Proceedings World Geothermal Congress 2023
Beijing, China, April 17 – 21, 2023

Robust Numerical Modeling for Geothermal Reservoirs Based on 1D Well Model and Machine Learning Techniques

Zhihong Zhao[a,*], Jiacheng Wang[a], Jinfan Chen[a], Jie He[a], Xianfeng Tan[b]

[a]Department of Civil Engineering, Tsinghua University, Beijing, China

[b]Shandong Provincial Lunan Geo-engineering Exploration Institute, Jining, Shandong, China

*zhzhao@tsinghua.edu.cn

Keywords: Geothermal reservoir, Integrated numerical model, Simplified well model, Machine Learning

ABSTRACT

An efficient numerical model, which can simulate the coupled fluid flow and heat transfer processes in geothermal reservoirs, is essentially required to evaluate the fate of geothermal wells and the performances of geothermal reservoirs in response to the long-term well operations. In this numerical study, a surrogate model based on the machine learning technique was developed to reduce the substantial computational burden of forward simulations, which was then combined with genetic algorithm to develop a robust optimization approach for operational parameters of geothermal doublets in heterogeneous geothermal reservoirs. A bench mark example was finally presented to demonstrate the robustness and accuracy of the developed numerical methods.

1. INTRODUCTION

Deep geothermal energy is a promising low-carbon alternative of domestic heating compared to the conventional hydrocarbon-based methods. Sustainable development of deep geothermal energy is an emerging paradigm, the challenges of which involve reducing negative environmental impacts, whilst simultaneously increasing energy access, affordability, security and efficiency of energy use. Numerical modeling, which is capable of simulating the coupled fluid flow and heat transfer processes in geothermal reservoirs, has been widely used to evaluate the recoverable geothermal energy (Liu et al. 2019, 2022). Combining sensitivity analysis with numerical modeling is increasingly adopted to optimize operational parameters such as well position, injection and production rates, and injection temperature (Liang et al. 2018). However, this simulation-based optimization method usually requires a large number of intensive forward simulations to evaluate the reservoir performance considering every possible well configuration, which is time-consuming (Wang et al. 2022). To reduce the substantial computational burden of numerical model of geothermal reservoirs, machine learning (ML) techniques such as artificial neural network, decision tree, and random forest have been recently used to robustly predict the reservoir performance subjected to varying operational parameters (Moraga et al. 2022; Okoroafor et al. 2022).

The main aim of this numerical study is to develop a robust optimization method to determine the optimal operational parameters in heterogeneous geothermal reservoirs. An economical objective function was first defined to maximize the profit over the lifetime of geothermal wells. Based on ML techniques, surrogate models were constructed and selected to predict the reservoir performance under the possible operational parameters. The robust optimization method for operational parameters was formulated by combining the genetic algorithm method with the constructed surrogate model. A bench mark example was finally presented to demonstrate the robustness and accuracy of the developed numerical methods.

2. METHODOLOGY

2.1 Objective function

Recoverable geothermal energy means the amount of geothermal resource that can be recovered in a sustainable way with current technology, and the standards to ensure the sustainable exploitation of geothermal energy were defined in Liu et al. (2022). In this study, the benefit from geothermal development and the cost due to environmental impacts (e.g., hydraulic head drawdown) are considered, and the following objective of optimizing operational parameters is defined to maximize the economic profit (P) over the lifetime of geothermal wells (e.g., N years),

$$P = \sum_{t=0}^{N}\left(P_T - C_P\right) \tag{1a}$$

$$P_T = \eta\left(\rho_f C_f q \Delta T\right) p_h \tag{1b}$$

$$C_P = q \Delta P p_e \tag{1c}$$

where P_T is the profit from geothermal energy development, C_P is the cost of electricity due to hydraulic head drawdown, η is the efficiency with which the heat energy can be used, ρ_f is the water density, C_f is the water heat capacity, q is the injection and production rate, ΔT is the temperature change between produced and injected water, p_h is the heat price, ΔP is the change of hydraulic head in the production well, p_e is the electricity price. Note that pumping water to the surface requires additional electricity due to reservoir pressure drawdown in the production well (Eq. (1c)), and this cost is calculated based on the local electricity price (Kong et al. 2017). The profit and cost are calculated in US dollars ($) according to the local price. In this study, p_h = 0.09 \$/(kW·h) and p_e = 13.68 \$/GJ are considered.

2.2 Surrogate model based on machine learning techniques

Constructing a surrogate model mainly includes the following two steps: (1) data generation based on the integrated geothermal reservoir model, (2) surrogate model training and testing based on machine learning techniques. Coupled fluid flow and heat transfer modeling on a doublet system in the heterogeneous geothermal reservoir has been well developed in Liu et al. (2019), and a simplified geothermal well model was proposed to improve the computational efficiency without losing accuracy (Wang et al. 2019). The integrated geothermal reservoir model with the simplified geothermal well model was first built and calibrated, and then used to generate the training and test data sets (Wang et al. 2022). In this study, it is assumed that the position of the production well is fixed, while other operation parameters including the position of the injection well, the injection and production rates and the injection temperature are varied. For each scenario, the value of the objection function is calculated using the integrated geothermal reservoir model, and the training and test data sets are formulated by repeating the calculation process. In addition, the maximum and minimum permeabilities in the three representative areas are also calculated as the predictor variables (Figure 1).

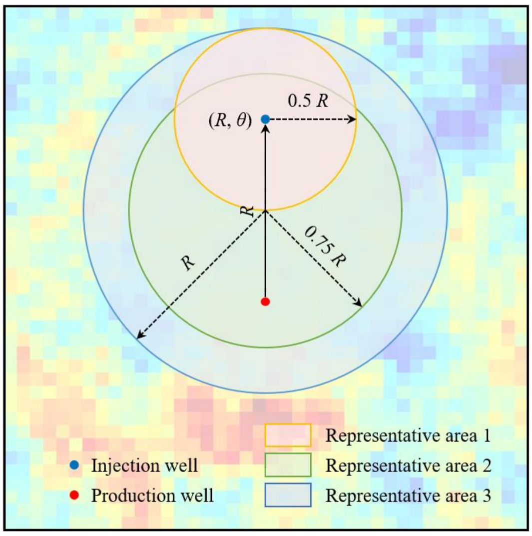

Figure 1: Schematic diagram of the predictor variables in the three representative areas (Wang et al. 2022).

Four tree-based and network-based ML algorithms, namely decision regression tree (DRT), random forest (RF), light gradient boosting machine (LightGBM), and multilayer perceptron (MLP) are used to construct the surrogate model between the operation parameters and the value of the objective function, respectively. The main ideas of the four ML algorithms are presented below.

(1) DRT

DRT is the regression equivalence of decision tree. Internal nodes only decide true or false by indicators, and transfer input from last nodes into next nodes until reaching leaf nodes corresponding to specific regression values. Strategies preventing overfit include pre-pruning and post-pruning.

(2) RF

RF adopts bagging strategy based on DRT, via bootstrap sampling and random feature selection in order to enlarge the diversity of base learners and enhance learning effect.

(3) LightGBM

LightGBM is the derivative method of gradient boosting decision tree, which searches the optimal node splitting position by histogram algorithm, to reduce memory consumption and training time.

(4) MLP

MLP belongs to artificial neural network, which consists of input layer, hidden layer and output layer. The activation function can improve the nonlinearities of the network. The network uses back propagation algorithm to update the weights and biases of each layer based on automatic differentiation and gradient descent. The whole process is repeated until the network converges.

Four primary measures including *R*-Square (R^2), root mean squared error (*RMSE*), mean absolute error (*MAE*), and mean absolute percentage error (*MAPE*) are used to evaluate the performance of the surrogate model (Eqs. (2a-d)).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{2a}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2b}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{2c}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i' - \hat{y}_i'}{y_i'}\right| \tag{2d}$$

where i is dataset sample number, n is the number of samples, y' and y are true value before and after normalization respectively, $\hat{y}'$ and $\hat{y}$ are predicted value before and after normalization respectively.

2.3 Optimization approach for operational parameters

The validated surrogate models are coupled with genetic algorithm (GA) to efficiently search for the maximum objective function, which avoids running a numerical simulation for each individual response. In GA, a population of candidate solutions to an optimization problem is evolved towards better solutions. The evolution starts from a population with a number of randomly generated individuals, and is an iterative process with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated. The more fit individuals are stochastically selected from the current population, and each individual's genome is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

3. BENCH MARK EXAMPLE

3.1 Geothermal reservoir model

An area of 16 km^2 was selected as the numerical model area, with the depth of 1130~1430 m. The 3D heterogeneous reservoir (Figure 2) was generated by the turning bands method, which was divided into cubic domains with the side length of 50 m and the correlation length of 200 m. The average reservoir permeability was $4.83\times10^{-11.8}$ m^2 and the variance was 0.5. The length of the open hole section of geothermal wells was assumed as 100 m. The values of doublet system parameters are shown in Table 1.

The initial hydraulic pressure field in the reservoir was generated under gravity, and thus the water pressure linearly related to the depth was applied on the lateral boundaries. The initial temperature field was assumed as a constant of $60 + 0.0025x$ ℃, where x was the x-coordinate value in the model. For the fluid flow process, a constant value of pressure was applied on the lateral boundaries, and impermeable boundary conditions were applied on the top and bottom boundaries. For the heat transfer process, open thermal boundary conditions were applied on the lateral boundaries, and adiabatic boundary conditions were applied on the top and bottom boundaries.

The reservoir was simulated as 3D fully saturated domains following the thermo-poroelasticity theory, which was discretized into finite tetrahedron elements. Geothermal wells were simulated as 1D line elements, which considered the heat convection and conduction along well axis and heat and mass exchange between geothermal fluid and reservoir rocks in radial direction. Both computation accuracy and efficiency were taken into account. The total numbers of elements were of about 48,000 tetrahedron elements for the reservoir and 400 1D elements for geothermal wells.

In this study, we considered a duration of continuous injection/production of 50 years, during which the evolutions of temperature and water table were continuously monitored in order to calculate the objective function P. The time step was set as one month.

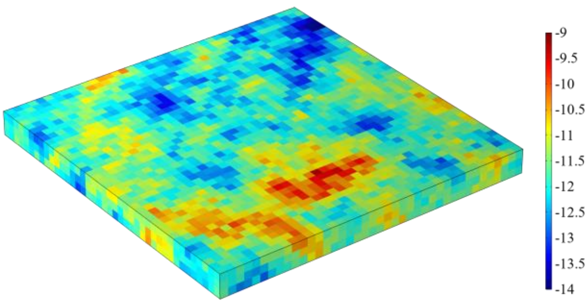

Figure 2: Normal distribution of permeability for the bench mark example (unit: log10(κ) m^2).

Table 1: Parameters of the bench mark example

	Parameters	Value	Unit
Reservoir rocks	Density (ρ_s)	2600	kg/m^3
	Permeability (log10(κ))*	-11.8	m^2
	Specific heat capacity ($C_{p,s}$)	880	J/(kg·K)
	Thermal conductivity ($k_{p,s}$)	1.68	W/(m·K)
	Porosity (φ)	0.26	
Water	Fluid density (ρ_f)	1000	kg/m^3
	Specific heat capacity of fluid ($C_{p,f}$)	4200	J/(kg·K)
	Thermal conductivity of fluid ($k_{p,f}$)	0.58	W/(m·K)

3.2 Surrogate model

Data generation was the first step in constructing the surrogate model. The production well was in the center of the squared numerical reservoir model. The injection well was arranged within 200 m to 1000 m away from the production well, the injection and production rates were between 80 m^3/h and 120 m^3/h, and the injection temperature was between 25 ℃ and 35 ℃ (Figure 3). Uniform random numbers were set to ensure the randomness and make the operational parameters cover the entire optimization space as much as possible. A total of 600 observations were generated, including 85% of training data and 15% of test data. Input variables included four operational parameters, namely, the well spacing (R), the angle of the injection well with respect to the horizontal line (θ), the injection and production rates (Q), the injection temperature (T_{in}), and 6 extreme values of the permeabilities. Values of the objective function were calculated by reservoir responses, which formed the data set together with input variables.

Then, surrogate models were constructed based on different machine learning techniques, including decision regression tree, random forest, light gradient boosting machine, and multilayer perceptron. The results of surrogate models trained by different techniques are shown in Table 2. The four evaluation indicators R^2, *RMSE*, *MAE* and *MAPE* showed that the surrogate model based on multi-layer perceptron was optimal. When the training data size was 510, the value of R^2 was close to 1, which meant the surrogate model could predict the objective function with sufficient accuracy (Figure 4). Therefore, the surrogate model based on MLP was used to predict the optimal operational parameters through genetic algorithm.

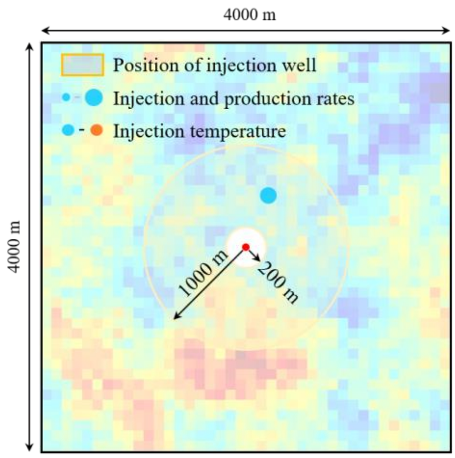

Figure 3: Schematic diagram of the operational parameters

Table 2: The evaluation indicators of surrogate models

ML algorithm	DRT	RF	LGBM	MLP
R^2	0.969	0.992	0.991	0.999
RMSE	0.034	0.017	0.018	0.007
MAE	0.027	0.012	0.014	0.004
MAPE	0.020	0.009	0.010	0.003

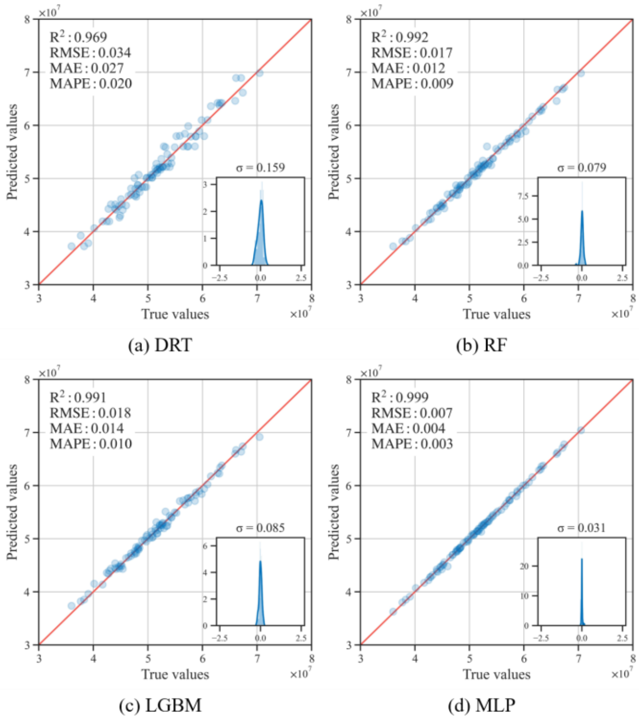

Figure 4: Scatter plot and histogram of the objective function calculated from the surrogate model.

3.3 Optimal operation parameters

The number of population and generation in genetic algorithm is determined according to the complexity of the doublet system. In this study, operation parameters including the position of the injection well, the injection and production rates and the injection temperature were optimized simultaneously. Therefore, the population size was 40, the maximum number of generations was 1000, crossover probability and mutation probability were set to 0.7 and 0.05, respectively.

The surrogate model based on multi-layer perceptron was combined with genetic algorithm to predict the optimal operational parameters. Each run of the surrogate model took about 60 s, and the total time was about 10 h (600 observations), including the optimization time of about 5 minutes. The results showed that the optimal placement of the injection well was x = 408.742 m and y = −214.040 m, the optimal rate of injection and production was Q = 120 m^3/h, and the injection temperature was T_{in} = 25 °C, with the objective function value of \$ 7.126×10^7. Using the numerical simulation model, we obtained the objective function value of \$ 7.260×10^7 for the same operational parameters, which was close to the predicted results. We also applied the simulation-based optimization method to this case with the same population size and the maximum number of generations. After running for about 67 h, the number of generations reached the maximum preset value. The final optimization results were x = 410.080 m , y = −217.005 m, Q = 120 m^3/h and T_{in} = 25 °C, with the economic profit value of \$ 7.251×10^7. Figure 5 shows the iterative results of two methods. The good agreement between the proposed optimization method and simulation-based optimization method demonstrated that the surrogate model-based method was reasonable and accurate with less computational cost. In addition, for the heterogeneous geothermal reservoir used in this bench mark, the rate and the injection temperature were the extrema within the specified optimization scope, which exhibited that although the larger rate would lead to pressure drawdown and increase of electricity cost, and lower injection temperature would induce thermal breakthrough and decrease of profit, the negative effects of them were far less than the raise of objective function, namely the economic profit.

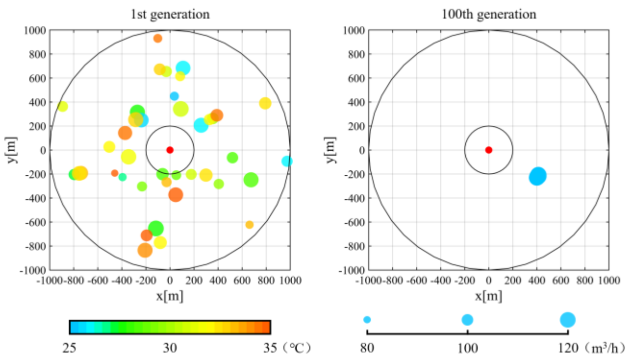

Figure 5: Iterative results of the operation parameters.

4. CONCLUSION

In this study, we developed a surrogate model-based optimization method to find the optimal operational parameters in geothermal reservoirs with less computational cost compared with the simulation-based optimization method. Based on machine learning techniques and predictor variables, the trained surrogate models could accurately predict the economical objective function in heterogeneous geothermal reservoirs. Combining the optimal surrogate model with genetic algorithm led to the surrogate model-based optimization method, the robustness and accuracy of which were demonstrated through the bench mark example of a doublet system. The results showed that the surrogate model based on multi-layer perceptron had the best prediction performance compared with other machine learning techniques. The surrogate model-based optimization method could find the optimal operation parameters including the position of the injection well, the injection and production rates and the injection temperature, corresponding to the maximum economic profit value, and the error of the optimization results was within an acceptable range.

ACKNOWLEDGEMENT

The study was jointly supported by the National Key R&D Program of China (No. 2021YFB1507302) and the National Natural Science Foundation of China (U2067203).

REFERENCES

Kong Y, Pang Z, Shao H, Kolditz O. Optimization of well-doublet placement in geothermal reservoirs using numerical simulation and economic analysis. *Environmental Earth Sciences*, **76**, (2017), 118.

Liang X, Xu T, Feng B, Jiang Z, Optimization of heat extraction strategies in fault-controlled hydro-geothermal reservoirs, *Energy*, **164**, (2018), 853-870.

Liu G, Pu H, Zhao Z, Liu Y. Coupled thermo-hydro-mechanical modeling on well pairs in heterogeneous porous geothermal reservoirs, *Energy*, **171**, (2019), 631-653.

Liu G, Zhao Z, Xu H, Zhang J, Kong X, Yuan L. A robust assessment method of recoverable geothermal energy considering optimal development parameters, *Renewable Energy*, **201**, (2022), 426-440.

Moraga J, Duzgun HS, Cavur M, Soydan H. The Geothermal Artificial Intelligence for geothermal exploration, *Renewable Energy*, **192**, (2022), 134-149.

Okoroafor ER, Smith CM, Ochie KI, Nwosu CJ, Gudmundsdottir H, Aljubran M. Machine learning in subsurface geothermal energy: Two decades in review, *Geothermics*, **102**, (2022), 102401.

Wang G, Liu G, Zhao Z, Liu Y, Pu H. A robust numerical method for modeling multiple wells in city-scale geothermal field based on simplified one-dimensional well model, *Renewable Energy*, **139**, (2019), 873-894.

Wang J, Zhao Z, Liu G, Xu H. A robust optimization approach of well placement for doublet in heterogeneous geothermal reservoirs using random forest technique and genetic algorithm, *Energy*, **254**, (2022), 124427.