Advances in Geo-Energy Research

Vol. 18, No. 1, p. 84-98, 2025

Original article

Efficient optimization of coupled geothermal reservoir modeling and power plant off-design based on deep learning

Ziyou Liu, Manojkumar Gudala, Bicheng Yan*

Physical Science and Engineering (PSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Keywords:
Geothermal
power plant
coupling
deep learning
optimization

Cited as:
Liu, Z., Gudala, M., Yan, B. Efficient optimization of coupled geothermal reservoir modeling and power plant off-design based on deep learning. Advances in Geo-Energy Research, 2025, 18(1): 84-98.
https://doi.org/10.46690/ager.2025.10.07

Abstract:
The accurate evaluation of the electricity output of geothermal power plants requires effective coupling between the geothermal reservoir and power plant. Existing coupling models integrate numerical simulation models of the reservoir and power plant; however, they are computationally expensive for electricity prediction (forward modeling) and integrated reservoir-power plant optimization. Therefore, this study aimed to enhance the efficiency of the coupled reservoir-power plant model for forward modeling and optimization by replacing simulation forward models with deep-learning-based surrogate models. Two independent surrogate models of the reservoir and power plant were trained and assembled into one coupled forward model. Moreover, a multiobjective optimizer was integrated with the coupled forward model to optimize reservoir operations and power plant designs to achieve the highest electricity output or the best economic outcome. Surrogate models for the reservoir and power plant accurately predicted the geothermal production temperature and electricity output while approximately achieving speedups of 1.23×10^5 and 1.77×10^5 times over those of the corresponding simulation models, respectively. Furthermore, optimization using our surrogate-based coupled model was 1.31×10^6 times faster than that using the simulation-based coupled models. Optimization results revealed that low injection temperature, large well distance, and stable reservoir injection and production rates contributed to better power plant performance. High design geothermal temperature, mass flow rate, and ambient temperature favored electricity generation, particularly in power plants located in hot regions. Our work remarkably accelerates the feasibility assessment and decision-making procedures for geothermal reservoirs and power plants.

1. Introduction

In 2017, the Kingdom of Saudi Arabia (KSA) launched a 500-billion-USD project to create an innovative economic zone called "NEOM", with the goal of achieving zero carbon emissions. Therefore, the development of renewable energy in NEOM is crucial, and geothermal energy represents a promising option. NEOM has abundant geothermal sources, where the geothermal gradient ranges from 35 to 45 °C/km (Aboud et al., 2023), which exceeds the typical geothermal gradient (25-30 °C/km). The primary geothermal reservoir in NEOM is a hydrothermal reservoir composed of permeable porous media saturated with hot brine; it is located 200-3,000 m beneath the Earth's surface.

High-fidelity reservoir simulations can effectively simulate the thermal front movement and predict the production temperature. However, geothermal reservoir simulation involves high computational costs because of reservoir heterogeneity, multiphysics coupling phenomena, and matrix-fracture interactions (Yan et al., 2024). The high computational costs in-

Yandy Scientific Press

*Corresponding author.
E-mail address: ziyou.liu@kaust.edu.sa (Z. Liu); manojkumar.gudala@kaust.edu.sa (M. Gudala); bicheng.yan@kaust.edu.sa (B. Yan).

Received August 15, 2025; revised September 8, 2025; accepted September 24, 2025; available online September 27, 2025.

hibit conventional simulation-based optimizations. In contrast, a deep learning (DL)-based surrogate model can simulate underlying nonlinear relationships between input and output parameters and directly predict outputs without solving partial differential equations, which is a more efficient alternative to simulation models in both forward and inverse modeling (Chen et al., 2025).

During geothermal production, the production temperature decreases with time after the cold waterfront reaches the production well (cold-water breakthrough). DL architectures excelling at time-series forecasts, such as recurrent neural networks (RNNs), gated recurrent unit (GRUs), and long short-term memory (LSTM) networks, are commonly used for predictions. Qin et al. (2023) predicted and optimized the net power generation of a geothermal field using an RNN-based surrogate model. Li et al. (2024) used a bidirectional GRU to predict the production temperature of a fractured geothermal reservoir. Wang et al. (2023) developed a convolution neural network (CNN)-LSTM hybrid surrogate model to predict the production rate and temperature. However, RNN-based models cannot fully exploit the parallel computing capability of modern graphics processing units (GPUs) and can be vulnerable to vanishing and exploding gradients. In comparison, the transformer architecture is built entirely based on the attention mechanism and processes all timesteps simultaneously, which mitigates vanishing and exploding gradients and allows the maximum utilization of the parallelism capability of GPUs. Xue and Chen (2023) compared the performances of an RNN, an LSTM network, and a transformer architecture in geothermal electricity output predictions and demonstrated that the transformer architecture had the best predictive accuracy. However, the transformer architecture used in their study was designed to solve autoregressive problems, in which the current timestep is predicted based on previous timesteps. The transformer's performance in solving nonautoregressive problems and its hybridization capability with other architectures remain to be studied. In this study, a hybrid CNN-nonautoregressive transformer (NAT) architecture was constructed to predict the geothermal production temperature. The hybrid architecture mitigated vanishing and exploding gradients while allowing multimodal input features, including three-dimensional (3D) heterogeneous geological fields (permeability and porosity), time-varying injection schedules, and other scaler parameters.

The geothermal fluid produced by a reservoir with a temperature > 80 °C can be utilized by a power plant to generate electricity (Hu et al., 2021). The reservoir productivity (production temperature and rate) has a significant impact on the output of a power plant. Therefore, the coupling between the subsurface and surface units must be studied because it plays a critical role in accurately evaluating the performance of a power plant. Garapati et al. (2017) integrated a multiphase reservoir simulation model with a power plant model to examine the electricity output potential of a CO_2-based geothermal system. Jiansheng et al. (2022) built a coupling model integrating an enhanced geothermal system (EGS) with a horizontal well and an organic Rankine cycle (ORC) power plant. Hsieh et al. (2024) coupled an EGS model with a flash-binary hybrid power plant model. Meng et al. (2024) proposed a time-sequential coupling model of EGS and ORC power plants. However, existing coupled models face two major challenges. First, these coupled models integrate numerical simulation models of the reservoir and power plant, which results in high computational costs. Reservoir models are commonly developed in simulators such as COMSOL Multiphysics or TOUGH2, whereas power plant models are simulated in ASPEN PLUS or in-house simulators. However, the coupling of such computationally intensive simulators significantly increases the computation time of the models for electricity output predictions and optimization. This may be the reason why the aforementioned studies did not apply their models in optimization contexts. Second, current studies assume constant injection and production rates. However, realistic injection and production rates of a geothermal reservoir are rarely constant throughout the lifetime. The injectivity can decrease due to clogging triggered by particle migration, mineral precipitation, and microbial activities (Markó et al., 2021; Luo et al., 2023). The injection scheme must be changed under such circumstances, resulting in time-varying production rates. Besides, the investors may deliberately change the production rate based on the real-time energy price and market demand (Haklıdır, 2020). Therefore, dynamic well control in geothermal applications must be investigated for flexible reservoir management and productivity adjustments. Besides, the effects of dynamic injection and production rates on power plant performance cannot be examined using existing models.

This study primarily aimed to accelerate the forward modeling and optimization procedures of a coupled reservoir-power plant model by integrating the coupled model with surrogate models instead of simulation models. In reservoir modeling, a CNN-NAT hybrid surrogate model was constructed for the 3D heterogeneous geothermal reservoir in NEOM to efficiently predict the wellhead production temperature. Dynamic well controls were implemented in the reservoir model to better align with realistic production schedules. In power plant study, a DL-based power plant surrogate model was built to predict the maximum electricity output under actual heat-source and heat-sink conditions. More importantly, the reservoir and power plant surrogate models were assembled into one coupled forward model through heat-source conditions, efficiently coupling the subsurface and surface units. Finally, the forward model was combined with a multiobjective optimizer to efficiently optimize the coupled system to achieve the highest electricity output or best economic outcome. The optimal decision parameters of the geothermal reservoir and the optimal power plant design conditions within the framework are discussed herein.

The novelty of our work lies in the following three aspects:

1) A CNN-NAT reservoir surrogate model is established. The NAT can predict production temperatures at all timesteps simultaneously without dependence on previous timesteps, while mitigating vanishing or exploding gradients observed in RNN-based architectures. Hybridization with a CNN allows the architecture to process complex 3D spatial data beside sequential data.

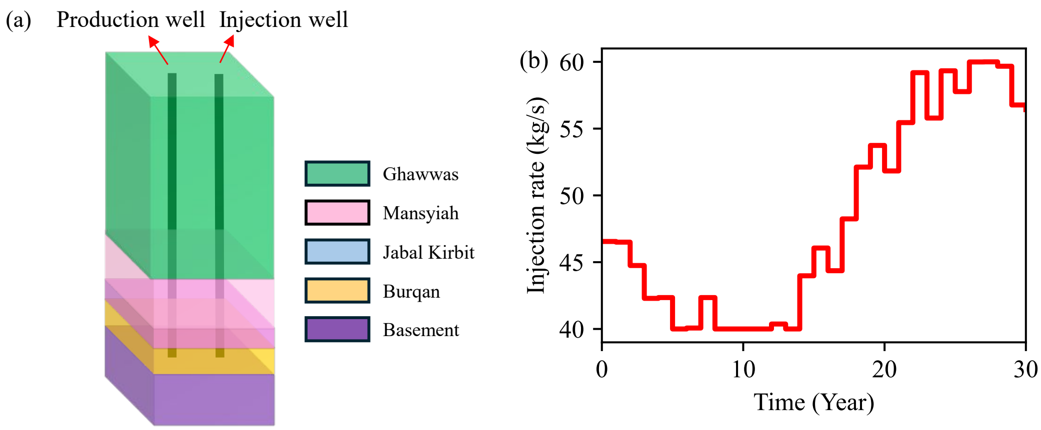

Fig. 1. Geothermal reservoir model: (a) Stratigraphy of the Ifal Basin and (b) dynamic injection rates.

2) A data-driven forward model coupling the reservoir and power plant is proposed, which is much more efficient than simulation-based coupled models reported in existing studies. The effects of dynamic reservoir injection and production rates on the performance of the power plant are examined.
3) The multiobjective optimizer combined with the forward model efficiently optimizes the coupled system considering reservoir uncertainty. The optimal decision parameters of the reservoir and power plant are determined based on reservoir productivity variations and power plant outputs throughout the lifespan.

Our work remarkably accelerates the feasibility assessment and decision-making procedures of geothermal reservoirs and power plants. The remainder of the manuscript is organized as follows: Section 2 discusses the methodology of the surrogate models and optimization framework. Section 3 presents the prediction performance of the surrogate models. Section 4 discusses the optimization results. Section 5 presents the conclusions and highlights of this work.

2. Methodology

2.1 Geothermal reservoir simulation model

The Ifal Basin is one of the well-defined basins along the Red Sea coast in the NEOM region, northwestern Saudi Arabia. Because the geology of the Ifal Basin has been well studied, a hydrothermal reservoir model based on the geological data of the basin was developed to study the geothermal energy potential of this region.

The stratigraphy of the Ifal Basin is shown in Fig. 1(a). Over the granitic basement, there are four sedimentary formations characterized by distinct lithology and thickness. From top to bottom, they are the Ghawwas, Mansyiah, Jabal Kirbit, and Burqan formations (Hughes and Johnson, 2005). Among them, the Burqan formation is the deepest sandstone formation with the highest initial reservoir temperature and permeability. Therefore, it was selected as the target reservoir for geothermal production. The other three formations serve as the overburden in the model. The lithology, thickness, and other properties of the four formations are summarized in Table S5 (Supplementary file).

The target reservoir has a volume of $1,000 \times 1,000 \times 200$ m^3 in the x, y, and z directions, respectively. It is located at a depth of 2,286 m beneath the surface. A doublet system – a pair of injection and production wells – was employed for reservoir development. The production rate is set equal to the injection rate, which is a common setup in a doublet system (Ke et al., 2021), assuming high permeability between the production and injection wells and negligible fluid losses to the formation. Dynamic well controls were implemented on the doublet system. The injection rate changes dynamically each year, as shown in Fig. 1(b). In addition, the surface temperature is 30 °C, and constant pressure and temperature conditions are applied to the external boundaries.

Considering reservoir heterogeneity in terms of permeability and porosity, 3D heterogeneous permeability and porosity fields were generated using the open-source Python library GSTools (Müller et al., 2022). The permeability fields were first generated, which followed the Gaussian distribution. Further, the porosity-permeability correlation for sandstone regressed by Gudala and Govindarajan (2020) was used to generate the porosity fields. The correlation is expressed as follows:

$$\phi = \frac{\log_{10} K + 16.97}{19.12} \tag{1}$$

where ϕ is the porosity and K is the permeability.

Examples of 3D heterogeneous permeability and porosity fields are shown in Fig. 2. The porosity field follows the same spatial distribution pattern as the permeability field owing to the correlation.

The reservoir model was built using the simulation software COMSOL Multiphysics (Multiphysics, 1998). The COMSOL modules used in reservoir simulation included the “heat transfer”, “flow in porous media”, “rock mechanics”, and “poroelasticity” modules, which simulated the thermo-hydro-mechanical (THM) coupling mechanism during geothermal recovery (Wan et al., 2023). Moreover, a nonisothermal wellbore flow model was developed using the nonisothermal pipe flow module and coupled with the reservoir model to calculate heat

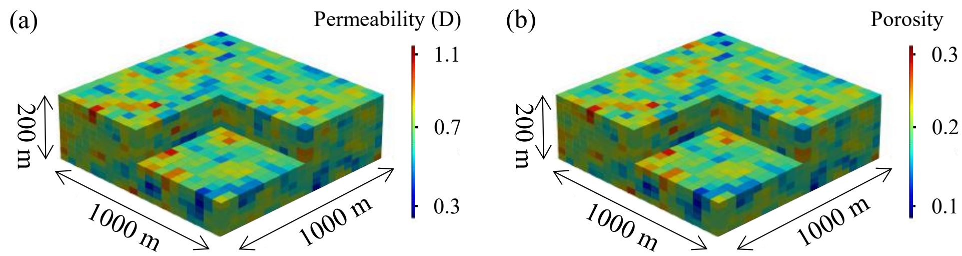

Fig. 2. Heterogeneous geological fields: (a) permeability and (b) porosity fields.

Table 1. Ranges of scaler reservoir parameters (Al-Laboun et al., 2014; Hoteit et al., 2023).

Parameter	Unit	Lower bound	Upper bound
$K_{\min}$	mD	20	80
$K_{\max}$	mD	4,000	11,000
I	-	0.1	0.5
c_x	m	50	200
c_y	m	50	200
c_z	m	25	50
T_{inj}	°C	20	100
L_w	m	100	600
G	°C/km	35	50
λ	$\mathrm{W/(m \cdot K)}$	1	5
C_p	$\mathrm{J/(kg \cdot K)}$	700	1,200
α	-	0.2	1
E	GPa	10	40
ν	-	0.2	0.4

loss through the production wellbore. The governing equations of the THM reservoir model and the nonisothermal wellbore model are presented in Appendix A in Supplementary file. The physical model of the production wellbore has been discussed by Shi et al. (2019), as shown in Fig. 3. The properties of each component in the wellbore model are listed in Table S6.

2.2 Geothermal reservoir surrogate model

2.2.1 Data preparation

To build a database for reservoir surrogate model training, 2,000 cases based on the numerical model discussed in Section 2.1 were generated and simulated. Eleven parameters, including three operational parameters and eight reservoir property parameters, were selected as the input variables of the surrogate model: Injection temperature T_{inj}, injection rate q_{inj}, well distance L_w, geothermal gradient G, thermal conductivity λ, specific heat capacity C_p, Biot coefficient α, Young's modulus E, Poisson ratio ν, porosity ϕ, and permeability K. The injection rates of all cases were sampled by following a uniform distribution between 40 and 60 kg/s. The rate changes between two neighboring timesteps were limited to < 5 kg/s

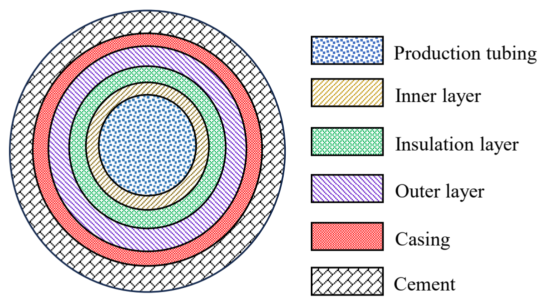

Fig. 3. Structure of a geothermal production wellbore.

to avoid drastic injectivity or productivity variations and allow the reservoir to fully respond to productivity changes. Moreover, 2,000 permeability fields were generated with GSTools. Six parameters determined the permeability field (minimum permeability $K_{\min}$; maximum permeability $K_{\max}$; anisotropic ratio I; and correlation lengths in the x, y, and z directions, denoted as c_x, c_y and c_z, respectively). A smaller correlation length indicates greater spatial heterogeneity and vice versa. These parameters were sampled using the Latin Hypercube sampling method, with ranges listed in Table 1. After the permeability fields were generated, the same number of porosity fields were generated based on the porosity-permeability correlation expressed in Eq. (1).

After running all the simulations in COMSOL, the wellhead production temperature data of 30-year duration were collected. To better capture the cold-water breakthrough time, the temperature data were collected using a small timestep (0.5 year). Overall, a database containing input and output data of 2,000 cases was developed.

2.2.2 CNN-NAT

A hybrid CNN-NAT reservoir surrogate model (denoted as $\mathscr{S}_r$) was constructed to predict the wellhead production temperature $T_p(t)$ from multimodal input features, including heterogeneous geological fields $\boldsymbol{m}_g$, time-sequential well control $\boldsymbol{m}_w$, and scaler parameters $\boldsymbol{m}_s$. A CNN encoder ($\mathscr{C}$) was used to capture the spatial information, and NAT was employed to predict the production temperature. The prediction procedure was as follows:

$$T_p(t) = \mathscr{S}_r(\boldsymbol{m}_g, \boldsymbol{m}_w, \boldsymbol{m}_s) \tag{2}$$

Fig. 4 shows the architecture of $\mathscr{S}_r$. The input of the CNN encoder was 3D heterogeneous geological data. In this study, because permeability and porosity were correlated, either of them could be input into the model (the permeability

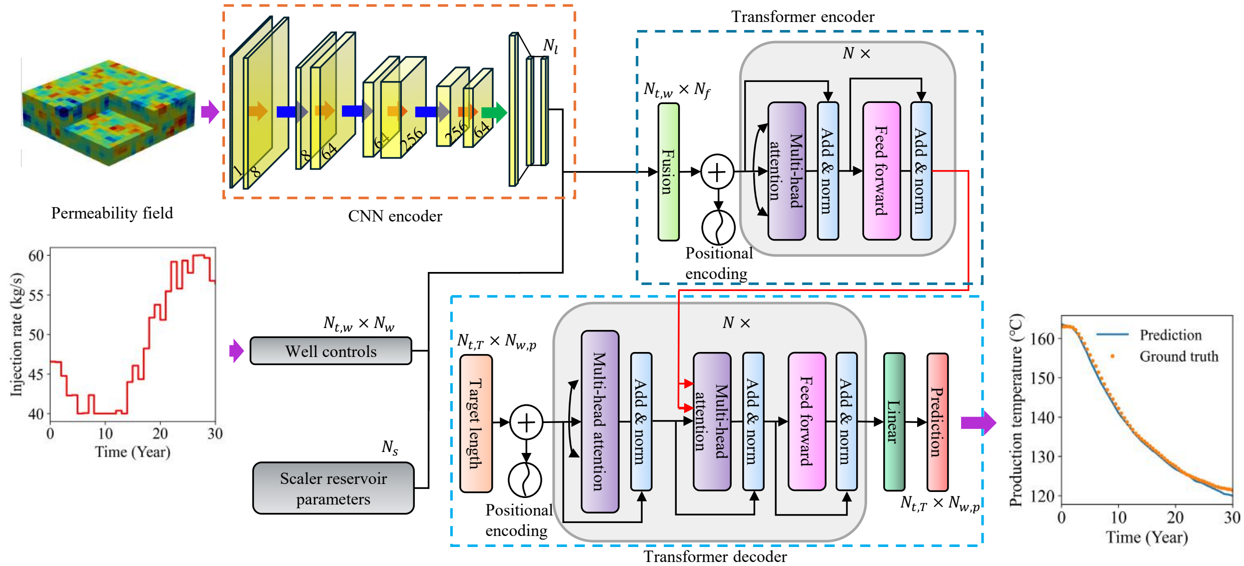

Fig. 4. CNN-transformer architecture.

field was input in this study). After processing through the convolutional blocks (red arrows) and pooling layers (blue arrows) of the CNN encoder, the high-dimensional geological data were flattened into a 1D latent vector, which is a dense representation of the original permeability field with a reduced number of parameters (Feng et al., 2025). The aforementioned procedure is formulated as:

$$\tilde{\boldsymbol{m}}_g = \mathscr{C}(m_g) \tag{3}$$

where $m_g \in \mathbb{R}^{N_z \times N_x \times N_y}$ is the 3D permeability field; N_z, N_x, and N_y are dimensions in the z, x, and y directions, respectively, which equal the number of grids in each direction in reservoir discretization; and $\tilde{\boldsymbol{m}}_g \in \mathbb{R}^{N_l}$ is the latent vector with dimension N_l.

The latent vector obtained from the CNN encoder is one of the input features of the transformer architecture, together with well-control data ($\boldsymbol{m}_w \in \mathbb{R}^{N_{t,w} \times N_w}$) and other scaler input variables ($\boldsymbol{m}_s \in \mathbb{R}^{N_s}$). $N_{t,w}$ is the number of timesteps of the well-control data, and N_w is the number of wells. In this study, $N_w = 1$ because the injection and production wells shared the same schedule. N_s is the number of scaler input variables. The input vectors are broadcasted and projected into a shared embedding space of dimension N_f and ultimately fused into one latent representation $\tilde{\boldsymbol{m}} \in \mathbb{R}^{N_{t,w} \times N_f}$ by addition.

Because the attention-based transformer architecture does not have an inherent sense of sequence, positional encoding (PE) is added as the temporal information of data points. Sinusoidal positional encodings are used, and they are expressed as follows (Vaswani et al., 2017):

$$PE_{(\mathscr{P},2i)} = \sin \frac{\mathscr{P}}{10000^{2i/N_f}} \tag{4}$$

$$PE_{(\mathscr{P},2i+1)} = \cos \frac{\mathscr{P}}{10000^{2i/N_f}} \tag{5}$$

where $\mathscr{P}$ is the temporal position of a data point in a time sequence and i is the index of a feature in the projected space.

After augmentation with PE, the resulting sequence $\tilde{\boldsymbol{m}} + PE \in \mathbb{R}^{N_{t,w} \times N_f}$ is input to the transformer encoder. The encoder generates contextual representations that capture long-range dependencies within the input sequence through the multihead self-attention mechanism. The contextual representations are passed to the decoder through the cross-attention mechanism (red elbow arrow from the transformer encoder to the transformer decoder in Fig. 4). Subsequently, they are further transformed by the feed-forward network and the linear projection layer to generate the final geothermal production temperature predictions with the targeted length.

Compared with recurrent architectures such as LSTM networks and RNNs, the transformer architecture based on the multihead attention mechanism processes all timesteps in parallel instead of processing them sequentially, which better utilizes the GPU parallelism ability and effectively mitigates vanishing and exploding gradients. Besides, hybridization with a CNN allows the transformer architecture to handle complicated 3D spatial information instead of handling only sequential data.

The mean square error (MSE) was used as the loss function L_r during training, and the nondimensional coefficient of determination (denoted as R^2) was employed as a performance evaluation metric of the surrogate model, which are expressed by Eqs. (6) and (7), respectively:

$$\mathscr{L}_r = \frac{1}{N_b N_{t,T}} \sum_{j=1}^{N_b} \sum_{k=1}^{N_{t,T}} \left(\hat{T}_{j,k} - T_{j,k}\right)^2 \tag{6}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{j=1}^{N_b} \sum_{k=1}^{N_{t,T}} \left(\hat{T}_{j,k} - T_{j,k}\right)^2}{\sum_{j=1}^{N_b} \sum_{k=1}^{N_{t,T}} \left(T_{j,k} - \bar{T}\right)^2} \tag{7}$$

where $\hat{T}_{j,k}$ and $\hat{T}_{j,k}$ are the predicted temperature and ground truth of sample j at timestep k, respectively; N_b is the total number of samples (i.e., batch size); and $\bar{T}$ is the mean ground truth temperature over all samples and timesteps.

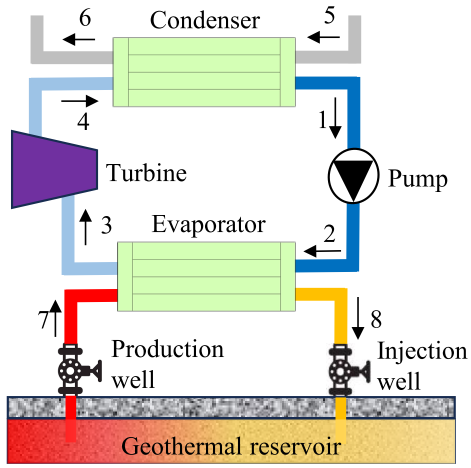

Fig. 5. ORC layout.

2.3 Power plant simulation model

The subcritical binary ORC power plant was studied because it is currently the most typical and widely used configuration, accounting for 42.8% of the 673 power units worldwide (Gutiérrez-Negrín, 2024). The layout of the ORC, comprising an evaporator, a turbine, a condenser, and a pump, is illustrated in Fig. 5.

The net electricity output of the ORC P_{net} was calculated as follows:

$$P_{\text{net}} = P_t - P_p - bm_c \tag{8}$$

where P_t is the raw electricity output of the turbine, P_p is the electricity consumption of the pump, and m_c is the mass flow rate of the cooling fluid. Air was used as the cooling fluid because of water scarcity in Saudi Arabia. Finally, b is the electricity consumption of the condenser per mass flow rate of air.

The levelized cost of electricity (LCOE) and net present value (NPV) were employed as key metrics to evaluate the cost-effectiveness of the power plant. The LCOE is defined as the ratio of the discounted costs over the lifetime of a power plant to the discounted actual cumulative electricity generated (Lai and McCulloch, 2017), which is expressed by:

$$\text{LCOE} = \frac{C_{\text{in}} + C_d + \int_0^{t_{ab}} C_{om,a} e^{-i_r t} \mathrm{d}t}{\int_0^{t_{ab}} P_{cum,a} e^{-i_r t} \mathrm{d}t} \tag{9}$$

where C_{in} is the installation cost of the power plant, C_d is the drilling cost of geothermal wells, t_{ab} is the abandonment time of the power plant, $C_{om,a}$ is the annual operation and maintenance (O&M) costs of the power plant, i_r is the interest rate (currently 5% in KSA), and $P_{cum,a}$ is the annual cumulative electricity output.

The NPV is expressed as:

$$NPV = -C_{\text{in}} - C_d + \int_0^{t_{ab}} (BP_{cum,a} - C_{om,a}) e^{-i_r t} \mathrm{d}t \tag{10}$$

where B is the electricity price.

The price of the electricity generated from geothermal energy comprises the market price and feed-in tariff (FIT). The FIT is an incentive policy that guarantees a specific price (normally higher than the market price) for electricity generated from renewable sources for a long period of time (normally 15-20 years). This policy reduces risks associated with renewable energy installation and production, which encourages investments in renewable energy. In 1990, Germany was the first country to apply the FIT to renewable energy. Currently, 93 countries have applied incentives for renewable energy, and 44 countries have FITs, including China, the UK, and France (Murdock et al., 2020). However, the KSA has not yet implemented a FIT or any other incentive policy for renewable energy. Detailed expressions used to calculate the parameters in P_{net}, the LCOE, and NPV are presented in Appendix A (Supplementary file).

In the design stage of the power plant, the design values of the heat-source (geothermal) temperature, heat-source mass flow rate, and heat-sink (ambient environment) temperature were specified. Based on the design heat-source and heat-sink conditions, the operational parameters of the power plant, such as the condensing temperature, working fluid temperature at the turbine inlet, cooling fluid temperature at the condenser outlet, and pump pressure, were optimized. The optimization goal was to achieve the maximum net electricity output.

The power plant is expected to operate as per the optimal operational parameters under the design conditions. However, when the heat-source or heat-sink conditions deviate from their design values, such as due to production temperature decline, production rate variations, or ambient temperature fluctuations, the power plant will operate under off-design conditions. Hence, the operational parameters should be optimized to achieve the highest net electricity output under the new heat-source and heat-sink conditions. Further details on the design and off-design procedures are available in our previous work (Liu et al., 2025). Both the design and off-design models were validated by comparing the results of this study with those reported in literature, demonstrating a strong correlation. The validation results are presented in Appendix B (Supplementary file).

2.4 Power plant surrogate model

2.4.1 Data preparation

The in-house power plant simulator took approximately 3-10 min to run one design and off-design simulation, which was computationally very expensive to be employed in optimization. Therefore, an efficient power plant surrogate model was developed.

The input variables of the surrogate model (denoted as $\boldsymbol{m}_p$) included:

1) Design values of the geothermal production temperature $T_{g,d}$, geothermal mass flow rate $q_{g,d}$, and ambient temperature $T_{am,d}$, which were used to design the power plant.
2) Off-design values of the geothermal production temperature $T_{g,o}$, geothermal mass flow rate $q_{g,o}$, and ambient temperature $T_{am,o}$, which determined the off-design conditions of the power plant.
3) The geothermal fluid re-injection temperature T_{re}, which

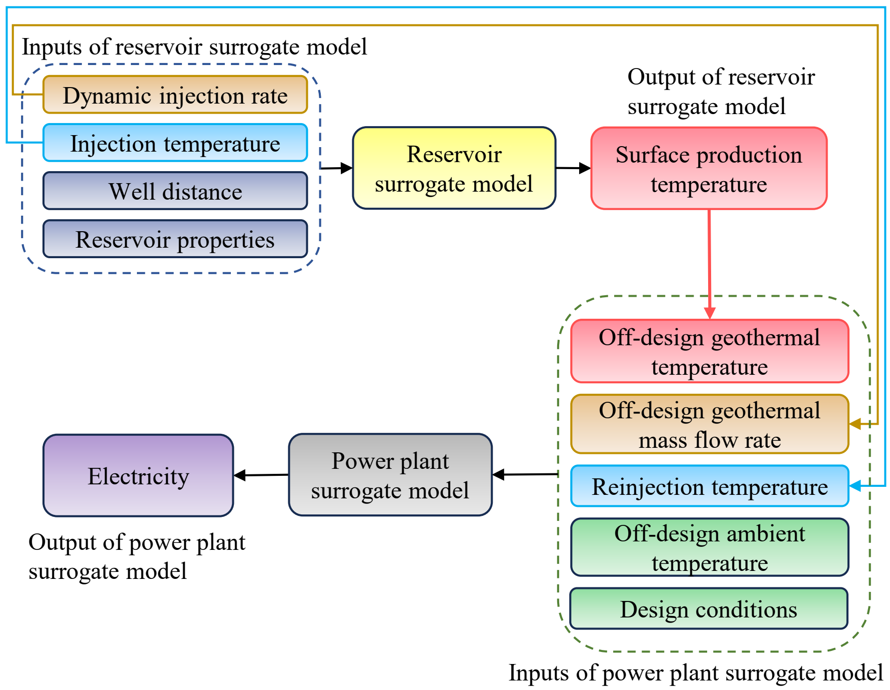

Fig. 6. Coupling relationships of the reservoir and power plant.

constrained the geothermal fluid temperature at the evaporator outlet.

The input parameters were sampled using the Latin Hypercube sampling method, with parameter ranges listed in Table 2. Notably, 2,000 cases were generated and simulated in the in-house power plant simulator. The resulting data on off-design net electricity output were collected as the ground truth to train the surrogate model. Consequently, the power plant surrogate model could flexibly predict the electricity output under any off-design condition for any power plant design.

Table 2. Ranges of power plant input parameters.

Parameter	Unit	Lower bound	Upper bound
$T_{g,d}$	°C	80	160
$q_{g,d}$	kg/s	20	70
$T_{am,d}$	°C	0	40
$T_{g,o}$	°C	80	160
$q_{g,o}$	kg/s	20	70
$T_{am,o}$	°C	0	40
T_{re}	°C	30	95

2.4.2 Surrogate model architecture

The power plant surrogate model (denoted as $\mathscr{S}_p$) was constructed based on a fully connected neural network (FCNN). The FCNN comprises a sequence of stacked layers, including input, hidden, and output layers, where each layer (denoted as $l_i, i = 1, 2, 3, \cdots, n$) contains numerous neurons. Every neuron in a layer receives and processes inputs from all neurons in the previous layer, making them "fully connected". The outputs of each layer are passed through a nonlinear activation function σ, which introduces nonlinearity and allows the network to learn complex patterns. The procedure can be formulated as follows:

$$\hat{p}_{\text{net}} = \mathscr{S}_p(\boldsymbol{m}_p; \theta) = \sigma_{\tanh}\left(l_n \sigma_{\tanh}\left(\ldots \sigma_{\tanh}\left(l_2 \sigma_{\tanh}\left(l_1(\boldsymbol{m}_p)\right)\right)\ldots\right)\right) \tag{11}$$

where θ denotes learnable parameters, including the bias and weight, and $\sigma_{\tanh}$ is the adaptive hyperbolic tangential activation function, which can effectively mitigate exploding or vanishing gradients (Yan et al., 2023).

The MSE is used as the loss function $\mathscr{L}_p$ in the FCNN:

$$\mathscr{L}_p = \frac{1}{n_b} \sum_{j=1}^{n_b} \left(\hat{p}_{\text{net},j} - p_{\text{net},j}\right)^2 \tag{12}$$

where $\hat{P}_{\text{net},j}$ and $P_{\text{net},j}$ are the prediction and ground truth of the net electricity output of sample j in a batch, respectively, and n_b is the batch size.

The training of $\mathscr{S}_p$ is based on the Adam optimizer (Paszke et al., 2019). The goal of the training is to determine learnable parameters by minimizing the loss function using:

$$\theta^* = \arg\min_{\theta} \mathscr{L}_p(\theta) \tag{13}$$

2.4.3 Coupled reservoir-power plant forward model

The coupling relationships between the reservoir and power plant surrogate models are illustrated in Fig. 6. Some input variables of the power plant surrogate model $\mathscr{S}_p$ can be obtained from the input or output variables of the reservoir surrogate model $\mathscr{S}_r$. For example, the off-design geothermal mass flow rate $q_{g,o}$ equals the reservoir injection rate q_{inj} and the re-injection temperature T_{re} equals the reservoir injection temperature T_{inj}, which can be obtained from the inputs of the reservoir surrogate model. Further, the off-design geothermal production temperature $T_{g,o}$ equal the reservoir production temperature T_p, which is the output variable of the reservoir surrogate model. Therefore, a data-driven forward model (denoted as $\mathscr{S}$), which couples the surrogate models of the reservoir ($\mathscr{S}_r$) and the power plant ($\mathscr{S}_p$) considering heat-source conditions, was developed. Any reservoir productivity change, such as production temperature decline or injection rate variation, can impact the electricity output or economic outcome of the power plant.

2.5 Multiobjective optimization

A multiobjective optimization (MOO) framework was developed to optimize the performance of the power plant while accounting for reservoir uncertainty. The general formulation of the MOO framework is expressed as follows:

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathscr{X}} F_i(\boldsymbol{x}), \quad i=1,2,\ldots,f \\ &\text{s. t. } L_j(\boldsymbol{x})\le 0, \quad j=1,2,\ldots,l \\ &H_k(\boldsymbol{x})=0, \quad k=1,2,\ldots,h \end{aligned} \tag{14}$$

where $\boldsymbol{x}$ is the vector of decision parameters optimized within the space $\mathscr{X}$; F_i is the i-th objective function; L_j is the j-th inequality constraint; H_k is the k-th equality constraint; and f, l, and h are the number of objective functions, inequality constraints, and equality constraints, respectively.

The MOO framework based on nondominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002) was adopted. This algorithm mainly operates as follows:

1) Initialization: A random population M_0 is generated within the searching space.
2) Offspring creation: Parents are selected from M_0 using binary tournament selection and offspring O_0 are generated through crossover and mutation operations.
3) Pareto front selection: The parent and offspring populations are merged (D_0). The solutions of all the populations are ranked through nondominated sorting, where one solution is considered dominant over another if it performs better in at least one objective and not worse in all others. Solutions that are not dominated by any other solutions are selected as nondominated solutions, which are also the optimal solutions, forming the Pareto front.
4) New population creation: Best solutions from D_0 are selected as the new parent population M_1. Steps 2-4 are then repeated with the new population until the maximum number of iterations or other convergence criteria are satisfied.

The ultimate output of the MOO framework is the Pareto front, which is a set of optimal solutions that offer tradeoffs between objective functions.

The parameters involved in the MOO are divided into three categories:

1) Uncertain parameter vector $\boldsymbol{m}_u$: This is the set of parameters with high uncertainty. Compared with the power plant, the reservoir has significantly higher uncertainty. Therefore, uncertain parameters in this study mainly refer to reservoir parameters, which are expressed as:

$$\boldsymbol{m}_u=[\lambda,C_p,\alpha,E,v,\phi(x,y,z),K(x,y,z)]^T \tag{15}$$

2) Decision parameter vector $\boldsymbol{m}_d$. This is the set of parameters to be optimized in the MOO framework, which is usually related to the operations. The decision parameters of the reservoir $\boldsymbol{m}_{d,r}$ include the injection temperature T_{inj}, injection rate q_{inj}, and well distance L_w. These operational parameters determine the decline behavior of the production temperature. The decision parameters of the power plant $\boldsymbol{m}_{d,p}$ comprise the geothermal temperature $T_{g,d}$, geothermal mass flow rate $q_{g,d}$, and ambient temperature $T_{am,d}$. These three design parameters determine the installed capacity of the power plant and sizes of the heat exchangers, which further affect the off-design performance of the power plant throughout the lifespan. Besides, the operational parameters of the power plant (e.g., condensing temperature and pump pressure) do not have to be explicitly optimized by the MOO framework because the simulator of the power plant and the surrogate model based on the simulator have inherently optimized them. The decision parameters in the MOO can be expressed as:

$$\boldsymbol{m}_{d,r}=[T_{inj},q_{inj},L_w]^T \tag{16}$$

$$\boldsymbol{m}_{d,p}=\left[T_{g,d},q_{g,d},T_{am,d}\right]^T \tag{17}$$

$$\boldsymbol{m}_d=\begin{bmatrix}\boldsymbol{m}_{d,r}\\ \boldsymbol{m}_{d,p}\end{bmatrix} \tag{18}$$

3) Given parameter vector $\boldsymbol{m}_g$: This is the set of parameters that are given or can be directly measured. The given parameters of the reservoir $\boldsymbol{m}_{g,r}$ only include the geothermal gradient G, which can be measured at the geothermal well site. The given parameters of the power plant $\boldsymbol{m}_{g,p}$ contain the off-design ambient temperature $T_{am,o}$, which can be measured in the ambient environment. Besides, owing to the coupling between the power plant and reservoir, the decision parameters and output of the reservoir, including T_{inj} (equivalent to T_{re}), q_{inj} (equivalent to $q_{g,o}$), and geothermal production temperature T_p (equivalent to $T_{g,o}$), can serve as given parameters of the power plant. The given parameters can be expressed as:

$$\boldsymbol{m}_{g,r}=[G]^T \tag{19}$$

$$\boldsymbol{m}_{g,p}=[T_{re},q_{g,o},T_{am,o},T_{g,o}]^T \tag{20}$$

$$\boldsymbol{m}_g=\begin{bmatrix}\boldsymbol{m}_{g,r}\\ \boldsymbol{m}_{g,p}\end{bmatrix} \tag{21}$$

Further, the input variables of the surrogate models of the reservoir $\boldsymbol{m}_r$ and power plant $\boldsymbol{m}_p$ can be re-assembled as:

$$\boldsymbol{m}_r=\begin{bmatrix}\boldsymbol{m}_u\\ \boldsymbol{m}_{d,r}\\ \boldsymbol{m}_{g,r}\end{bmatrix} \tag{22}$$

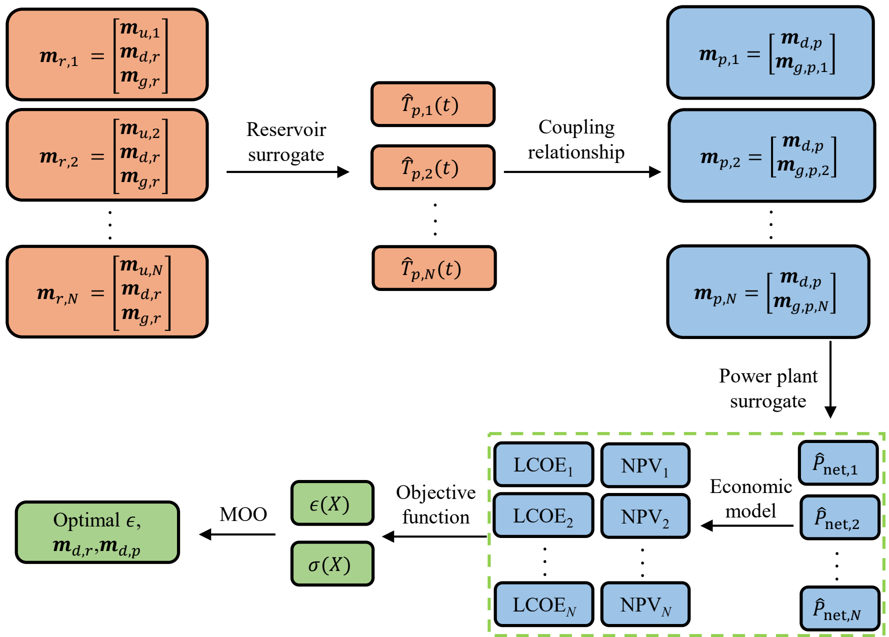

Fig. 7. Workflow of multiobjective optimization considering reservoir uncertainty.

$$\boldsymbol{m}_p = \begin{bmatrix} \boldsymbol{m}_{d,p} \\ \boldsymbol{m}_{g,p} \end{bmatrix} \tag{23}$$

The MOO workflow considering reservoir uncertainty is illustrated in Fig. 7.

The uncertain parameters $\boldsymbol{m}_u$ are assumed to follow a Gaussian distribution, and N equal-probable realizations ($N = 50$ in this work) are established to account for reservoir uncertainty. Each realization can be denoted as $\boldsymbol{m}_{u,i}, i = 1,2,...,N$. Consequently, N sets of input parameters $\boldsymbol{m}_{r,i}$ exist for the reservoir surrogate model, which results in N production temperature profiles $\hat{T}_{p,i}(t)$ and N sets of input parameters $\boldsymbol{m}_{p,i}$ for the power plant surrogate model. Finally, N different electricity and economy results are obtained corresponding to each realization. The objective functions in the MOO can be defined as:

$$\varepsilon(X) = \frac{\sum_{i=1}^{N} X_i}{N} \tag{24}$$

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^{N} (\varepsilon(X) - X_i)^2}{N}} \tag{25}$$

where X can be the cumulative electricity P_{cum}, the LCOE, or NPV. $\varepsilon(X)$ and $\sigma(X)$ are the expectation and standard deviation of X over N realizations, respectively.

In this study, the expectation of each metric (P_{cum}, the LCOE, or NPV) was the primary objective. The standard deviation is interpreted as the risk in achieving the primary objective. Three basic optimization scenarios were tested, including electricity, LCOE, and NPV optimization. In addition, a nonlinear constraint was imposed in LCOE optimization, which ensured that the cumulative electricity generated in any realization was > 0 to prevent negative LCOE results from being selected as the lowest (optimal) results. The constraint was formulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{x} \in \mathscr{X}} \; F_i(\boldsymbol{x}), \quad i = 1,2,\dots,f \\ &\text{s. t. } L_j(\boldsymbol{x}) \leq 0, \quad j = 1,2,\dots,l \\ &\quad H_k(\boldsymbol{x}) = 0, \quad k = 1,2,\dots,h \end{aligned} \tag{26}$$

3. Performance of surrogate models

3.1 Performance of the reservoir surrogate model

For the reservoir surrogate model, the learning rate and batch size were set as 1×10^{-3} and 24, respectively. The total number of epochs was set to 200. The training stopped early when there was no further improvement in the validation loss for 30 epochs. The training and validation losses decreased quickly with increasing number of epochs and converged after approximately 100 epochs, as shown in Fig. 8(a). The training process of the surrogate model was efficient, requiring 475.33 s. When using the testing dataset, the model's predictions closely matched the simulation ground truth, yielding an R^2 score of 0.994 and a relative error of $0.79\% \pm 0.74\%$, as shown in Figs. 8(b) and 8(c). Moreover, the prediction results accurately captured various production temperature decline patterns, showing high agreements with simulation results, as shown in Fig. 8(d). The high prediction accuracy indicated that the reservoir surrogate model can provide reliable production temperature data to be input to the power plant surrogate model in the coupled forward model. The average central processing unit (CPU) time for prediction was 0.0039 s per prediction, which was approximately 1.23×10^5 times faster than the numerical simulation model (~ 8 min per simulation).

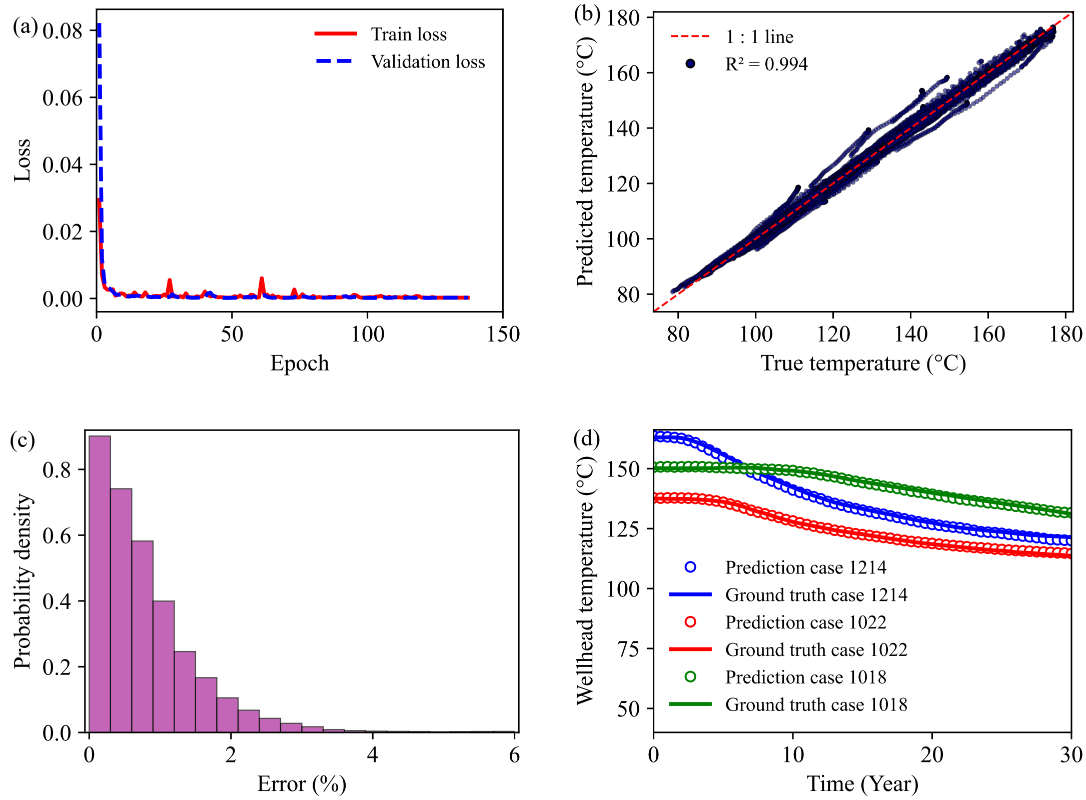

Fig. 8. Performance of the reservoir surrogate model: (a) Loss, (b) R^2 score based on the testing dataset, (c) distribution of the relative error based on the testing dataset and (d) comparisons between simulation and prediction results for typical testing cases.

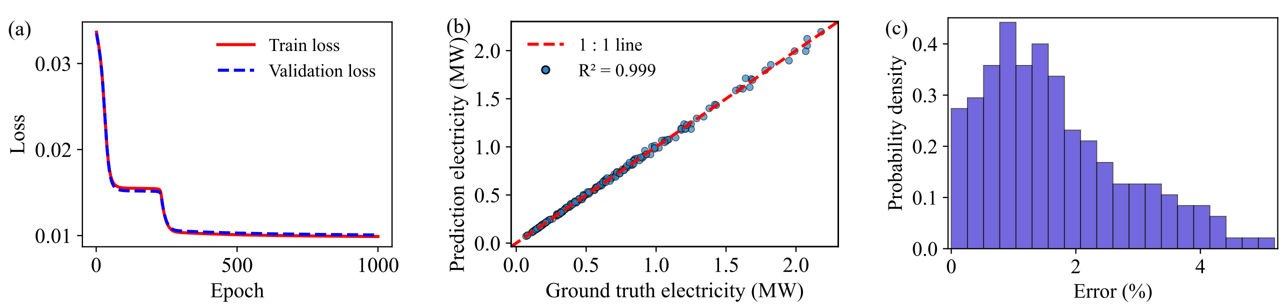

Fig. 9. Performance of the power plant surrogate model: (a) Loss, (b) R^2 score based on the testing dataset, and (c) distribution of the relative error.

3.2 Performance of the power plant surrogate model

The power plant surrogate model had seven hidden layers, with 30 neurons in each layer. The learning rate and batch size were set as 1×10^{-3} and 100, respectively. The total number of epochs was 1,000. The training and validation losses changed similarly and converged well after approximately 400 epochs, as shown in Fig. 9(a). The training process required 206.42 s, exhibiting high efficiency. The well-trained surrogate model accurately predicted the net electricity output, with an R^2 score of 0.999 on the testing dataset, as shown in Fig. 9(b). The predicted electricity output had a low relative error of $1.67\% \pm 1.15\%$, as shown in Fig. 9(c). The average CPU computation time was 0.0017 s per prediction, which achieved approximately 1.77×10^5 times speedup over the in-house power plant simulator (which required 5 min per simulation on average).

4. Optimization results

The high accuracy of the reservoir and power plant surrogate models ensures the reliability of the coupled forward model for optimization.

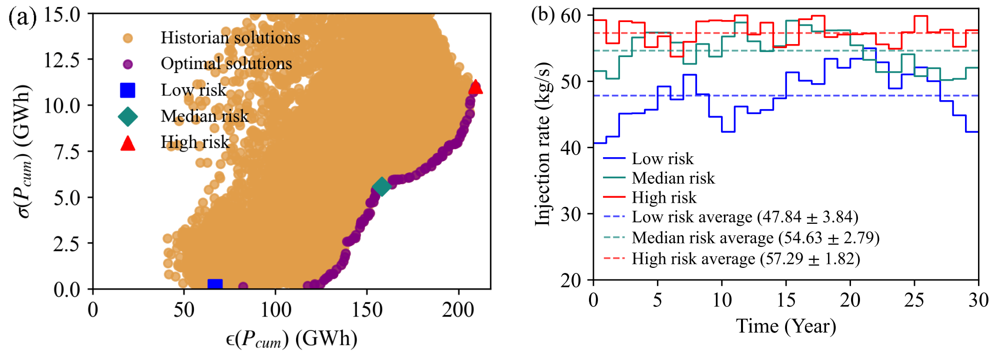

Fig. 10. Solutions of electricity output optimization: (a) Pareto plot and (b) optimal injection rates.

Table 3. Means and standard deviations of uncertain parameters.

Parameter	Unit	Mean	Standard deviation
λ	W/(m·K)	3	0.3
C_p	J/(kg·K)	900	50
α	-	0.6	0.1
E	GPa	25	2.5
ν	-	0.3	0.01

Table 4. Searching ranges of decision parameters.

Parameter	Unit	Lower bound	Upper bound
T_{inj}	°C	70	95
q_{inj}	kg/s	40	60
L_w	m	200	500
$T_{g,d}$	°C	80	160
$q_{g,d}$	kg/s	30	70
$T_{am,d}$	°C	5	40

In the MOO, the given parameter vectors of the reservoir $\boldsymbol{m}_{g,r}$ and power plant $\boldsymbol{m}_{g,p}$, realizations of an uncertain parameter vector $\boldsymbol{m}_{u,i}$, and searching ranges of the decision parameter vectors of the reservoir $\boldsymbol{m}_{d,r}$ and power plant $\boldsymbol{m}_{d,p}$ were consistent across all optimization scenarios. The geothermal gradient G in $\boldsymbol{m}_{g,r}$ was set to 40 °C/km, which is the highest geothermal gradient in the Ifal Basin. The initial reservoir temperature was approximately 129.44 °C, representing a medium-temperature geothermal source for the power plant. For the off-design ambient temperature $T_{am,o}$ in $\boldsymbol{m}_{g,p}$, a significant distinction was observed between the ambient temperatures of the summer (May-October) and winter seasons (November-April) in NEOM (Aljohani et al., 2024), as shown in Fig. S1 (Supplementary file). Therefore, $T_{am,o}$ was set as 31.79 °C and 21.93 °C for summer and winter, respectively. Moreover, $T_{am,o}$ remained consistent throughout the lifespan of the power plant. The uncertain parameters were sampled by following a Gaussian distribution based on their means and standard deviations, as listed in Table 3. The searching ranges of the decision parameters are presented in Table 4. The injection rate q_{inj} was a time sequence with 31 timesteps, and the injection rate at each timestep was a decision parameter.

4.1 Electricity output optimization

The optimization goal was to achieve the highest cumulative electricity output after 30 years while minimizing the standard deviation (risk). The optimized solutions are shown in Fig. 10(a). Among the optimal solutions on the Pareto front (purple points), low-risk (blue square), median-risk (green diamond), and high-risk (red triangle) solutions were selected based on the minimum, median, and maximum standard deviations. The high-risk solutions yield the highest expectation but are accompanied by the highest risk and uncertainty. In contrast, the low-risk solution yields the lowest expectation. Consequently, the median-risk solution was selected. The optimal injection rate is plotted in Fig. 10(b). The mean (dash lines) and standard deviation of the injection rate at each risk level were calculated, as presented in the plot. The expectation, risk, and other decision parameters (except for the injection rate) at each risk level are listed in Table 5.

The solutions at the median- and high-risk levels had similar optimal reservoir decision parameters, indicating that the difference in the cumulative electricity output at different risk levels was mainly due to different power plant designs. For the high-risk solution, the design geothermal temperature $T_{g,d}$ (127.88 °C) was close to the initial wellhead production temperature. Moreover, the design geothermal mass flow rate $q_{g,d}$ (60.40 kg/s) and design ambient temperature $T_{am,d}$ (26.02 °C) were close to the average injection rate (57.29 kg/s) and the average ambient temperature (26.86 °C), respectively. Such values of design geothermal temperature and mass flow rate allow the evaporator to fully utilize the geothermal energy at the early stages of its lifetime when there is no significant decline in the production temperature. Besides, using a design ambient temperature that approximates the average ambient temperature guarantees a satisfactory overall performance of the condenser when the actual ambient temperature (21.93 °C in winter and 31.79 °C in summer) fluctuates around the design value. However, using a high $T_{g,d}$ close to the initial

Table 5. Optimal results of the electricity output optimization scenario.

Risk	T_{inj} (°C)	L_w (m)	$T_{g,d}$ (°C)	$q_{g,d}$ (kg/s)	$T_{am,d}$ (°C)	$\varepsilon(P_{cum})$ (GWh)	$\sigma(P_{cum})$ (GWh)
Low	81.24	337.58	97.73	33.40	7.12	67.14	0.066
Median	73.58	395.10	122.07	51.61	17.13	158.26	5.87
High	72.31	410.64	127.88	60.40	26.02	203.22	10.72

production temperature is risky because the evaporator may perform poorly at later stages of its lifetime when there is a significant production temperature drop. Moreover, using a high $q_{g,d}$ above the upper limit of the actual injection rate (60.00 kg/s) can significantly limit evaporator performance when there is a drastic reduction in the injection rate. In summary, using a high geothermal temperature and mass flow rate to design the power plant is favorable for electricity generation but also increases risk and uncertainty.

In comparison, the optimal $T_{g,d}$, $q_{g,d}$, and $T_{am,d}$ values of the median-risk solution were lower than those of the high-risk solution, where $T_{g,d}$ (122.07 °C) approximated the production temperature at the middle stage and $q_{g,d}$ (51.61 kg/s) was below the average injection rate (54.63 kg/s). To a certain extent, such design conditions limit the capability of the evaporator to utilize the best geothermal source at the early stage but result in a stable performance throughout the lifetime of the power plant. These design conditions generate a moderate amount of electricity with less risk than those in the high-risk solution. The low-risk solution had the lowest optimal $T_{g,d}$, $q_{g,d}$, and $T_{am,d}$ values, which were close to their lower limits. Such a power plant design is safe and has the lowest risk because the evaporator is designed to utilize the geothermal energy even if its temperature or mass flow rate drops to a low value. However, it significantly limits the power plant's capacity during most of its lifetime, resulting in the most conservative electricity output. Additionally, when using a low $T_{am,a}$ value to design the condenser, the condenser must increase the cooling fluid mass flow rate to maintain proper condensation under the high-ambient-temperature condition in summer, which increases electricity consumption and reduces net electricity output.

4.2 Economic optimization

This section discusses the optimization of the reservoir and power plant to achieve the best economic outcomes (lowest LCOE or highest NPV) with the lowest risks. In NPV optimization, two scenarios with and without the FIT are considered. Pareto plots and optimal injection rates of the three optimization scenarios are presented in Fig. 11. The optimal solutions of each scenario are summarized in Table S9 (Supplementary file).

In the LCOE optimization scenario, the optimal decision parameters of the reservoir and the power plant at the median- and high-risk levels followed the same patterns as those in the electricity optimization scenario. Furthermore, the standard deviations of all three risk levels were low; $\sigma(LCOE)$ of the high-risk solution was 0.49, indicating that there is little risk in choosing the high-risk solution as the optimal result, although the median-risk solution is often recommended. Therefore, the optimal LCOE result was 14.81 US¢/kWh. This value is higher than the average LCOE value of a geothermal power plant (6.40-10.60 US¢/kWh) reported by Lazard® (Bilicic and Scroggins, 2023). The main reason was that the LCOE values reported by Lazard® were based on scalable, commercial geothermal projects, wherein several production and injection wells are drilled to achieve high geothermal mass flow rates. However, in our study, only one pair of injection and production wells with a maximum production rate of 60 kg/s was considered, which limited the electricity generation capacity of the power plant and raised the cost of electricity generation.

NPV optimization without the FIT conformed to the current policy implemented in the KSA. The average local electricity price is 0.26 SAR/kWh (6.93 US¢/kWh). Detailed electricity prices in the KSA are presented in Table S8 (Supplementary file). The results of NPV optimization without the FIT were negative at all risk levels, where the best result was US$ -7.92 million, as shown in Fig. 11(b). In this scenario, incentive policies such as FITs become important. The design geothermal temperatures $T_{g,d}$ of all risk levels in this scenario were close to the lower bound (90 °C). This was because electricity generation is a negative factor when the power plant does not make profits. Higher electricity outputs lead to higher installation and O&M costs. Therefore, a low design geothermal temperature is selected to restrict the electricity generation capacity.

To further illustrate the importance of the FIT, another NPV optimization scenario was tested, where the FIT was added to the electricity price as a subsidy. Because no FIT has been implemented in the KSA so far, the reference value of the FIT in Germany was adopted, which is 24 €-ct/kWh (25.2 US¢/kWh). With the incorporation of the FIT, the optimal NPV result changed substantially from US$ -7.92 to 19.53 million. Thus, the FIT will bring more profits to the power plant, encouraging more investments in geothermal electricity and boosting the development of geothermal energy in NEOM, KSA.

In the optimization scenarios of electricity, the LCOE, and NPV with the FIT, the injection rates of the median- and high-risk solutions showed smaller standard deviations than those of the low-risk solutions, indicating that maintaining stable injection and production rates improves power plant performance. This is because the isentropic efficiency of the turbine decreases in the off-design conditions. If the geothermal mass flow rate (i.e., production rate) varies drastically from the

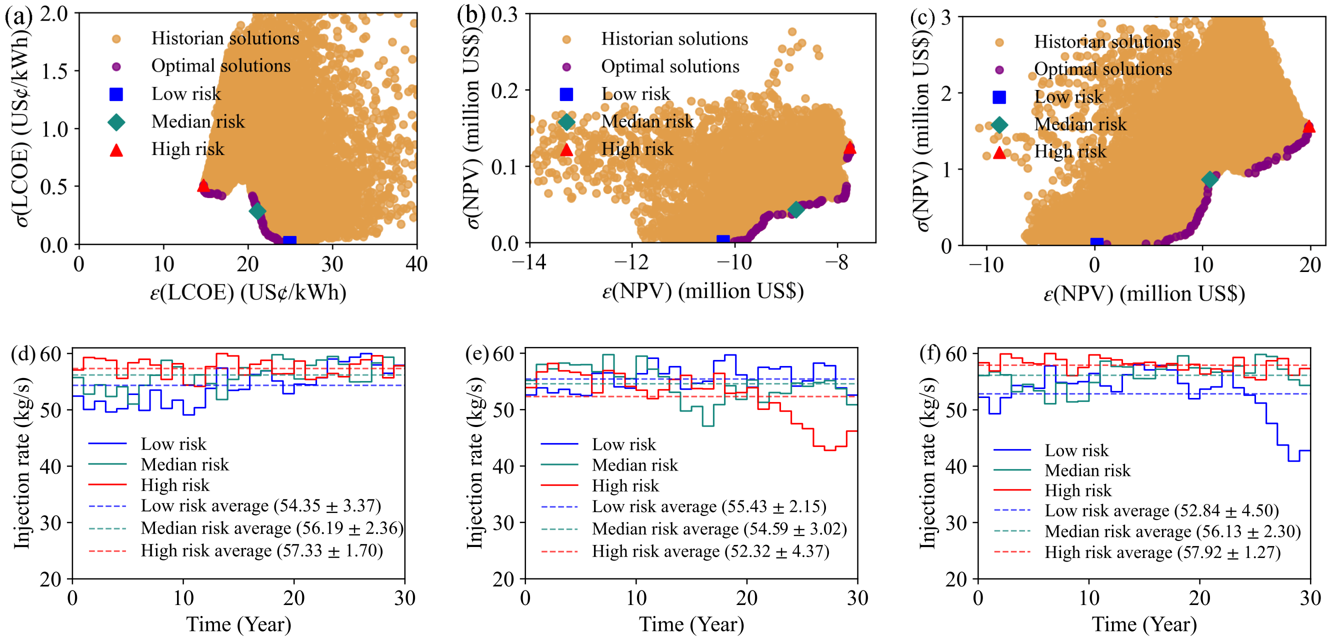

Fig. 11. Solutions of economic optimizations: Pareto plots obtained after (a) LCOE optimization, (b) NPV optimization without the FIT, and (c) NPV optimization with the FIT. Optimal injection rates obtained after (d) LCOE optimization, (e) NPV optimization without the FIT and (f) NPV optimization with the FIT.

design value, the power plant will continuously operate under off-design conditions with a low turbine efficiency during its lifetime, which will result in low electricity outputs and profits. The solutions of NPV optimization without the FIT (Fig. 11(e)) further validated the aforementioned conclusion. In this scenario, the high-risk solution showed a higher standard deviation of injection rates than the low and median-risk solutions. As mentioned before, electricity generation is a negative factor when the power plant does not make profits. Therefore, the high-risk solution changes the injection rate drastically to operate under off-design conditions with low turbine efficiency, which decreases the electricity output to reduce revenue loss. In practice, when there is a significant decline in productivity or injectivity, engineers may consider drilling several infill wells to maintain the productivity and guarantee stable geothermal energy supply to the power plant.

Regarding the computational performance, the average CPU time for one MOO task when using our coupled model was approximately 25.91 min, with 0.014 s per simulation. In comparison, MOO using the simulation models was estimated to require approximately 313 min per simulation, including 8 min for reservoir simulation (in COMSOL) and 305 min for power plant simulation via the in-house simulator. As the simulation consumes most of the CPU time during MOO, it can be concluded that the surrogate-based MOO framework developed in this work can achieve a speedup of $> 1.31 \times 10^6$ times that achieved by the simulator-based MOO framework.

5. Conclusions

This study developed a coupled reservoir-power plant model based on surrogate models to address the high computational costs associated with forward modeling and optimization in conventional simulation-based coupled models. The major conclusions are summarized below.

1) The reservoir and power plant surrogate models accurately predicted the wellhead geothermal production temperature and electricity output with mean relative errors of 0.79% and 1.67%, respectively, while achieving speedups of 1.23×10^5 and 1.77×10^5 times over those of the corresponding simulation models, respectively. The coupled forward model achieved considerable efficiency in geothermal electricity or economy predictions with minor tradeoffs in accuracy, which further accelerates the feasibility assessment procedures of geothermal projects.
2) The implementation of dynamic well controls (injection and production rates) in the reservoir model better aligned with realistic well-control schedules than constant well-control schedules typically assumed in existing studies. Optimization results indicated that low reservoir injection temperature, large well distance, and stable injection rates contributed to good thermodynamic and economic performances of the power plant. The results provide guidance for infill well drilling in geothermal fields.
3) High design geothermal temperatures close to production temperatures in the early production stage and high design mass flow rates close to the average production rate can result in satisfactory electricity outputs and economic outcomes with moderate risks. For power plants in regions with high annual average ambient temperatures, such as NEOM, the design ambient temperature should be close to the ambient temperature in hot seasons (e.g., summer) to allow the condenser to work efficiently

without excessive electricity consumption.

4) The optimal NPV increased significantly from US$ -7.92 to 19.53 million with the support of the FIT. Therefore, the FIT is a crucial incentive policy to ensure the successful operation of the geothermal power plant in NEOM, KSA. The optimization using our surrogate-based coupled forward model was 1.31×10^6 times faster than that using conventional simulation-based coupled models, which significantly accelerates the decision-making procedures of the reservoir and the power plant.

The major limitation of this work is that the working fluid in the binary power plant is R134a, which is commonly investigated in previous studies (Hu et al., 2021) but not necessarily the optimal one for this work. However, the results of different working fluids are not compared and the optimal working fluid is not selected in this study. The main obstacle was that power plant simulation using different working fluids will significantly increase computational costs compared to that when using only one working fluid. This hampered dataset generation for the surrogate models and their training. In future, the optimal working fluid will be determined. The major bottleneck remains the utilization of more efficient optimization algorithms or CPU parallelism in our in-house power plant design and off-design simulator to enhance simulation efficiency and reduce dataset-generation costs.

Acknowledgements

The authors acknowledge funding support from King Abdullah University of Science and Technology (KAUST), Saudi Arabia (No. BAS/1/1423-01-01).

Supplementary file

https://doi.org/10.46690/ager.2025.10.07

Conflict of interest

The authors declare no competing interest.

References

Aboud, E., Alqahtani, F., Abdulfarraj, M., et al. Geothermal imaging of the saudi cross-border city of neom deduced from magnetic data. Sustainability, 2023, 15(5): 4549.

Al-Laboun, A., Al-Quraishi, A., Zaman, H., et al. Reservoir characterization of the burqan formation sandstone from midyan basin, northwestern Saudi Arabia. Turkish Journal of Earth Sciences, 2014, 23(2): 204-214.

Aljohani, Z., Asiri, A., Al-Awlaqi, S., et al. Assessment of solar energy availability and its potential applications in neom region. Renewable Energy Research and Applications, 2024, 5(1): 11-19.

Bilicic, G., Scroggins, S. Lazard's levelized cost of energy analysis-version 16.0. Lazard, April, 2023.

Chen, G., Jiao, J., Wang, Z., et al. Multi-fidelity machine learning with knowledge transfer enhances geothermal energy system design and optimization. Advances in Geo-Energy Research, 2025, 16(3): 244-259.

Deb, K., Pratap, A., Agarwal, S., et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

Feng, Z., Yan, B., Shen, X., et al. A hybrid CNN-transformer surrogate model for the multi-objective robust optimization of geological carbon sequestration. Advances in Water Resources, 2025, 196: 104897.

Garapati, N., Adams, B. M., Bielicki, J. M., et al. A hybrid geothermal energy conversion technology-a potential solution for production of electricity from shallow geothermal resources. Energy Procedia, 2017, 114: 7107-7117.

Gudala, M., Govindarajan, S. K. Numerical modeling of coupled fluid flow and geomechanical stresses in a petroleum reservoir. Journal of Energy Resources Technology, 2020, 142(6): 063006.

Gutiérrez-Negrín, L. C. Evolution of worldwide geothermal power 2020-2023. Geothermal Energy, 2024, 12(1): 14.

Haklıdır, F. S. T. The importance of long-term well management in geothermal power systems using fuzzy control: A western Anatolia (Turkey) case study. Energy, 2020, 213: 118817.

Hoteit, H., He, X., Yan, B., et al. Uncertainty quantification and optimization method applied to time-continuous geothermal energy extraction. Geothermics, 2023, 110: 102675.

Hsieh, J. C., Li, Y., Lin, Y., et al. Off-design performance and economic analysis in coupled binary cycle with geothermal reservoir and turbo-expander. Energy, 2024, 305: 132274.

Hu, S., Yang, Z., Li, J., et al. Thermo-economic optimization of the hybrid geothermal-solar power system: A data-driven method based on lifetime off-design operation. Energy Conversion and Management, 2021, 229: 113738.

Hughes, G. W. a. G., Johnson, R. S. Lithostratigraphy of the red sea region. GeoArabia, 2005, 10(3): 49-126.

Jiansheng, W., Lide, S., Qiang, Z., et al. Numerical investigation on power generation performance of enhanced geothermal system with horizontal well. Applied Energy, 2022, 325: 119865.

Ke, T., Huang, S., Xu, W., et al. Study on heat extraction performance of multiple-doublet system in hot sedimentary aquifers: Case study from the Xianyang geothermal field, northwest China. Geothermics, 2021, 94: 102131.

Lai, C. S., McCulloch, M. D. Levelized cost of electricity for solar photovoltaic and electrical energy storage. Applied Energy, 2017, 190: 191-203.

Li, Y., Peng, G., Du, T., et al. Advancing fractured geothermal system modeling with artificial neural network and bidirectional gated recurrent unit. Applied Energy, 2024, 372: 123826.

Liu, Z., Gudala, M., Katterbauer, K., et al. Robust optimization of fully coupled geothermal reservoir and power plant system. Available at SSRN 5318891, 2025.

Luo, W., Kottsova, A., Vardon, P. J., et al. Mechanisms causing injectivity decline and enhancement in geothermal projects. Renewable and Sustainable Energy Reviews, 2023, 185: 113623.

Markó, Á., Mádl-Szőnyi, J., Brehme, M. Injection related issues of a doublet system in a sandstone aquifer-a generalized concept to understand and avoid problem sources in geothermal systems. Geothermics, 2021, 97: 102234.

Meng, N., Gao, X., Wang, Z., et al. Numerical investigation and optimization on dynamic power generation performance of enhanced geothermal system. Energy, 2024, 288: 129901.

Müller, S., Schüler, L., Zech, A., et al. Gstools v1. 3: A toolbox for geostatistical modelling in python. Geoscientific Model Development, 2022, 15(7): 3161-3182.

Multiphysics, C. Introduction to comsol multiphysics®. COMSOL Multiphysics, Burlington, MA, 1998, 9(2018): 32.

Murdock, H. E., Gibb, D., André, T., et al. Renewables 2020-global status report. Paris, REN21 Secretariat, 2020,

Paszke, A., Gross, S., Massa, F., et al. Pytorch: An imperative style, high-performance deep learning library. ArXiv Preprint ArXiv: 1912.01703v1, 2019.

Qin, Z., Jiang, A., Faulder, D., et al. Efficient optimization of energy recovery from geothermal reservoirs with recurrent neural network predictive models. Water Resources Research, 2023, 59(3): e2022WR032653.

Shi, Y., Song, X., Wang, G., et al. Study on wellbore fluid flow and heat transfer of a multilateral-well CO_2 enhanced geothermal system. Applied Energy, 2019, 249: 14-27.

Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need. ArXiv Preprint ArXiv: 1706.03762, 2017.

Wan, Y., Yuan, Y., Zhou, C., et al. Multiphysics coupling in exploitation and utilization of geo-energy: State-of-the-art and future perspectives. Advances in Geo-Energy Research, 2023, 10(1): 7-13.

Wang, N., Chang, H., Kong, X., et al. Deep learning based closed-loop well control optimization of geothermal reservoir with uncertain permeability. Renewable Energy, 2023, 211: 379-394.

Xue, Z., Chen, Z. Deep learning based production prediction for an enhanced geothermal system (EGS). Paper SPE-212754-MS Presented at SPE Canadian Energy Technology Conference, Calgary, Alberta, Canada, 15-16 March, 2023.

Yan, B., Gudala, M., Hoteit, H., et al. Physics-informed machine learning for noniterative optimization in geothermal energy recovery. Applied Energy, 2024, 365: 123179.

Yan, B., Gudala, M., Sun, S. Robust optimization of geothermal recovery based on a generalized thermal decline model and deep learning. Energy Conversion and Management, 2023, 286: 117033.