and the sensitivity analysis. We opted to group selected variables to accommodate dependency among them. This step is important to ensure physical consistency of the low-fidelity model.

Step 2 - Selection of significant parameters: in this step, a 2-level DoE, such as folded Plackett-Burman, 2-level Full Factorial, or 2-level Fractional Factorial, is used to perform the screening process to select the most significant parameters in the objective functions. The concept is to measure the sensitivity of the time-continuous coefficients w_i , representing the regression functions of the responses, to the input

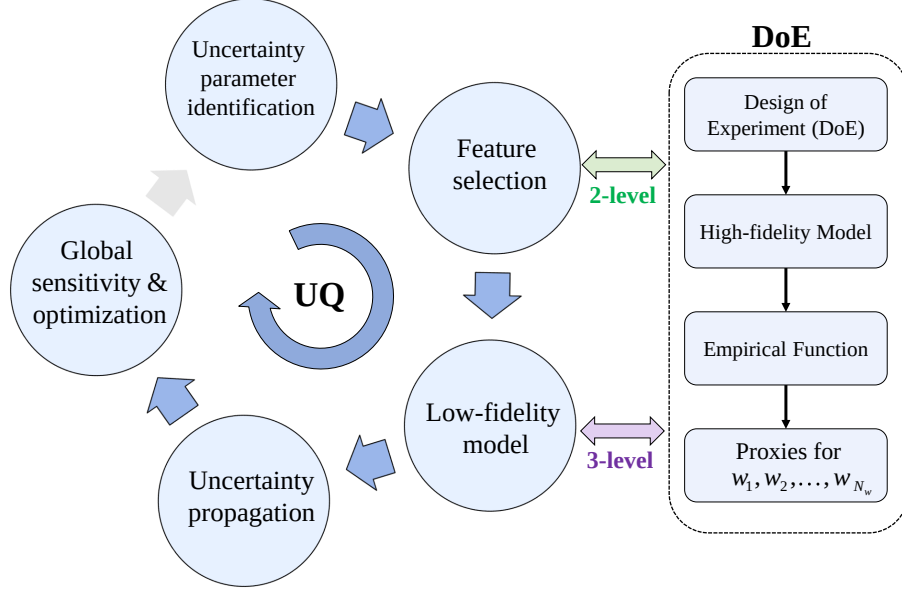


Figure 3: Proposed Uncertainty Quantification (UQ) workflow to develop time-continuous, regression-based uncertainty quantification. The processes within the dashed line consist of fitting, training, and testing of the low-fidelity model from several realizations of the high-fidelity model. The parameters within the processes are sampled with either 2-level (for feature selection) or 3-level (for final response) DoE. The final response describes time-continuous objective functions, F_{TR} and F_{EP} . Monte Carlo simulations are conducted for uncertainty propagation. The global sensitivity analysis uses ANCOVA technique.

parameters (modifiers). Since the 2-level DoE adopt first-order polynomials as the basis for the low-fidelity model, the sensitivity measure is also of the first-order [52]. This can be inferred directly from the regression coefficients. The reduced-order model from a 2-level DoE can be expressed by:

$$f(\xi) = f_0 + \sum_{i=1}^d f_i \Psi_1(\xi_i). \quad (20)$$

The variance of f is defined by:

$$Var[f] = \sum_{i=1}^d f_i^2 Var[\xi_i] \quad (21)$$

The sensitivity is of the first-order, that is,

$$S_i = \frac{f_i^2 Var[\xi_i]}{Var[f]}, \quad i = 1, \dots, d, \quad (22)$$

with,

$$\sum_{i=1}^d S_i = 1. \quad (23)$$

By conducting a hypothesis testing of 95% confidence based on the sensitivity indices, insignificant input parameters can be eliminated.

Step 3 - Low-fidelity model: after selecting the significant parameters, a more accurate 3-level DoE, such as D-Optimal, Taguchi, Latin Hypercube, or Box-Behken, is used to sample the variables. The responses corresponding to the DoE samples are obtained from the high-fidelity simulation model. The input-output samples are then used to train the low-fidelity model with a Least-Square scheme, using the Least Angle Regression (LARS) algorithm in UQLab [53]. Variable degrees of the polynomial-basis-function were investigated using two metrics to evaluate the low-fidelity model: 1) R^2 for measuring the accuracy of the training, and 2) Q^2 for measuring the predictability of the model, such that,

$$R^2 = \frac{\sum_{i=1, i \in T}^{N_s} \left(f_i^{true} - \sum_{\alpha \in A} f_{\alpha} \Psi_{\alpha}(\xi) \right)^2}{\sum_{i=1, i \in T}^{N_s} \left(f_i^{true} - \frac{1}{N_s} \sum_{k=1, k \in T}^{N_s} f_k^{true} \right)^2}, \quad (24)$$

The difference between R^2 and Q^2 is in the data sets used for their evaluation, where R^2 is evaluated using the training data T , and Q^2 is evaluated using the testing data D . The testing set contains sampling from the space of modifiers, which are different from the training set.

Step 4 - Uncertainty propagation: Monte Carlo simulations are applied to the low-fidelity model to capture uncertainty propagation. The sampled realizations are collected in the form of distributions from which the probabilistic forecast of percentiles, P10, P50, and P90 are evaluated.

Step 5 - Global sensitivity analysis using ANCOVA: this step aims to evaluate the global sensitivity of the objective functions, including the time-dependent thermal-recovery and enthalpy-production factors. Since the objective functions are in the form of nested functions (see Eqs. 19 and 20), direct evaluation of sensitivity analysis (Eq. 21 to 23) cannot be performed [48]. The commonly used ANOVA (Analysis of Variance) or Sobol sensitivity indices may miss-predict the sensitivity value due to the complex dependency among variables. To overcome this issue, we utilized ANCOVA indices, which provide triplet sensitivity values: individual (uncorrelative) sensitivity, interactive, and correlative sensitivity [48]. The uncorrelative sensitivity represents the uncorrelated structural contribution of a parameter, which is expressed as:

$$S_i^U = \frac{Var(f_i(\xi_i))}{Var[f]}, \quad i = 1, \dots, d. \quad (25)$$

Both interactive and correlative sensitivity denote the correlated contribution among parameters. The interactive sensitivity indices are given by:

$$S_i^I = \frac{Cov \left[f_i(\xi_i), \sum_{\substack{\mathbf{v} \subset \{1, \dots, N_r\} \\ \{i\} \in \mathbf{v}}} f_{\mathbf{v}}(\xi_{\mathbf{v}}) \right]}{Var[f]}, \quad i = 1, \dots, d. \quad (26)$$

While the correlative sensitivity indices are defined by:

$$S_i^C = \frac{Cov \left[f_i(\xi_i), \sum_{\substack{\mathbf{w} \subset \{1, \dots, N_r\} \\ \{i\} \notin \mathbf{w}}} f_{\mathbf{w}}(\xi_{\mathbf{w}}) \right]}{Var[f]}, \quad (27)$$

$$i = 1, \dots, d,$$

In the above equations, N_r is the number of Monte Carlo realizations, and $\xi_{\mathbf{v}}$ and $\xi_{\mathbf{w}}$ are independent realizations used to approximate the covariance and variance [48]. The total contribution of uncorrelative (U), interactive (I), and correlative (C) sensitivity is of the first-order, which is expressed by:

$$S_i = S_i^U + S_i^I + S_i^C. \quad (28)$$

This step is critical to identify the governing factors that impact the objective functions. The ANCOVA calculations were performed using UQLab [53].

5 Simulation model

5.1 Simulation domain

To assess the impact of subsurface uncertainties on the thermal recovery and enthalpy production, we consider a 3D mechanistic model with one vertical injector and one vertical producer, as illustrated in Figure 4. This setup follows the doublet arrangement, which is commonly implemented for production-injection recovery schemes in geothermal development [54].

The model consists of 12 geological layers with different permeabilities. The top and bottom layers in the 3D model are impermeable for fluid flow. However, they are conductive for heat to capture heat loss or gain to the overburden and the underburden. The model resembles a layered-cake configuration where each layer is assumed homogeneous. Besides the six uncertainty parameters discussed in Section 2, we add a parameter to reflect the vertical heterogeneity to account for multi-layer cross-flow behavior. The measure

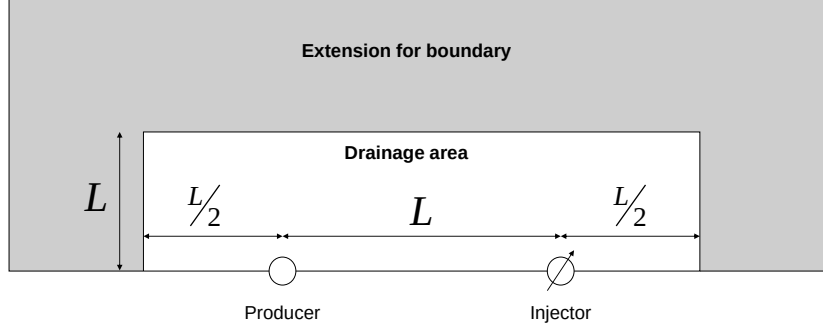


Figure 4: Illustration of the 3D mechanistic model with two wells (top view). The shaded zone is used to account for the thermal boundary conditions.

of the contrast in vertical permeability is quantified by the Dykstra-Parson coefficient [55], denoted by V_{DP} , and defined by:

$$V_{DP} = \frac{k_{0.5} - k_{0.16}}{k_{0.5}}, \quad (29)$$

where the subscripts 0.16 and 0.5 denote, respectively, the 16% and 50% probabilities from the Cumulative Distribution Function (CDF) of the permeability. In this work, V_{DP} varies between 0.56, representing an almost homogeneous case, to 0.94 representing a very heterogeneous case. Examples of the realizations are shown in Figure 5.

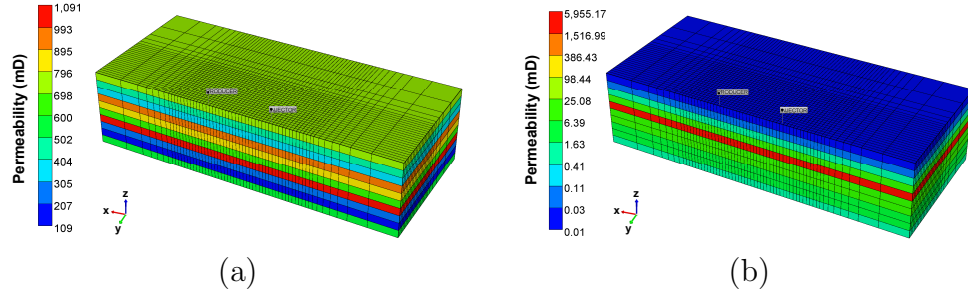


Figure 5: Two realizations of the model with vertical permeability distribution corresponding to Dykstra-Parson coefficients, $V_{DP} = 0.56$ (a), and $V_{DP} = 0.94$ (b).

Because of the symmetry across the vertical plane connecting the two wells, only one side of the model is considered in the simulations. To account for the

lateral boundary conditions, we extended the model and used a pore-volume multiplier on the boundary grid blocks to mimic infinite-acting regime [54]. The drainage volume used to calculate the recovery factor and energy in place is assumed to be relative to the well spacing L , as shown in Figure 4. The production rate is a function of permeability multiplied by the reservoir pay thickness (that is, kh) and varies between 3 l/s to 128 l/s , following the collected data ranges in Figure 2. The porosity is considered as a function of permeability. The 3D model is discretized into a $58 \times 25 \times 12$ irregular Cartesian grid. This grid resolution was verified and found to be adequate for this study. The model is set to re-inject the wastewater at a constant temperature of 60 °C. The initial reservoir pressure is set at 15000 kPa. In Figure 6, we show the results for a typical simulation case corresponding to 46.3 l/s production rate, 600 mD permeability, 113 °C initial temperature, 0.1 porosity, 2.4 $W/(m \cdot K)$ thermal conductivity, 700 m well-spacing, and various Dykstra-Parson coefficients with values at 0.56 and 0.94, respectively.

A typical flow behavior occurring in the reservoir during water recycling is gravity-dominated. Since all layers are hydraulically connected, the injected cold water tends to flow downward because of buoyancy. Figure 6 also shows that the thermal breakthrough is slower than the water-front breakthrough due to heat conduction from the reservoir and the surrounding formation. Figure 7a and 7b show, respectively, the decline in the energy in place and the enthalpy production rate as a function of the dimensionless time t_D for two different Dykstra Parson coefficients. The energy in place reflects the change in temperature, pressure, and mass at the reservoir conditions. In this case, water is re-injected, and therefore, no significant change in the mass of the fluid occurs in the reservoir. Since the reservoir pressure is almost constant, the change in the internal reservoir energy is mainly influenced by reservoir cooling. The energy decline can be simply calculated from the temperature decline and vice-versa using Eqs. 5 to 7. In the case of significant heterogeneity, the average temperature and energy in place decline faster, which is due to the presence of thief layers that can lead to a faster breakthrough, resulting in lower efficiency [56]. Note that thief layers (or high conduit fractures) serve to maintain high injectivity/productivity. However, water breakthrough along thief layers is a key factor in diminishing the efficiency of heat recovery, similar to hydrocarbon production from oilfields and water flow in fractured media [57, 58, 59]. Therefore, well spacing between the injectors and producers should be optimized to avoid premature water

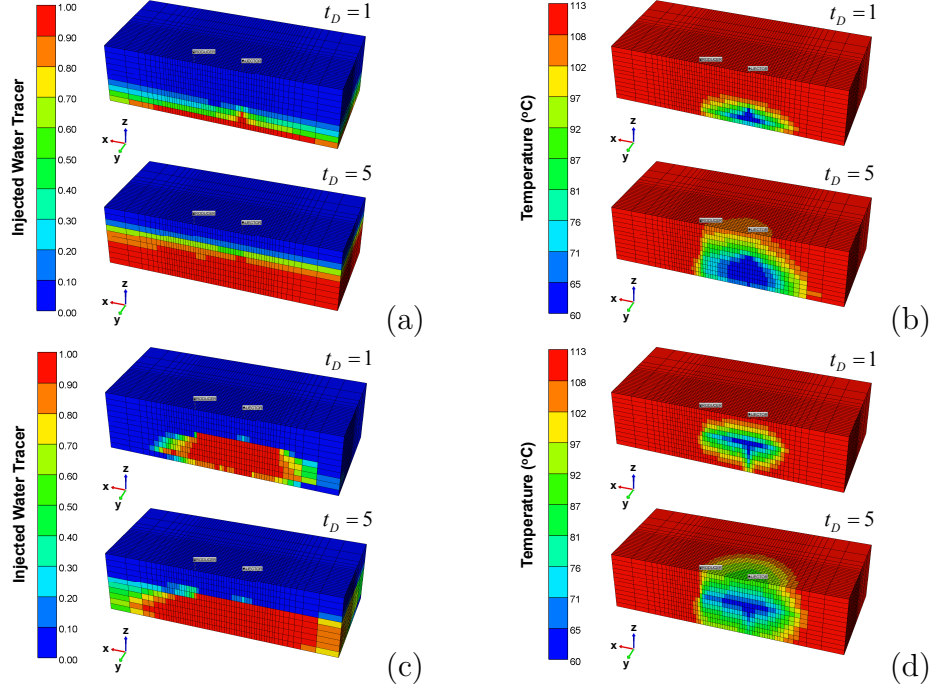


Figure 6: Simulation results for the propagation of injected water-front, captured by a tracer (a, and c) and temperature profiles (b, and d) at different dimensionless times $t_D = 1$ and $t_D = 5$, respectively. The maps in (a) and (b) are for a low-heterogeneous case with $V_{DP} = 0.56$, and the rest are for a high-heterogeneous case with $V_{DP} = 0.94$.

breakthrough and maximize thermal recovery [60, 8].

Figure 7b shows the enthalpy production rates versus the dimensionless time. The produced enthalpy rate is constant at constant production mass rate and temperature. The enthalpy decline is a result of the cold water-front breakthrough, which leads to a drop in the production fluid temperature. This concept is illustrated in Figure 6, which shows the thermal front propagation from the injector to the producer at different times. Reservoir heterogeneity causes early breakthrough, which decreases the temperature at the producer leading to a faster decline in enthalpy production rate. These simulation cases represent typical behaviors of the simulation model, which is used in the optimization process.

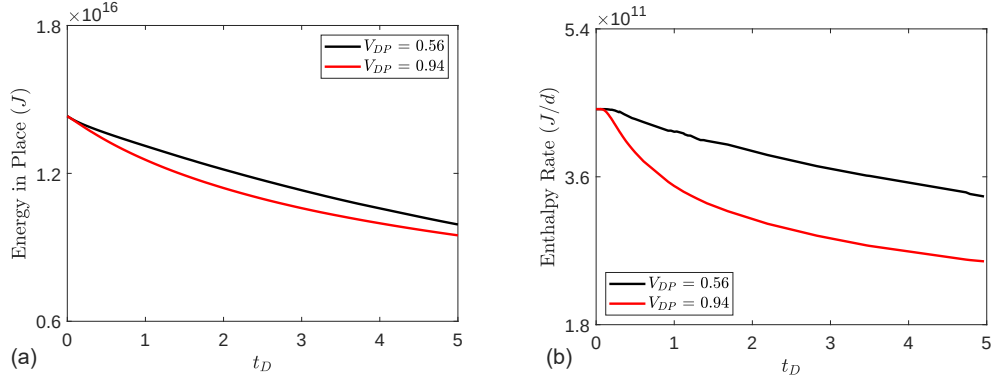


Figure 7: Simulation results for two scenarios with different heterogeneity showing the reservoir energy decline versus the dimensionless time (a), and enthalpy production rate (b).

5.2 Capturing the physics with the low-fidelity model

We demonstrate the applicability of the low-fidelity models in capturing the physics-based solutions. For a given set of simulation results, regression is performed to determine the fitting parameters $\omega_{i,i=1,\dots,5}$, corresponding to the analytical expressions of the thermal recovery factor and enthalpy production factors, given in Eqs. 18 and 19. The low-fidelity models for F_{TR} and F_{EP} can adequately fit the change in thermal recovery and enthalpy production rate versus time, as demonstrated in Figure 8 for two test cases with different heterogeneities. In the workflow, the fitting coefficients $\omega_{i,i=1,\dots,5}$ are automatically determined as functions of the uncertainty parameters, including permeability, heterogeneity, well spacing, and temperature.

6 Results

The quantified uncertainty and design parameters are the permeability (k), temperature (T), well spacing (L), heterogeneity (V_{DP}), and thermal conductivity (σ). An efficient 2-level folded Plackett-Burman DoE is initially used to determine the significant parameters (feature selection). Reservoir porosity and production rate are considered as dependent parameters on the reservoir permeability to avoid physical inconsistency in the sampling [61, 62]. The feature selection process, used to identify the significant parameters, evaluates the sensitivity of the objective functions, represented by the fitting

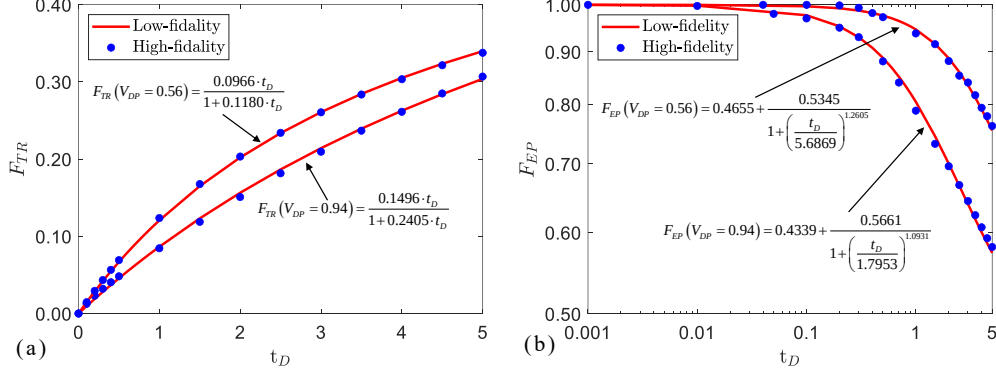


Figure 8: Fitting the physics-based (high-fidelity) solutions for two test cases with different heterogeneities by the low-fidelity models for (a) thermal recovery factor F_{TR} , and (b) enthalpy production factor F_{EP} .

coefficients $\omega_{i,i=1,\dots,5}$. The design requires 13 simulation cases including the endpoints and the center point of the variation intervals of the permeability k , temperature T , well spacing L , heterogeneity V_{DP} , and thermal conductivity σ . Some of the simulation cases from the design and the corresponding fitted low-fidelity model are shown in Figure 9. Figure 10 shows that the p-values for k , T , L , and V_{DP} are above 0.05, while the one for σ is below 0.05, indicating insignificance of its variation on the objective functions. Therefore, σ was not included as an uncertainty parameter in the construction of the low-fidelity models for F_{TR} and F_{EP} . Instead, the value of σ is fixed at its P50.

After eliminating the insignificant parameters, a Latin Hypercube design is used to construct the responses of the time-independent coefficients w_i using 4 groups of parameters. These coefficients are then used to assemble the objective functions F_{TR} and F_{EP} . Convergence analysis was conducted to select the order of approximation of the polynomial chaos expansions (PCE), as well as the size of the design. Increasing the order of PCE can improve the accuracy of the fitting of the training sample. However, it may cause over-fitting, leading to poor model predictability. Figure 11a shows that as the order of PCE increasing, the accuracy of reproducing the training set increases, while predictivity may decrease. The predictability of the model, measured by the blind-test samples including 1000 cases, shows that a PCE of order 3 is optimum, and higher-order leads to overfitting. On the other

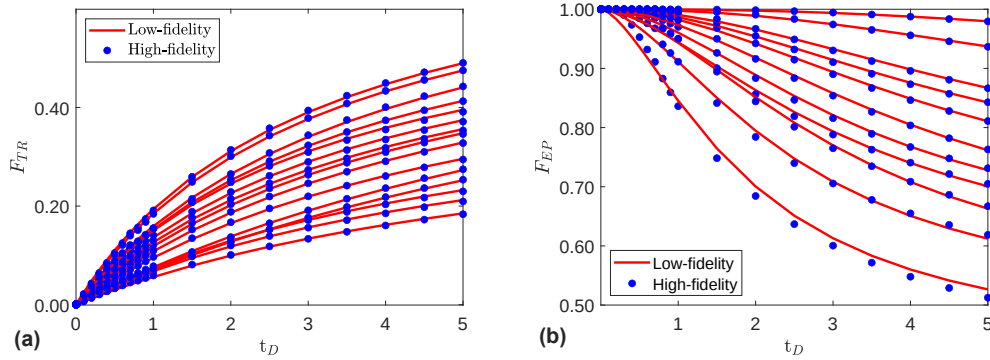


Figure 9: Results showing the reproduction of the high-fidelity model versus time by the low-fidelity model for several cases used in the sensitivity analysis by the low-fidelity model for (a) F_{TR} and (b) F_{EP} .

hand, the size of the training sample in the Latin Hypercube design should be high enough to guarantee the desired accuracy in the blind test. Figure 11b shows no significant improvement in the accuracy of the design with more than 150 cases, which was the selected size. Parity plots showing the accuracy of the proxy model in reproducing the training set and in predicting the blind testing set are shown in Figure 12. The performance of other designs, including Full Factorial, Box Behken, Taguchi, Central Composite, and D-Optimal are discussed in Appendix A. To avoid over-fitting from the training set, emphasis should be given to predictability using blind testing. The obtained training accuracy (R^2) for the low-fidelity models for the 150 cases is above 0.99, while the predictability accuracy (Q^2) for 1000 blind cases is above 0.92. The low-fidelity models provide the PCE-based proxy to estimate the time-independent w_i as a function of the well spacing L , permeability k , heterogeneity V_{DP} , and temperature T . Figure 13 shows examples of the behavior of four coefficients, w_1 , w_2 , w_4 , and w_5 , as a function of L and V_{DP} . The approximated fitting coefficients are then used in Eqs. 18 and 19 to calculate the time-dependent objective functions, F_{TR} and F_{EP} . The quality of the predictions of F_{TR} and F_{EP} compared to the physics-based simulation results at $t_D = 2$ are shown in Figure 14. The predictability of the model other times showed similar behavior.

The pdf-s for F_{TR} and F_{EP} are shown versus time in Figure 15. The predictions are generated by performing Monte Carlo simulations using the low-fidelity models with 10,000 samples. The mean prediction of F_{TR} has the

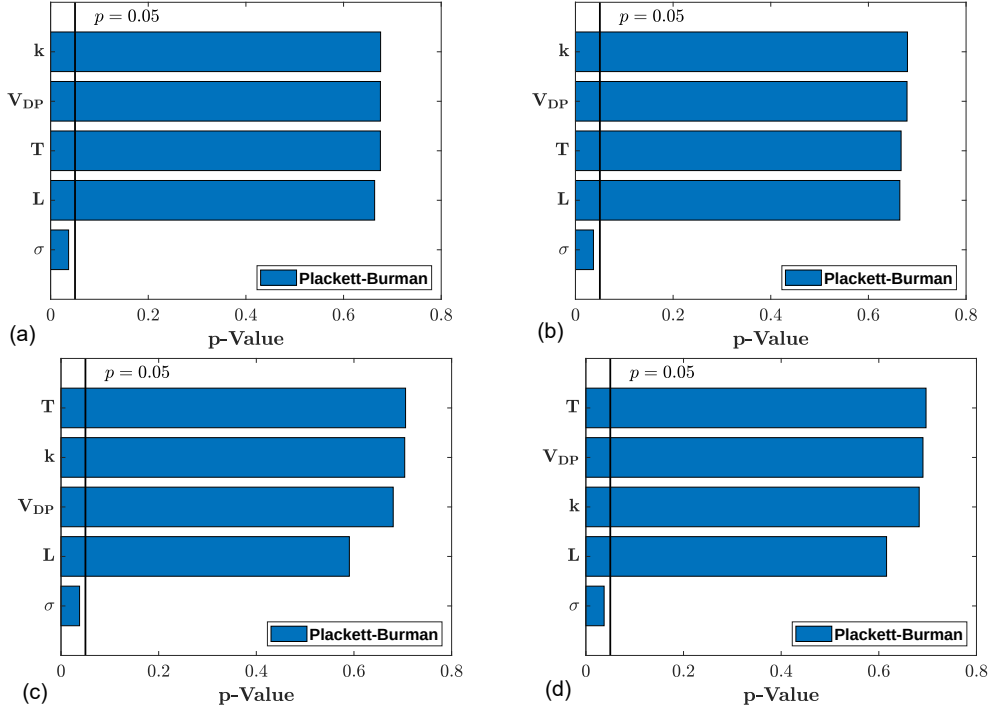


Figure 10: Feature selection results from folded Plackett-Burman design showing the uncertainty parameters to the fitting coefficients (a) w_1 , (b) w_2 , (c) w_3 , (d) w_4 ; the behavior for w_5 is similar (figure not shown). The black line denotes a cut-off for 0.05 probability, reflecting 95% confidence.

square-root-of-time behavior, similar to the trend in diffusion problem [63]. At $t_D = 1$, the percentiles P10, P50, and P90 of F_{TR} are 0.04, 0.11, and 0.18, respectively, and the corresponding F_{EP} 's are 0.87, 0.95, and 0.997. The ANCOVA sensitivity result (Figure 16) shows that permeability and well spacing influence both the thermal recovery factor and enthalpy production factor at all times. For the thermal recovery factor, both permeability and well spacing stay dominant until the end of the process. There is a slight change in sensitivity values after the breakthrough time, but there is no change in the ranking. For the enthalpy production factor, permeability stays dominant at all times. However, there is a change in parameter ranking after the breakthrough time. Dykstra-Parson coefficient and well spacing switch places after fluid breakthrough, which indicates a change in recovery

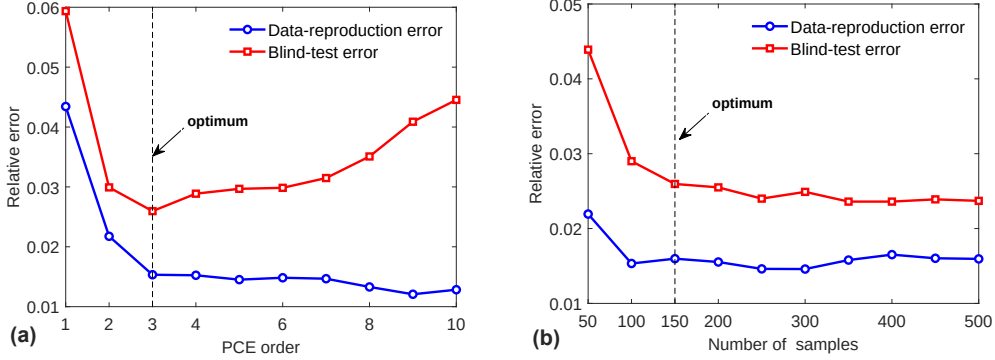


Figure 11: Accuracy of the low-fidelity model as a function of the order of PCE in reproducing the training set (blue line) and in predicting the blind test set (red line) are shown in the left plot(a). The plot on the right (b) shows the accuracy behavior of the low-fidelity model as a function of the size of the design.

mechanisms after the breakthrough time from “heterogeneity-influenced” to “well-spacing-influenced”. We show the response surfaces of the w_i coefficients versus several uncertainty parameters in Figure 13. The surfaces reflect the variability of the coefficients with respect to the uncertainty parameters which show quasi-linear to non-linear behaviors. To capture the significant uncertainty variables, we emphasize the use of ANCOVA in conducting the sensitivity analysis. We demonstrate the use of our workflow to analyze the optimum injector-producer spacing (L) for cases with different reservoir permeabilities (k), as shown in Figure 17. The parameters k and L are selected based on the ANCOVA result. Figure 17a and b show the distributions of $F_{TR}(t_D = 1)$ and $F_{EP}(t_D = 1)$ in terms of the variations in k and L . Since the optimization should be conducted on both functions, $F_{TR}(t_D = 1)$ and $F_{EP}(t_D = 1)$, simultaneously, the optimum well spacing should maximize both F_{TR} and F_{EP} , which can be captured by maximizing an objective function Obj , such that:

$$Obj = w_1 F_{TR} + w_2 F_{EP}, \quad (30)$$

where w_1 and w_2 are weighting factors relative to the field conditions. For instance, selecting $w_1 = 0.4$ and $w_2 = 0.6$ for a reservoir case with $k = 750mD$ ($V_{DP} = 0.55$), the corresponding optimum spacing is $L = 655m$ (see Fig.17c and d). The optimization problem becomes multidimensional when including

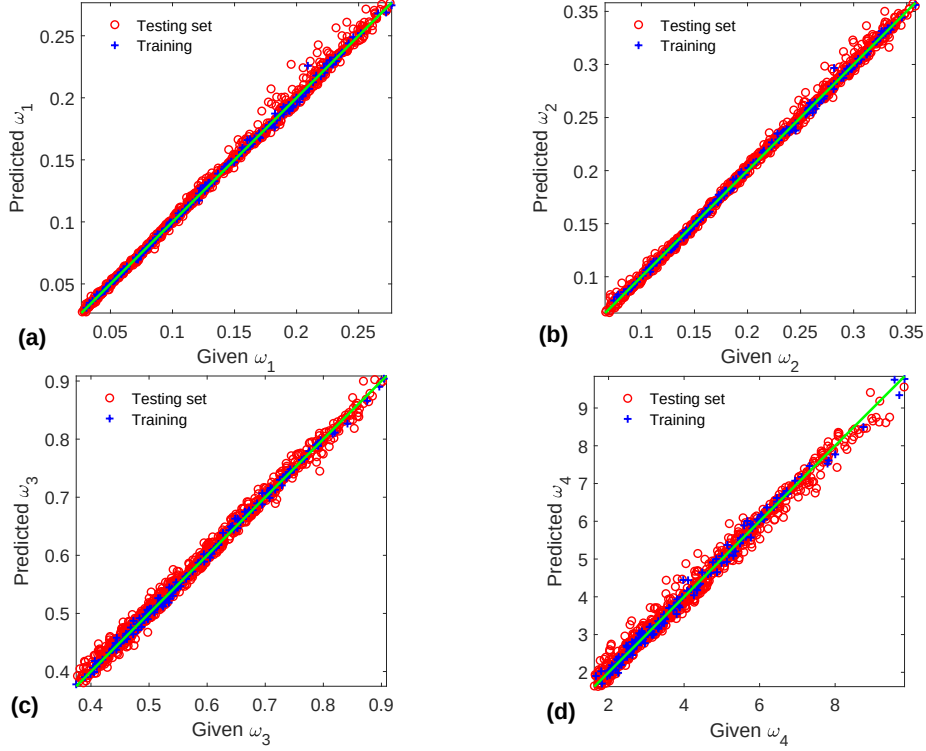


Figure 12: Training accuracy and predictability of the low-fidelity models for the time-independent fitting coefficients including (a) w_1 , (b) w_2 , (c) w_3 , (d) w_4 . Similar behavior was obtained for w_5 ; figure is not shown.

other variables, such as reservoir heterogeneity (V_{DP}). Figure 18, shows the distribution of the objective function in 3D versus the uncertainty variable k , and V_{DP} , and the design variable L . The selected objective function is case dependent, which reflects the desired maximization of F_{TR} and F_{EP} . In this case, the optimum well spacing corresponds to maximum value "Obj" as indicated in Fig. 18.

To demonstrate the applicability of the optimization approach, we applied the model to various geothermal fields considering their temperature, permeability, and heterogeneity ranges [5, 64, 65]. The estimated optimum well-spacing between injectors and producers is compared to the actual average spacing in the fields, as shown in Fig. 19. The predictions are in reasonable agreement with the actual data. However, several exceptions showed significant discrepancies, which could be related to various reasons. For instance,

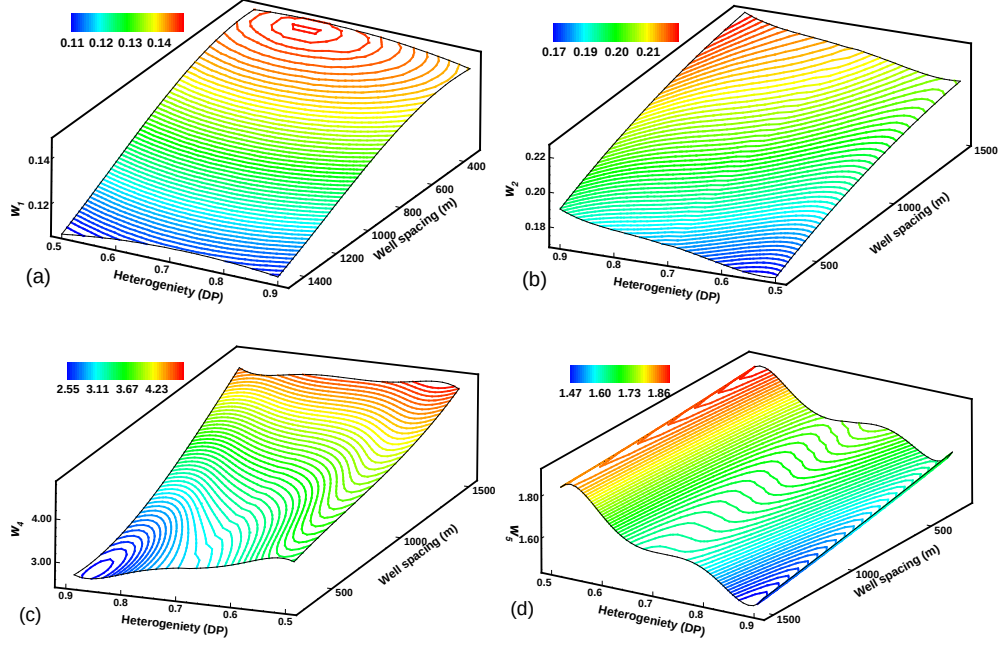


Figure 13: Response surfaces of the time-independent w_i coefficients as functions of the uncertain parameters, where several examples are shown including (a) w_1 , (b) w_2 , (c) w_4 , and (d) w_5 , versus well spacing L and heterogeneity V_{DP} .

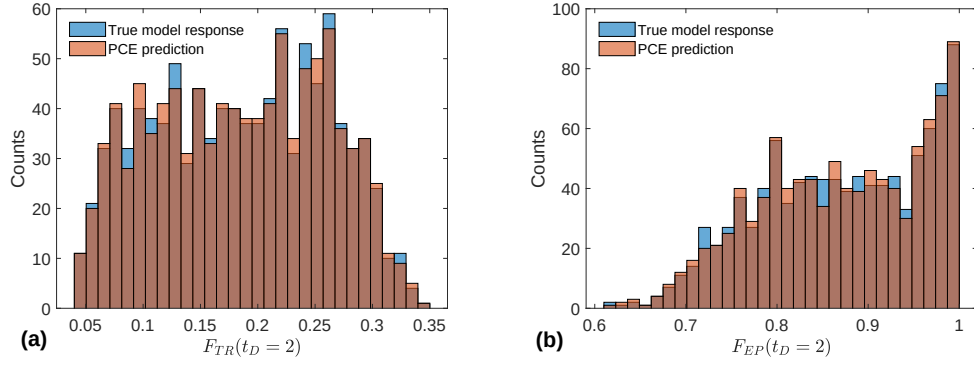


Figure 14: Prediction accuracy of time-dependent objective response (PCE prediction) compared to the blind-testing data including 1000 cases (True model) at $t_D = 2$ corresponding to (a) thermal recovery factor F_{TR} , and (b) enthalpy production factor F_{EP} .

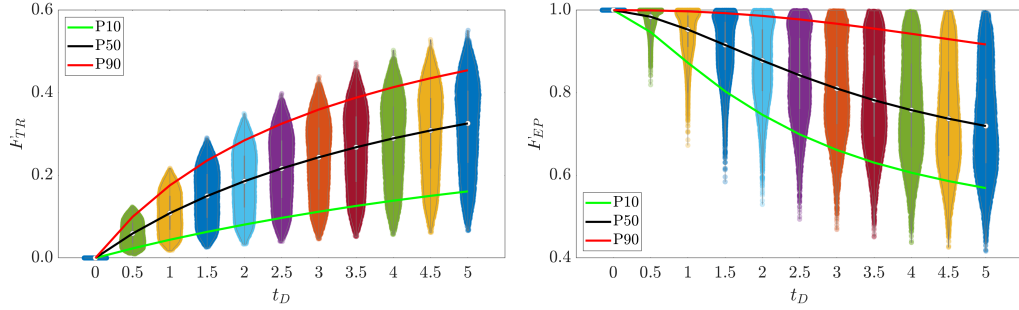


Figure 15: Uncertainty propagation results for the time-dependent objective function F_{TR} (left) and F_{EP} (right) using Monte Carlo simulations applied to the low-fidelity models.

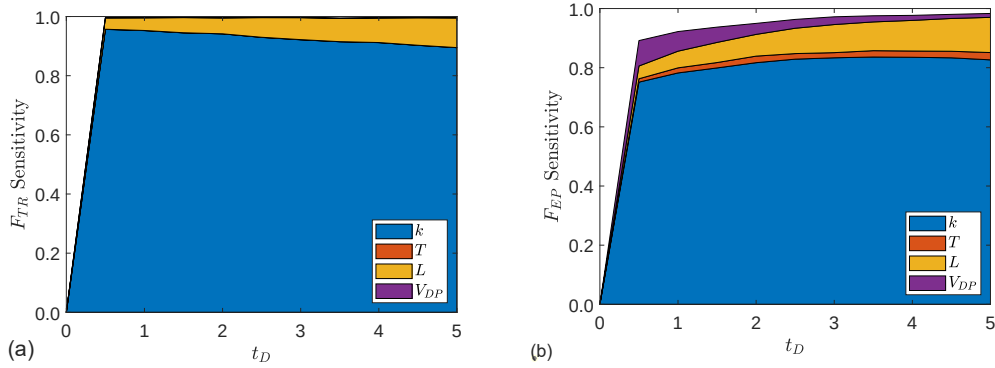


Figure 16: First-order ANCOVA sensitivity indices of the low-fidelity models: F_{TR} (a) and F_{EP} (b).

water re-injection in Dieng field (Java) has experienced mineral deposition at the injection wells, resulting in injectivity loss where the current injection to production volume ratio is about 68% [65]. The actual injection-production well spacing is around 1500m, and with a permeability of 0.1-30 md, temperature range 240-300 °C [66], the calculated well spacing is around 400m. Momotombo field ([64]), on the other hand, has full re-injection capacity to wastewater with a well spacing of 350m, resulting in enthalpy decline [65, 67]. With a permeability range of 1-600md and temperature 180-200 °C ([68]), the corresponding estimated optimum spacing is around 500m, indicating a potentially too tight well spacing in the current well development configuration.

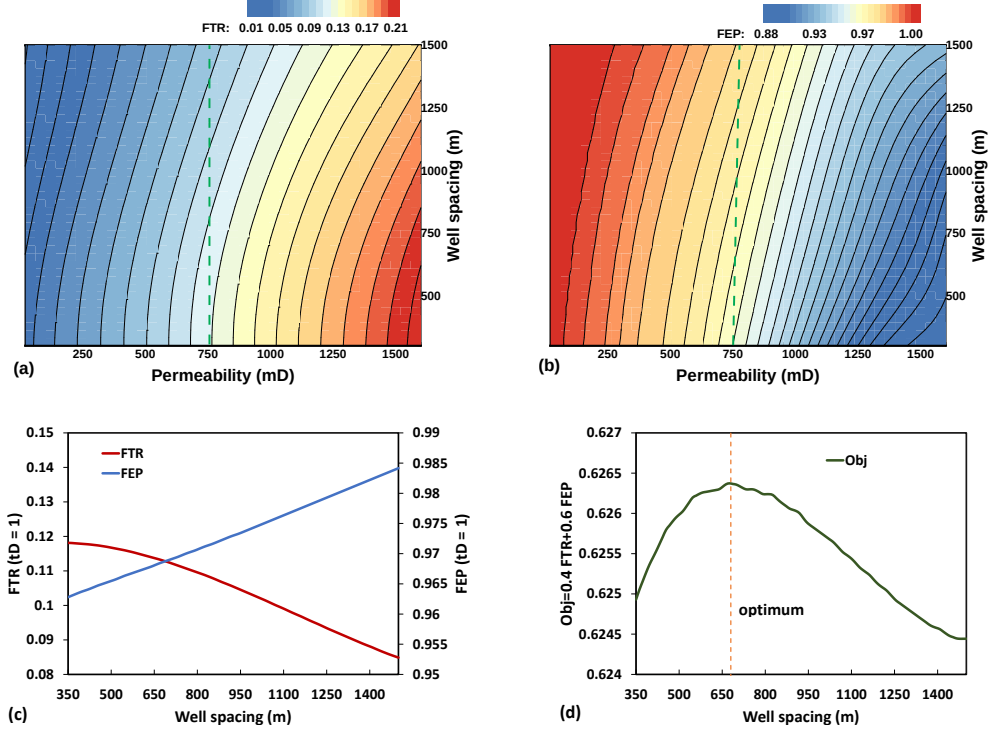


Figure 17: Distribution of F_{TR} (a) and F_{EP} (b) at $t_D = 1$ as a function for reservoir permeability and well spacing showing opposite trends. Extractions of F_{TR} and F_{EP} trends versus well spacing at permeability $k = 750$ mD is shown in (c). A multi-objective optimization given in Eq. 30 shows an optimum well spacing at $L = 655$ m (d).

7 Discussion

The proposed simulation approach utilizes reservoir conditions compiled from worldwide geothermal field data, which may serve as a quick analog to assess field production efficiency. The flow behavior in geothermal reservoir under re-injection can be described by two dimensionless number: Peclet and Rayleigh [69, 70, 71, 72]. The Peclet number describes the significance of convective versus conduction heat transfer, while the Rayleigh number reflects the ratio between buoyant-driven and viscous-driven flow [73].

The feature selection process using folded Plackett Burman shows the insignificance of thermal conductivity, as an uncertainty parameter. The Peclet number within our realizations is in the range of 5 - 110. The reported Peclet

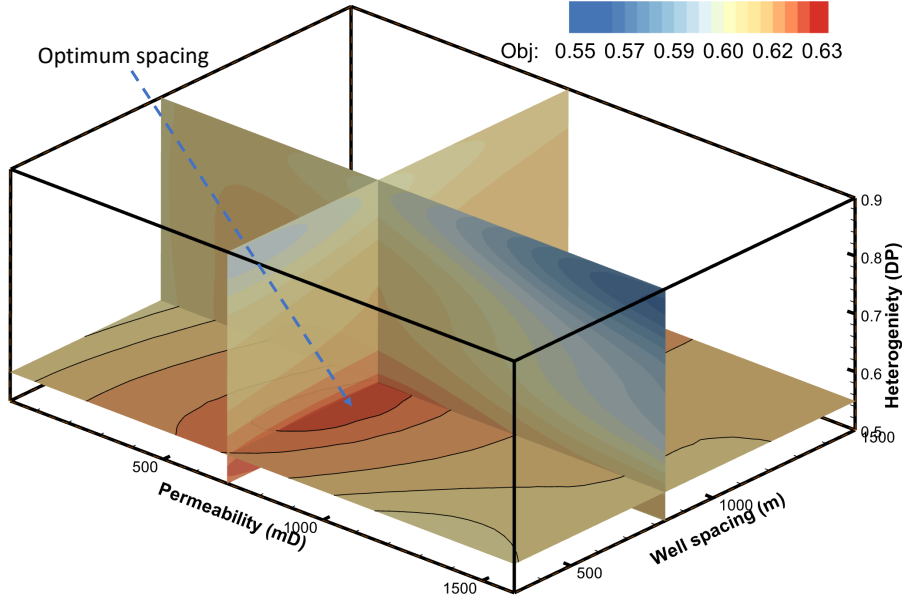


Figure 18: A 3D heat map showing an objective function (Eq. 30) versus reservoir permeability, heterogeneity, and well spacing. The indicated zone corresponds to the optimum well spacing corresponding to a specific reservoir's permeability and heterogeneity condition.

number for tracer injection in geothermal reservoirs is 3 - 80 [74]. It denotes that in most geothermal re-injection process, the dominant heat transfer process is through convection. The dimensionless time term reflects the injected/produced volume relative to the reservoir's pore volume. Therefore, the practical ranges for injection are around $t_D = 1$. The obtained thermal recovery factor at $t_D = 1$ is within 0.05 and 0.18, with the most likely case at 0.11, which is in agreement with the average worldwide thermal recovery of 0.13 - 0.14 [5].

The ANCOVA enables to identify the parameters that govern the physical phenomenon at different times, where the permeability, rate, and well spacing are identified as the most significant parameters over time. This result is supported by the obtained Rayleigh number in our realizations within the range of 1 - 440. The reported Rayleigh number for geothermal reservoirs is less than 500 [75, 76, 73, 77]. Furthermore, in a buoyant-dominated setting, permeability, rate, and well spacing are to trigger the convection [78]. Therefore, our results are in agreement with field observations, which indicate

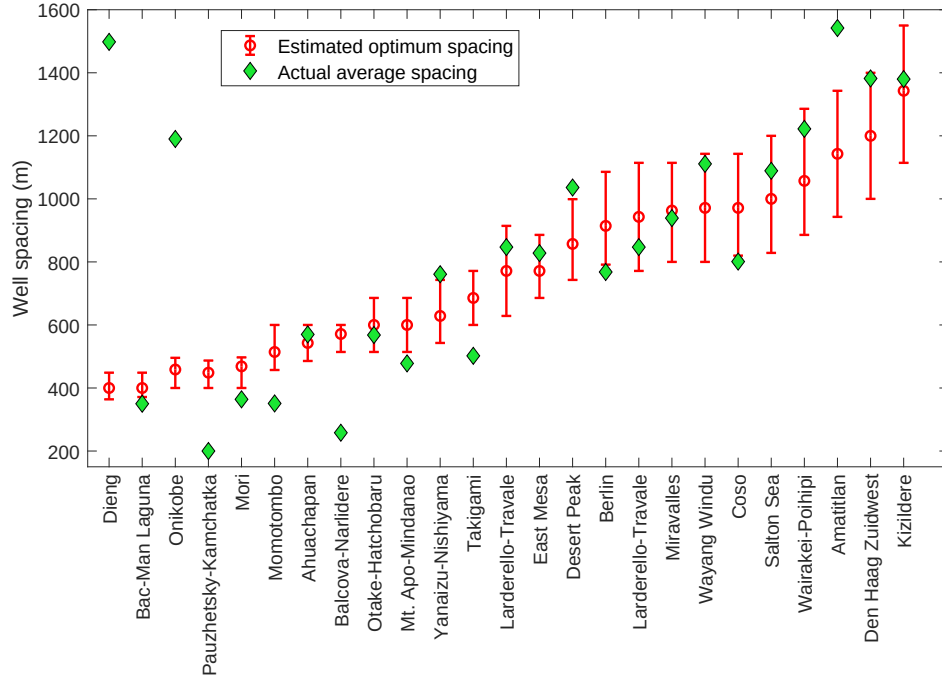


Figure 19: Computed ranges of optimum well spacings for various geothermal fields and compared the to actual average spacing in the fields .

that flow in a geothermal reservoir during re-injection is gravity-dominated. The workflow provides a novel insight for constructing continuous space-time low-fidelity models for geothermal-related problems. It is useful to accelerate complex and expensive computations [79, 80] and perform efficient optimization and history matching [81, 82, 83, 15]. However, this model does not capture the effect of detailed mechanisms such as injectivity loss from mineral precipitation or other field specifics such as the economic aspects, including the minimum operational production rates and enthalpy.

8 Conclusions

In this work, we present a new workflow to develop time-continuous, multi-objectives uncertainty quantification of geothermal production under re-injection. We identify the parameters and their ranges using a database reflecting multiple analogs from literature. The feature selection step is to select the sig-

nificant parameters utilizing a simple design. The time-continuous response is constructed using a nested-function approach and optimized under multi-objectives setting. We also propose the use of ANCOVA indices to perform global sensitivity analysis with the Design of Experiment (DoE). Emphasis should be given to assess the proxy predictability Q^2 , and not to solely rely on training accuracy R^2 to avoid overfitting. Improved predictability is achieved using space-filling design with an optimum number of realizations.

The feature selection results showed that thermal conductivity is insignificant to the re-injection process. The heat transfer process is dominated by convection (driven by fluid flow) with a high value of Peclet number. The thermal recovery and enthalpy production efficiency are significantly influenced by permeability, rate, porosity, and well spacing. As the flow is gravity-dominated, the Rayleigh number is relatively high. Monte Carlo simulations show that the most likely recoverable heat under re-injection is 11%. Such a reservoir is likely to operate within a 95% enthalpy production factor, which is in agreement with the reported geothermal recoveries under injection. The optimization workflow was applied to several field cases, which provided insights about the estimated well spacing versus the actual well configurations, and therefore highlight opportunities for field optimization. The proposed approach could also be used for screening of new field developments, whose initial parameters will have the biggest impact on how much would lowering uncertainty improve optimization of recovery.

Aknowledgments

We would like to thank CMG Ltd. for providing the STARS academic license, KAUST for the support, and UQLab for the software’s license.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A Sensitivity analysis

We discuss the effect of 1) different 2-level DoE's on feature selection, 2) different 3-level DoE on the construction of the low-fidelity models and uncertainty propagation, and 3) the effect of the number of different realizations on the accuracy of Latin Hypercube DoE.

A.1 Sensitivity of different 2-level DoEs

We use 2-level full factorial design, 2-level fractional factorial design, and folded Plackett-Burman design to investigate the process accuracy for feature selection. These designs inherently assume uniform probability for every realization [84], and use first-order regression. Therefore, we only quantify the effect of the linear terms associated with the first-order sensitivity. The full factorial design samples the low and high values within the range of uncertainty of each variable, following the permutation rule. For d variable, the total number of cases is $N_s = 2^d$. The fractional factorial design performs a similar sampling as the full factorial design but considers a resolution optimization to reduce the total number cases. On the other hand, the folded Plackett-Burman design samples some of the low and high sides of each variable, such that the total number of cases is $N_s = 4i, i=1,2,3,\dots$. The corresponding maximum number of variables for each case is $d = 4i - 1$. Figure 20 shows consistent ranking of the assessed variables, which confirms the robustness of the folded Plackett-Burman design. This design is recommended for variable ranking and feature selection. It requires a smaller number of cases than the full factorial design with comparable accuracy.

A.2 Sensitivity of different 3-level DoEs

We use 3-level Full Factorial, Central Composite, Box-Behnken, D-Optimal, Latin Hypercube, and Taguchi designs to investigate the effect of sampling and training on the reliability of the low-fidelity models, including predictability and uncertainty propagation. All the designs assume uniform probability for every realization [84]. The 3-level Full Factorial design requires $N_s = 3^d$ cases. The Central Composite design provides the same sample points as in Full Factorial design, added with the center point and augmented with a group of axial points. The axial points represent extreme values to allow estimation of the curvature. The total number of cases becomes $N_s = (2 \times 3^d) + 2$.

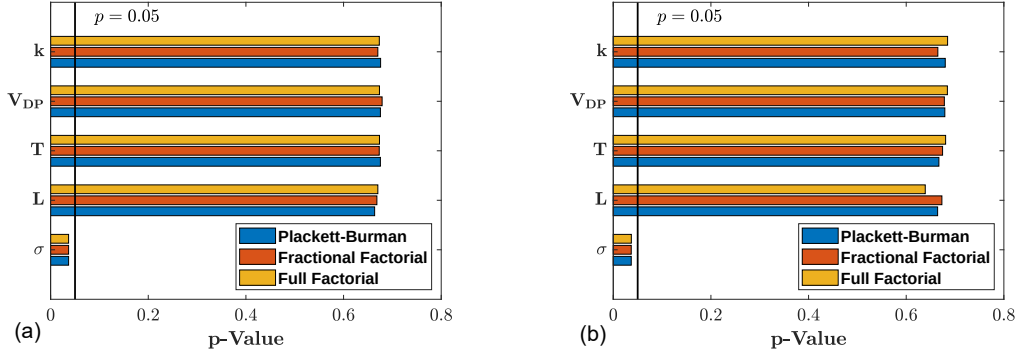


Figure 20: Selection of significant variables relative to w_1 (a) and w_2 (b), using the Plackett-Berman DoE, and compared with other designs for confirmation. The black line denotes a cut-off of 0.05 probability (95% confidence). The other (w_3 , w_4 , and w_5) showed similar behavior and therefore not shown.

The additional two cases correspond to the center and the axial points. Box-Behnken design includes the midpoint of the edges and the center point, while the D-Optimal design includes sample points by optimizing a D-optimality criterion. The Latin Hypercube design creates realizations by maximizing the distance among the samples, while the Taguchi design utilizes special orthogonal arrays to generate a minimal number of realizations to capture full information of all variables. The summary of the total number of realizations, training accuracy, and predictability for the coefficient low-fidelity models is shown in Figure 21. Achieving high predictability is more important than high training accuracy since high predictability enables accurate interpolation for the prediction. For instance, Taguchi design showed the highest training accuracy but produced the lowest predictability. It cannot provide accurate predictions for unknown points. Figure 21 shows that higher predictability Q^2 is achieved by increasing the total number of realizations under random and space-filling conditions. Randomization is critical to tackle unknown non-linearity within the model and correlation among parameters, while space-fillingness to obtain an equal response from the whole region [85]. Full Factorial, D-Optimal, and Taguchi design sample extreme/corner points. Therefore, they have low predictability compared to the other random and space-filling designs. Full Factorial has more realizations than Taguchi and D-Optimal, which gives higher predictability. Central composite de-

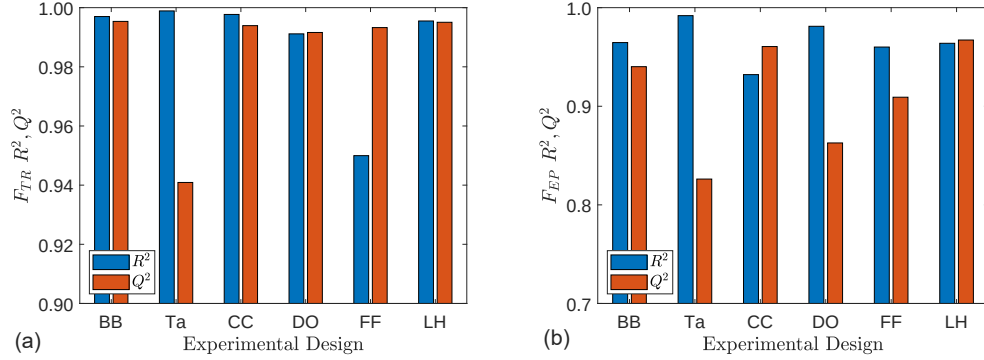


Figure 21: Accuracy of different DoE evaluated by average R^2 and Q^2 obtained for the thermal recovery factor F_{TR} (a) and enthalpy production factor F_{EP} (b). The corresponding symbols: FF is Full Factorial, CC is Central Composite, BB is Box Behken, DO is D-Optimal, LH is Latin Hypercube, and Ta is Taguchi. The number of realizations in the different designs are: BB=27, Ta=28, CC=36, DO=74, FF=81, LH = 1000.

sign is more space-filling than Box-Behken design since it is Full Factorial design equipped with axial points [14]. Therefore, Central Composite design gives higher predictability than Box-Behken design. Latin Hypercube has both random and space-filling properties, therefore, it gives the highest predictability.

A.3 Sensitivity of the size of Latin Hypercube design

We perform sensitivity of a different number of realizations to the Latin Hypercube design. The aim is to find the optimum condition which gives high predictability and small realizations. The comparison result for Monte Carlo simulations is shown in Figure 22. Change in the number of realizations gives a slightly different distribution in Monte Carlo results, as shown in Figure 22. All tested cases with different number of realizations produced consistent parameter ranking, which demonstrates the robustness of the method.

References

- [1] BP. BP Statistical Review of World Energy 2019. Technical report, British Petroleum, 2019.

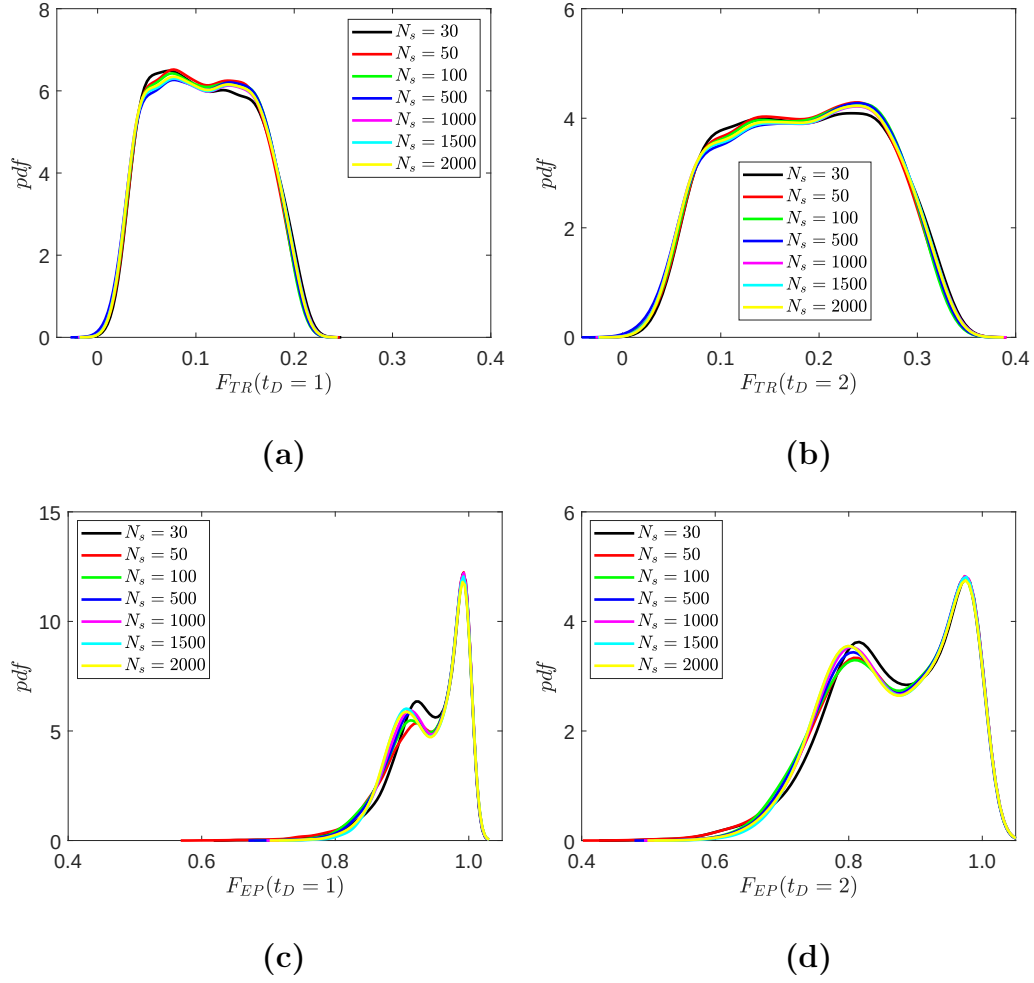


Figure 22: Monte Carlo results using different number of realizations on Latin Hypercube design: $F_{TR}(t_D = 1)$ (a), $F_{TR}(t_D = 2)$ (b), $F_{EP}(t_D = 1)$ (c), and $F_{EP}(t_D = 2)$ (d). The plots are produced by applying Kernel smoothing.

- [2] IEA. Market Report Series: Renewables 2019. Technical report, International Energy Agency, 2019.
- [3] Subir K Sanyal. CLASSIFICATION OF GEOTHERMAL SYSTEMS-A POSSIBLE SCHEME. In *Thirtieth Workshop on Geothermal Reservoir Engineering*, pages 1–8, Stanford, California, USA, 2005. Stanford University.

- [4] Inga S. Moeck. Catalog of geothermal play types based on geologic controls. *Renewable and Sustainable Energy Reviews*, 37:867–882, sep 2014.
- [5] Sadiq J. Zarrouk and Hyungsul Moon. Efficiency of geothermal power plants: A worldwide review. *Geothermics*, 51:142–153, jul 2014.
- [6] Egill Juliusson and Roland N. Horne. Optimization of injection scheduling in fractured geothermal reservoirs. *Geothermics*, 48:80–92, 2013.
- [7] Ryan Santoso, Hussein Hoteit, and Volker Vahrenkamp. Optimization of Energy Recovery from Geothermal Reservoirs Undergoing Re-Injection: Conceptual Application in Saudi Arabia. In *SPE Middle East Oil and Gas Show and Conference*, volume March 20, 2019, 03 2019. D031S027R003.
- [8] Daniel T. Birdsell, Benjamin M. Adams, and Martin O. Saar. Minimum transmissivity and optimal well spacing and flow rate for high-temperature aquifer thermal energy storage. *Applied Energy*, 289:116658, 2021.
- [9] Benjamin M. Adams, Thomas H. Kuehn, Jeffrey M. Bielicki, Jimmy B. Randolph, and Martin O. Saar. On the importance of the thermosiphon effect in cpg (co2 plume geothermal) power systems. *Energy*, 69:409–418, 2014.
- [10] Benjamin M. Adams, Thomas H. Kuehn, Jeffrey M. Bielicki, Jimmy B. Randolph, and Martin O. Saar. A comparison of electric power output of co2 plume geothermal (cpg) and brine geothermal systems for varying reservoir conditions. *Applied Energy*, 140:365–377, 2015.
- [11] Justin Ezekiel, Anozie Ebigbo, Benjamin M. Adams, and Martin O. Saar. Combining natural gas recovery and co2-based geothermal energy extraction for electric power generation. *Applied Energy*, 269:115012, 2020.
- [12] Jeoung Seok Yoon, Arno Zang, and Ove Stephansson. Numerical investigation on optimized stimulation of intact and naturally fractured deep geothermal reservoirs using hydro-mechanical coupled discrete particles joints model. *Geothermics*, 52:165–184, 2014. Analysis of Induced Seismicity in Geothermal Operations.

- [13] S.N. Pandey, A. Chaudhuri, and S. Kelkar. A coupled thermo-hydro-mechanical modeling of fracture aperture alteration and reservoir deformation during heat extraction from a geothermal reservoir. *Geothermics*, 65:17–31, 2017.
- [14] Esmail Ansari, Richard Hughes, and Christopher D. White. Statistical modeling of geopressured geothermal reservoirs. *Computers and Geosciences*, 103:36–50, jun 2017.
- [15] Daniel O. Schulte, Dan Arnold, Sebastian Geiger, Vasily Demyanov, and Ingo Sass. Multi-objective optimization under uncertainty of geothermal reservoirs using experimental design-based proxy models. *Geothermics*, 86:101792, jul 2020.
- [16] Egill Juliusson and Sigurdur Bjornsson. Optimizing production strategies for geothermal resources. *Geothermics*, 94:102091, 2021.
- [17] Philippe Pasquier, Angelo Zarrella, and Denis Marcotte. A multi-objective optimization strategy to reduce correlation and uncertainty for thermal response test analysis. *Geothermics*, 79:176–187, 2019.
- [18] Amir H. Keshavarzzadeh, Atefeh Maleki Zanjani, Kobra Gharali, and Maurice B. Dusseault. Multi-objective evolutionary-based optimization of a ground source heat exchanger geometry using various optimization techniques. *Geothermics*, 86:101861, 2020.
- [19] Dongkyu Park, Eunhee Lee, Dugin Kaown, Seong-Sun Lee, and Kang-Kun Lee. Determination of optimal well locations and pumping/injection rates for groundwater heat pump system. *Geothermics*, 92:102050, 2021.
- [20] Daniel O. Schulte, Dan Arnold, Sebastian Geiger, Vasily Demyanov, and Ingo Sass. Multi-objective optimization under uncertainty of geothermal reservoirs using experimental design-based proxy models. *Geothermics*, 86:101792, 2020.
- [21] Jeremy R. Patterson, Michael Cardiff, and Kurt L. Feigl. Optimizing geothermal production in fractured rock reservoirs under uncertainty. *Geothermics*, 88:101906, 2020.

- [22] P. Salinas, G. Regnier, C. Jacquemyn, C.C. Pain, and M.D. Jackson. Dynamic mesh optimisation for geothermal reservoir modelling. *Geothermics*, 94:102089, 2021.
- [23] Facheng Gong, Tiankui Guo, Wei Sun, Zhaomin Li, Bin Yang, Yimei Chen, and Zhanqing Qu. Evaluation of geothermal energy extraction in enhanced geothermal system (egs) with multiple fracturing horizontal wells (mfhw). *Renewable Energy*, 151:1339–1351, 2020.
- [24] Tiankui Guo, Songjun Tang, Jiang Sun, Facheng Gong, Xiaoqiang Liu, Zhanqing Qu, and Wei Zhang. A coupled thermal-hydraulic-mechanical modeling and evaluation of geothermal extraction in the enhanced geothermal system based on analytic hierarchy process and fuzzy comprehensive evaluation. *Applied Energy*, 258:113981, 2020.
- [25] Leo W. T. Ng and Karen E. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, 100(10):746–772, dec 2014.
- [26] A C Antoulas, D C Sorensen, and S Gugercin. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, 280, 2001.
- [27] G. Rozza, D. B.P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: Application to transport and continuum mechanics. *Archives of Computational Methods in Engineering*, 15(3):229–275, sep 2008.
- [28] Qian Shao, Anis Younes, Marwan Fahs, and Thierry A. Mara. Bayesian sparse polynomial chaos expansion for global sensitivity analysis. *Computer Methods in Applied Mechanics and Engineering*, 318:474–496, 2017.
- [29] Renee Swischuk, Laura Mainini, Benjamin Peherstorfer, and Karen Willcox. Projection-based model reduction: Formulations for physics-based machine learning. *Computers and Fluids*, 179:704–717, jan 2019.
- [30] L. Guo, M. Fahs, H. Hoteit, R. Gao, and Q. Shao. Uncertainty analysis of seepage-induced consolidation in a fractured porous medium. *Computer Modeling in Engineering and Sciences*, 129(1):279–297, 2021.

- [31] Renee Swischuk, Boris Kramer, Cheng Huang, and Karen Willcox. Learning physics-based reduced-order models for a single-injector combustion process. Technical report, 2019.
- [32] Ryan Santoso, Xupeng He, and Hussein Hoteit. Application of Machine-Learning to Construct Simulation Models from High-Resolution Fractured Formation. In *Abu Dhabi International Petroleum Exhibition and Conference*, volume November 12, 2019, 11 2019. D021S060R004.
- [33] Xupeng He, Ryan Santoso, and Hussein Hoteit. Application of Machine-Learning to Construct Equivalent Continuum Models from High-Resolution Discrete-Fracture Models. In *IPTC International Petroleum Technology Conference*, volume January 15, 2020, 01 2020. D031S075R003.
- [34] Abdirizak Omar, Mouadh Addassi, Volker Vahrenkamp, and Hussein Hoteit. Co-optimization of co2 storage and enhanced gas recovery using carbonated water and supercritical co2. *Energies*, 14(22), 2021.
- [35] Xupeng He, Weiwei Zhu, Ryan Santoso, Marwa Alsinan, Hyung Kwak, and Hussein Hoteit. CO2 Leakage Rate Forecasting Using Optimized Deep Learning. In *SPE Annual Technical Conference and Exhibition*, volume September 22, 2021, 09 2021. D021S023R001.
- [36] F Friedmann, A Chawathé, and D K Larue. Assessing Uncertainty in Channelized Reservoirs Using Experimental Designs. *SPE Reservoir Evaluation & Engineering*, pages 264–274, aug 2003.
- [37] O. P. Le Maitre and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer, New York, 2010.
- [38] Elvind Damsleth, Asmund Hage, and Rolf Volden. Maximum information at minimum cost: A north sea field development study with an experimental design. *SPE Economics and Management*, 3(2):110–117, apr 2011.
- [39] Mohammad Mahdi Rajabi, Marwan Fahs, Aref Panjehfouladgaran, Behzad Ataie-Ashtiani, Craig T. Simmons, and Benjamin Belfort. Uncertainty quantification and global sensitivity analysis of double-diffusive

- natural convection in a porous enclosure. *International Journal of Heat and Mass Transfer*, 162:120291, dec 2020.
- [40] Ryan Santoso, Victor Torrealba, and Hussein Hoteit. Investigation of an improved polymer flooding scheme by compositionally-tuned slugs. *Processes*, 8(2):197, 2020.
 - [41] H. Hoteit and M. Addassi. Integrated uncertainty quantification for reactive transport modeling of co2 mineralization in basalts. In *Third EAGE Geochemistry Workshop*, volume 2021, pages 1–5, 2022.
 - [42] Francisco Chinesta, Antonio Huerta, Gianluigi Rozza, and Karen Willcox. Model Order Reduction. *Encyclopedia of Computational Mechanics*, pages 1–59, 2004.
 - [43] Jaime Jose D. Quinao and Sadiq J. Zarrouk. Geothermal resource assessment using Experimental Design and Response Surface Methods: The Ngatamariki geothermal field, New Zealand. *Renewable Energy*, 116:324–334, 2018.
 - [44] Heru Berian Pratama, Marchel Christian Supijo, and Sutopo. Experimental design and response surface method in geothermal energy: A comprehensive study in probabilistic resource assessment. *Geothermics*, 87:101869, sep 2020.
 - [45] S. N. Pandey and Vikram Vishal. Sensitivity analysis of coupled processes and parameters on the performance of enhanced geothermal systems. *Scientific Reports*, 7(1):1–14, dec 2017.
 - [46] Alexandros Daniilidis, Hamidreza M. Nick, and David F. Bruhn. Interdependencies between physical, design and operational parameters for direct use geothermal heat in faulted hydrothermal reservoirs. *Geothermics*, 86:101806, jul 2020.
 - [47] Marta Fernández, Pablo Eguía, Enrique Granada, and Lara Febrero. Sensitivity analysis of a vertical geothermal heat exchanger dynamic simulation: Calibration and error determination. *Geothermics*, 70:249–259, nov 2017.

- [48] Yann Caniou. *Global sensitivity analysis for nested and multiscale modelling*. PhD thesis, Universite Blaise Pascal et Institut Francais de Mecanique Avancee, 2012.
- [49] R. Bird, W. Stewart, and E. Lightfoot. *Transport Phenomena*. Wiley, New York, 2nd edition, 2006.
- [50] M Grant and P Bixley. *Geothermal Reservoir Engineering*. Academic Press, Amsterdam, 2nd edition, 2011.
- [51] CMG. *STARS User Guide*. Computer Modeling Group, 2016.
- [52] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52(1):1–17, apr 1996.
- [53] Stefano Marelli and Bruno Sudret. UQLab: A Framework for Uncertainty Quantification in MATLAB. In *The 2nd International Conference on Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management - Proceedings of the 2nd International Conference on Vulnerability and Risk Analysis and Management, ICVRAM 2014 and the 6th International*, pages 2554–2563, University of Liverpool, United Kingdom, 2014. American Society of Civil Engineers (ASCE).
- [54] Subir K Sanyal and Steven J Butler. An Analysis of Power Generation Prospects from Enhanced Geothermal Systems. In *World Geothermal Congress*, pages 24–29, Antalya, Turkey, apr 2005.
- [55] Tzu-Rong Su, Tsui-Hsien Huang, Chia-Tze Kao, Hooi Yee Ng, Yung-Cheng Chiu, and Tuan-Ti Hsu. The Calcium Channel Affect Osteogenic Differentiation of Mesenchymal Stem Cells on Strontium-Substituted Calcium Silicate/Poly- ϵ -Caprolactone Scaffold. *Processes*, 8(2):198, feb 2020.
- [56] Sayantan Ganguly, Lippong Tan, Abhijit Date, and Mandalagiri Kumar. Numerical Investigation of Temperature Distribution in a Confined Heterogeneous Geothermal Reservoir Due to Injection-production. In *Energy Procedia*, volume 110, pages 143–148. Elsevier Ltd, mar 2017.

- [57] Xupeng He, Marwa Sinan, Hyung Kwak, and Hussein Hoteit. A corrected cubic law for single-phase laminar flow through rough-walled fractures. *Advances in Water Resources*, 154:103984, 2021.
- [58] Behshad Koohbor, Marwan Fahs, Hussein Hoteit, Joanna Doummar, Anis Younes, and Benjamin Belfort. An advanced discrete fracture model for variably saturated flow in fractured porous media. *Advances in Water Resources*, 140:103602, 2020.
- [59] Hussein Hoteit and Abbas Firoozabadi. An efficient numerical model for incompressible two-phase flow in fractured media. *Advances in Water Resources*, 31(6):891–905, 2008.
- [60] Wijnb Sommer, Johan Valstar, Pauline van Gaans, Tim Grotenhuis, and Huub Rijnaarts. The impact of aquifer heterogeneity on the performance of aquifer thermal energy storage. *Water Resources Research*, 49(12):8128–8138, 2013.
- [61] P. C. Carman. Fluid flow through granular beds. *Chemical Engineering Research and Design*, 75:S32–S48, dec 1997.
- [62] Alejandro Cardona and J. Carlos Santamarina. Carbonate rocks: Matrix permeability estimation. *AAPG Bulletin*, 103(1):131–144, jan 2020.
- [63] Tad W. Patzek, Frank Male, and Michael Marder. Gas production in the Barnett Shale obeys a simple scaling theory. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):19731–19736, dec 2013.
- [64] Alexandre Diaz, Eylem Kaya, and Sadiq J. Zarrouk. Reinjection in geothermal fields: A worldwide review update. *Renewable and Sustainable Energy Reviews*, 53:105–162, 2016.
- [65] Zahratul Kamila, Eylem Kaya, and Sadiq J. Zarrouk. Reinjection in geothermal fields: An updated worldwide review 2020. *Geothermics*, 89:101970, 2021.
- [66] Takeshi Hino, Ryuichi Itoi, Toshiaki Tanaka, Agung Pambudi, and Khasani. Natural state modeling of geothermal reservoir at dieng, central java, indonesia. In *GRC Transactions*, volume 37, Jan. 2013.

- [67] Dennis Kaspereit, William Rickard, William Osborn, Mary Mann, and Magdalena Perez. Field management and expansion potential of the momotombo geothermal field using numerical simulation and conceptual modeling. In *41st Workshop on Geothermal Reservoir Engineering*. Stanford University, Feb. 2016.
- [68] Enrique Porras, Toshiaki Tanaka, Hikari Fujii, and Ryuichi Itoi. Numerical modeling of the momotombo geothermal system, nicaragua. In *World Geothermal Congress 2005*. Turkey, Apri 2005.
- [69] V. Prasad, F. C. Lai, and F. A. Kulacki. Mixed convection in horizontal porous layers heated from below. *Journal of Heat Transfer*, 110(2):395–402, may 1988.
- [70] Ekkehard O. Holzbecher. *Modeling Density-Driven Flow in Porous Media*. Springer Berlin Heidelberg, 1998.
- [71] Martin O. Saar. Review: Geothermal heat as a tracer of large-scale groundwater flow and as a means to determine permeability fields. *Hydrogeology Journal*, 19(1):31–52, nov 2011.
- [72] Donald A. Nield and Adrian Bejan. *Convection in porous media*. Springer New York, 4th edition, jan 2013.
- [73] R. N. Horne and M. J. O’sullivan. Oscillatory convection in a porous medium heated from below. *Journal of Fluid Mechanics*, 66(2):339–352, 1974.
- [74] Gary D Lecain, Lawrence O Anna, and Michael F Fahy. Results from geothermal logging, air and core-water chemistry sampling, air-injection testing, and tracer testing in the northern Ghost Dance Fault, Yucca Mountain, Nevada, November 1996 to August 1998. Technical report, 2000.
- [75] I. G. Donaldson. Temperature gradients in the upper layers of the Earth’s crust due to convective water flows. *Journal of Geophysical Research*, 67(9):3449–3459, aug 1962.
- [76] I. G. Donaldson. The simulation of geothermal systems with a simple convective model. *Geothermics*, 2(PART 1):649–654, jan 1970.

- [77] M.L. Sorey. Numerical modeling of liquid geothermal systems. Technical Report 1044 -D, USDOE Office of Energy Efficiency and Renewable Energy Geothermal Tech Pgm, jan 1978.
- [78] F. C. Lai and F. A. Kulacki. Oscillatory mixed convection in horizontal porous layers locally heated from below. *International Journal of Heat and Mass Transfer*, 34(3):887–890, mar 1991.
- [79] Jincong He, Pallav Sarma, and Louis J. Durlofsky. Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Computers and Geosciences*, 55:54–69, jun 2013.
- [80] Meng Tang, Yimin Liu, and Louis J. Durlofsky. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413:109456, jul 2020.
- [81] Youssef M. Marzouk, Habib N. Najm, and Larry A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, jun 2007.
- [82] Elnur Aliyev and Louis J. Durlofsky. Multilevel Field Development Optimization Under Uncertainty Using a Sequence of Upscaled Models. *Mathematical Geosciences*, 49(3):307–339, apr 2017.
- [83] Jayanth Jagalur-Mohan, Birendra Jha, Zheng Wang, Ruben Juanes, and Youssef Marzouk. Inferring Fault Frictional and Reservoir Hydraulic Properties From Injection-Induced Seismicity. *Geophysical Research Letters*, 45(3):1313–1320, feb 2018.
- [84] Heng Li, Pallav Sarma, and Dongxiao Zhang. A comparative study of the probabilistic-collocation and experimental-design methods for petroleum-reservoir uncertainty quantification. *SPE Journal*, 16(2):429–439, jun 2011.
- [85] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer New York, 2003.