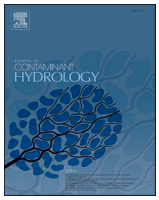
Journal of Contaminant Hydrology 277 (2026) 104876

Contents lists available at ScienceDirect

Journal of Contaminant Hydrology

journal homepage: www.elsevier.com/locate/jconhyd



A multiple surrogate simulation-optimization framework for designing pump-and-treat systems

Chaoqi Wang [a], Zhi Dou [a,*], Ning Chen [b], Yan Zhu [a], Zhihan Zou [a], Jian Song [a], Shen-Huan Lyu [b,c,d,**]

[a] School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
[b] Key Laboratory of Water Big Data Technology of Ministry of Water Resources, College of Computer Science and Software Engineering, Hohai University, Nanjing, China
[c] Department of Computer Science, City University of Hong Kong, Hong Kong, China
[d] State Key Laboratory for Novel Software Technology, Naniing University, Naniing, China

ARTICLE INFO

Keywords:
Groundwater contaminant remediation
Pump-and-treat optimization
Surrogate modeling
Multi-model framework
Machine learning
Genetic algorithm

ABSTRACT

Pump-and-treat (P&T) remediation is a widely adopted and effective method for groundwater contamination control. It is important to optimize the operation schemes (pumping well locations and pumping rates) to maximize contaminant removal efficiency and minimize operational costs. Recently, surrogate models have been integrated with optimization algorithms to formulate the remediation schemes. However, with various surrogate techniques available, their comparative performance in P&T remediation tasks and potential for combined usage of multiple surrogates require further exploration. In this study, five popular surrogate models—Kriging, Polynomial Interpolation, Support Vector Regression (SVR), Random Forest (RF), and Deep Neural Network (DNN)—were evaluated for their ability to predict contaminant removal efficiency under diverse schemes in a multi-contaminant site. The analysis revealed that, while DNN achieved the highest overall prediction accuracy in the validation stage across the 200 cases, no single surrogate model consistently outperformed the others in all individual cases. A multi-surrogate optimization framework, coupling all five models with a genetic algorithm, was developed to enhance P&T schemes. The usage of multiple surrogates finally brings benefits because the complementary strengths of diverse surrogate models are combined. We identified remediation schemes that achieved superior contaminant removal (17.5% residual contaminant) compared to the other results (19.2–21.7%). The framework offers a robust tool for environmental management and insights for advancing studies related to surrogate-based optimization.

1. Introduction

Groundwater is a vital natural resource (Liang et al., 2025; Masocha et al., 2020; Ning et al., 2024; Rabbani et al., 2025; Rao et al., 2022). On one hand, it provides primary supply for drinking water, industrial consumption and agricultural irrigation in many regions (Baskaran and Abraham, 2022; He et al., 2024; Li et al., 2024). On the other hand, it plays a key role in maintaining ecological balance by sustaining base-flows and wetland habitats (Fan et al., 2022; He et al., 2019, 2021). However, with expanding industrial activities and urban development, groundwater systems are being increasingly contaminated by human activities (Hamutoko et al., 2016; Uugwanga and Kgabi, 2021). The degraded water quality is posing significant risks to human health and aquatic ecosystems.

Various remediation methods have been developed to address the increasingly severe groundwater contamination phenomenon. These methods are broadly categorized into in-situ and ex-situ approaches (Truex et al., 2017; Truex et al., 2015). In-situ methods include permeable reactive barrier, chemical reduction, bio-reduction, and electrokinetic remediation techniques (Thornton et al., 2014; Wang et al., 2025; Xu et al., 2024; Zhu et al., 2020), which treat contaminants in place. In contrast, ex-situ methods involve the extraction of contaminated groundwater or soil for off-site treatment. As one of the ex-situ methods, pump-and-treat (P&T) technique is especially effective and

* Corresponding author.
** Correspondence to: Key Laboratory of Water Big Data Technology of Ministry of Water Resources, College of Computer Science and Software Engineering, Hohai University, Nanjing, China.
E-mail addresses: douz@hhu.edu.cn (Z. Dou), lvsh@hhu.edu.cn (S.-H. Lyu).

https://doi.org/10.1016/j.jconhyd.2026.104876
Received 17 September 2025; Received in revised form 29 January 2026; Accepted 31 January 2026
Available online 4 February 2026
0169-7722/

popular (Zhang et al., 2025).

The P&T technique has been deemed as one of the most critical, invaluable (Carroll et al., 2024), and widely applied (Truex et al., 2017; Truex and Johnson, 2017) method for groundwater contamination remediation. Simply put, the P&T technique works by using a series of extraction wells to pump contaminated groundwater to the surface for treatment (Zhang et al., 2025). The P&T technique is valid for both dissolved solute contaminants (Chang et al., 2007; Qiang et al., 2024; Song et al., 2025) and non-aqueous phase liquid (NAPL) contaminants (Bae et al., 2024; Ciampi et al., 2023). Besides, it has been proved valid under complex geological conditions (Ciampi et al., 2023; Zha et al., 2019). However, operational costs of pump-and-treat method is high due to energy-intensive pumping and long-term maintenance. Additionally, the P&T remediation efficiency typically exhibits a progressive decline in the later stages (Harvey et al., 1994). This is primarily because contaminants could be trapped in low-permeability zones such as silt and clay horizons or the matrix in fractured rocks. These trapped contaminants are released slowly through slow advection or back diffusion processes. Thus, the P&T efficiency subsequently enters a prolonged phase of persistently low extraction levels that may persist for years. This pattern is reflected by the tailing effect of contaminant concentrations in pumping wells (Carroll et al., 2024; Zha et al., 2019). As a result, a considerable amount of residual contamination may persist and require further treatment (Carroll et al., 2024; Zha et al., 2019). Empirical evidence indicates that residual contaminants persist after P&T treatment regardless of the aquifer conditions and implementation setups (Pedretti et al., 2014). Currently the P&T method still cannot achieve perfect remediation and fully restore aquifers to pre-contamination conditions. Therefore, it is necessary to optimize the P&T strategies to maximize contaminant removal efficiency (Kuo et al., 1992; Thornton et al., 2014) and minimize total remediation costs (Chang et al., 2007; Qiang et al., 2024).

Recently, P&T system design has been realized through simulation-optimization (S-O) frameworks. Firstly, numerical models are developed to represent the contaminated site's geometry and hydrogeological properties, and then applied to simulate the processes of groundwater flow and contaminant transport to the extraction pumping wells for removal. Secondly, the simulation model is integrated with optimization methods, which iteratively adjust well locations and pumping rates until it finds the optimal solution (either maximizing contaminant removal efficiency or minimizing total remediation costs). Various optimization algorithms have been employed to design the P&T systems. The early work by (Kuo et al., 1992) demonstrated that the simulated annealing (SA) algorithm is effective for P&T optimization. And they have specified that the simulated annealing algorithm is effective because it could avoid local optima and explore the global optimal. Chang et al. (2007) and Qiang et al. (2024) employed the genetic algorithm (GA) to optimize P&T scheme. Wang and Zheng (1997) highlighted that genetic algorithm as a global search method that is advantageous for designing groundwater pump-and-treat remediation schemes. Zheng and Wang (1999) further proposed an integrated optimization method that combines the advantages of tabu search and linear programming; this method achieved reduced computation cost and enhance remediation efficiency. (Elshall et al., 2020) employed the covariance matrix adaptation evolution strategy (CMA-ES), which is a multi-objective framework, to find the P&T scheme that achieves the treatment target with minimum pumping rate.

Nevertheless, the simulation-optimization framework typically demands numerous model evaluations for effective convergence. For the P&T design task, the simulation model is for the transient transport processes, which requires iterative computations at each time step (Hou et al., 2016). This generates a heavy calculation load and considerable time consumption, ultimately may create computationally prohibitive tasks (Li et al., 2021). To overcome this challenge of high computation cost, surrogate modeling has emerged as an essential technique (Qiang et al., 2024; Song et al., 2025).

Surrogate models are computationally cheaper models designed to approximate the dominant features of a complex model (Asher et al., 2015). Through various strategies, such as statistical methods or machine learning techniques, these models can learn the relationships between input parameters and output responses of the simulation model. Subsequently, the surrogate model generates predictions using input parameters, without performing the numerical simulations of the original model. The major motivation of using a surrogate model is to reduce the prohibitively high computation cost (Razavi et al., 2012). According to (Asher et al., 2015; Robinson et al., 2008), the current surrogate methods can be divided into three categories, including data-driven methods, projection-based methods and multi-fidelity methods. Luo et al. (2023) recommended using data-driven surrogate models for groundwater decision support problems (including remediation design, monitoring network design). Among them, machine learning methods like support vector machine (Ouyang et al., 2017) and artificial neural networks (Secci et al., 2022a; Zhou and Tartakovsky, 2021) have also been widely used. Recent Deep Neural network (DNN) surrogates have shown particular strength in high-dimensional groundwater problems that traditional methods fail to handle due to the curse of dimensionality (Mo et al., 2019). Notably, convolutional architectures excel at high-dimensional inverse modeling (e.g., simultaneous source and conductivity identification), while long short-term memory (LSTM) models perform exceptionally well in temporal contaminant forecasting (Li et al., 2021). Further, novel generative adversarial networks have been used to construct surrogates (Deng et al., 2025).

Recent advancements in multi-fidelity surrogate modeling integrate data from varying fidelity levels to boost accuracy and efficiency in computationally demanding scenarios. Giselle Fernández-Godino et al. (2019) provided decision criteria for multifidelity surrogates, emphasizing benefits in high-fidelity data scarcity, Lee et al. (2024) fused coupled and decoupled models, achieved increased surrogate accuracy under fixed budgets. Notably, (Lee et al., 2025) developed an adaptive quality-based multi-fidelity (AQBMF) framework that ranks and combines low-fidelity sources, surpassing traditional methods in benchmarks by filtering low-quality data and optimizing ensembles. Similarly, (Lee et al., 2026) utilized multi-fidelity techniques with similar low-fidelity data from different conditions, achieving up to 60% efficiency improvements and 15–20% gains in energy density.

Although surrogate models have been widely applied in hydrogeology, their use in the specific area of P&T design is limited. Currently, only a few studies have been identified that develop surrogate models for P&T system design: a Kriging model (Qiang et al., 2024; Zhang et al., 2022), neuro network models (Majumder and Eldho, 2020; Song et al., 2025), and the analytic element method (Matott et al., 2006), as well as several surrogates (Luo and Lu, 2014). The lack of systematic performance comparisons obscures the applicability of various surrogate methods to P&T design. On the other hand, current studies often rely on a single surrogate model (Qiang et al., 2024; Song et al., 2025), while (Forrester and Keane, 2009) emphasized that no surrogate method universally outperforms others, as each has unique strengths. Considering this, (Matott et al., 2006; Xing et al., 2019) suggest employing multiple surrogate models simultaneously. (Viana et al., 2009) proposed a framework using multiple surrogate models to enhance prediction accuracy.

This study aims to address the identified limitations related to surrogate modeling for pump-and-treat scheme design. Firstly, to tackle the scarcity of systematic comparisons, we assess the prediction accuracy of five common surrogate techniques: Kriging, Polynomial Interpolation, Support Vector Regression, Random Forest, and Deep Neural Network. Secondly, we investigate the joint use of these surrogate models: each trained model is integrated with optimization algorithms to search for the P&T scheme with the lowest residual contamination. The removal efficiencies of all surrogate models are compared to identify the most effective approach. This research represents an innovative exploration in pump-and-treat technique through novel testing of a multi-surrogate

framework. This approach delivers practical benefits directly for enhancing remediation efficiency and reducing environmental impact. The novel framework and findings may offer valuable insights for simulation-optimization applications in fields like aquifer characterization and contaminant source identification.

2. Methods

This study employs a simulation-optimization framework to optimally design the pump-and-treat (P&T) schemes. As illustrated in Fig. 1, the P&T scheme optimization process involves three steps.

First, a numerical simulation model is developed to simulate the groundwater flow and contaminant transport processes during the pump-and-treat operation. The operation scheme parameters, including well locations and pumping rates are randomly generated and incorporated into the simulation model to simulate the different contaminant removal processes. For each simulation, the final residual contaminant mass (normalized) in the synthetic aquifer is collected. These data are then compiled and integrated to form a training dataset.

Second, surrogate models are constructed and trained using the above training dataset, where the P&T configuration (well locations and pumping rates) are input and the residual contaminant is the output. This process enables the models to capture patterns from the data and replace the computationally intensive numerical simulation model. Five distinct surrogate models—Kriging, Polynomial Interpolation (PolyInterp), Support Vector Regression (SVR), Random Forest (RF), and Deep Neural Networks (DNN)—are tested to evaluate their accuracy. These models are selected because they have been used and proven to be effective in previous hydrogeological applications, also because these methods have various functioning mechanisms.

Third, an optimization algorithm is implemented, to determine the operation scheme with minimized the total contaminant mass in the synthetic aquifer. In this work, the genetic algorithm is employed as the optimization method. Because it is found to be able to effectively handle the complex, nonlinear nature of P&T design, to explore a wide range of well locations and pumping rates, as demonstrated in previous studies (Chang et al., 2007; Rudiyanto et al., 2023).

At last, the optimization results from different surrogate models are compared to establish a robust multi-surrogate simulation-optimization framework. The derived optimal well locations and pumping rates are subsequently validated through numerical simulation to assess their actual performance.

2.1. Numerical simulation model

2.1.1. Governing equations

The groundwater system was numerically modeled using a two-dimensional porous medium formulation. The governing equations for fluid flow and contaminant transport are implemented through the following coupled processes. The steady-state flow field was determined by combining Darcy's law with the continuity equation:

$$\boldsymbol{u} = -T\nabla h \tag{1}$$

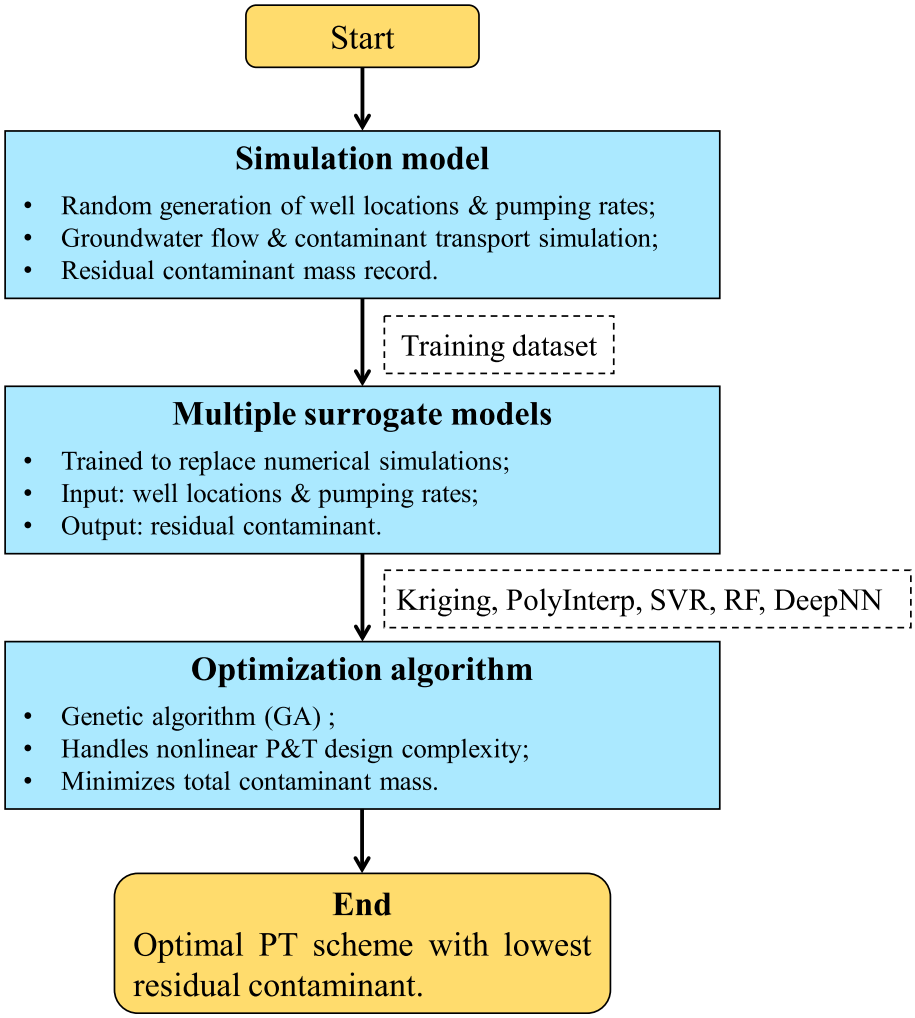


Fig. 1. The pump-and-treat scheme optimization using the multiple-surrogate simulation-optimization framework.

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \nabla\cdot(\rho\boldsymbol{u}) = Q_s \tag{2}$$

where h is the hydraulic head (L), Q_s is the mass source term (M/TL^3), T is hydraulic conductivity (L/T), $\boldsymbol{u}$ is the flow velocity vector (L/T), ε is porosity (/), ρ is fluid density (M/L^3). The derived velocity field was subsequently applied in the advection-dispersion equation:

$$\frac{\partial}{\partial t}(\varepsilon C) + \boldsymbol{u}\cdot\nabla(\theta C) = \nabla\cdot(\varepsilon(D_c + \boldsymbol{u}\alpha)\nabla C) + R \tag{3}$$

where C is concentration (M/L^3), D_c is the diffusion coefficient (L^2/T), R is the source term (M/TL^3), α is the dispersivity term (L). The partial differential equation system is discretized and solved numerically using COMSOL Multiphysics® simulation softwares (Multiphysics, 1998), employing the finite element method with appropriate boundary conditions and convergence criteria.

2.1.2. Setups

Section 2.1.2 Setups describes the configuration of the synthetic aquifer and the P&T system design. The aquifer properties, emission details, and P&T operation parameters are comprehensively summarized in Table 1. The two-dimentional synthetic aquifer is modeled as a 6 km × 6 km region (Fig. 2a) and a 2 km × 2 km site (Fig. 2b). We set up a relatively small site because when applied to larger regions, pump-and-treat (P&T) may face limitations in terms of operational costs and remediation efficiency. A 6 km × 6 km buffer region surrounds the site to simulate external conditions; the geometric center is at (0,0) in a Cartesian coordinate system. The aquifer has a thickness of 30 m, Porosity (ε) is 0.25, and fluid density (ρ) is 1000 kg/m^3.

As shown in Fig. 2, the hydraulic conductivity ($\boldsymbol{K}$) field is assumed to be log-normally distributed, ranging from 10^{-6} to 10^{-4} m/s. The heterogeneous conductivity field was generated using the geostatistical toolbox of the "GSTools Python library" developed by (Müller et al., 2022). A Gaussian variogram model was applied with a variance of 1 as the original conductivity field ($\mathbf{logK_0}$). The spatial correlation length is assigned to be 100 m. In the GSTools algorithm, the correlation length of 100 m represents moderate spatial variability: for distances <100 m, adjacent grid points exhibit stronger similarity in K values. Thus we can foster some stagnant zones that trap contaminants and slowly release. To adjust the values into specific bounds of ($\log K_{min} = -6$ and $\log K_{max} = -4$ m/s), the field $\mathbf{logK_0}$ then undergoes a global min-max scaling using the formula:

$$\mathbf{logK}_{scaled} = \frac{\mathbf{logK_0} - \min(\mathbf{logK_0})}{\max(\mathbf{logK_0}) - \min(\mathbf{logK_0})} \times (\log K_{max} - \log K_{min}) + \log K_{min} \tag{4}$$

In this scaling operation, every value is transformed proportionally without truncation and without altering relative spatial patterns. This allows convenient control over the maximum and minimum values.

To ensure the hydrogeological parameters for the synthetic aquifer are well-justified, we selected realistic values reflecting practical conditions. The aquifer dimensions, thickness, porosity, and fluid density represent typical characteristics of contaminated sandy aquifers. The hydraulic conductivity range and spatial correlation setups align with common field-scale log-normal distributions. The hydraulic gradient, calculated as (90 m - 60 m) / 6000 m = 0.005 (0.5%), mirrors gentle slopes found in natural aquifer systems.

Because the positions and pumping rates of each well remain constant throughout the entire extraction process in our design, we made a simplification assumption and implemented steady-state flow simulation. While it does not capture early transient drawdown or flow direction changes, we focus more on the overall extraction process. Steady-state flow is driven by constant-head boundary conditions: a higher hydraulic head (h = 90 m) is arranged at the left boundary (x = −3000 m) and a lower hydraulic head (h = 60) m is at the right boundary (x = 3000 km). Thus, a left-to-right hydraulic gradient is established. No-flow conditions are applied at the top and bottom (y = 3000, −3000 m) boundaries. The initial head field was set with a linear interpolation between 90 m at the left boundary and 60 m at the right boundary, ensuring a smooth transition across the domain. The steady-state groundwater flow dynamics are simulated in the synthetic aquifer system under prescribed boundary conditions.

The synthetic aquifer is assumed to be contaminated by five factories. As shown in Fig. 3, three contaminants (C1, C2, C3) are introduced by these factories: the first two factories emitting C1, the third and fourth emitting C2, and the last one emitting C3. All of the setups about the factories are randomly generated, including locations, emission rates, concentrations. The release durations are assumed to be 10 years for these factories.

The concentration fields in Fig. 3 will be used as the initial condition of our P&T system. Note that the emission concentrations for these factories are 77, 50, 17, 30 and 120 mol/m^3, respectively. The first two factories represent average conditions, whereas the third and fourth exhibit lower concentrations but cover larger areas. In contrast, the fifth factory shows higher concentrations within a smaller area. The contaminants with varied concentrations and spatial distributions should compose a complex contamination scenario. This complexity poses challenges for P&T design. However, addressing this complexity should enhance the significance and practical utility of this paper.

In this site, we have set up five pumping wells to extract the above contaminants. The locations of five wells can be anywhere within this site. Constrained by operational cost, the total flow rate of five pumping wells is fixed at 8000 m^3/day. The flow rates are distributed among the five wells in varying proportions. This results in 15 adjustable parameters: for each well, two coordinates (x, y) define its location, and one proportion determines its flow rate, where the actual flow rate is calculated by multiplying the proportion by 8000 m^3/day. These two kinds of parameters are also the key adjustable parameters for enhancing remediation efficiency in previous studies (Huang and Mayer, 1997; Song et al., 2025).

The P&T operation are simulated for a duration of 10 years and the contaminant removal result can be evaluated. The well locations and the flow rate proportions for the five wells will be adjusted collectively to optimize removal efficiency.

Table 1
Hydrogeological and Operational Parameters for P&T Scheme Optimization.

Parameter	Value/Description
Aquifer Dimensions	2 km × 2 km (4 km^2) site, 6 km × 6 km buffer
Aquifer Thickness	30 m
Porosity (ε)	0.25
Fluid Density (ρ)	1000 kg/m^3
Hydraulic Conductivity (K)	Log-normally distributed (Spatial correlation 100 m), 10^{-6} to 10^{-4} m/s
Boundary Conditions	Left: h = 90 m; Right: h = 60 m; Top/Bottom: No-flow
Contaminant Sources	5 factories emitting, three contaminant categories C1, C2, C3
Release Duration	10 years
P&T Wells	5 wells, locations adjustable within site
Adjustable Parameters	15 (2 coordinates +1 flow proportion per well)
P&T Operation Duration	10 years

2.1.3. Performance metrics

In this work, the total residual ratio ($R_{residual}$) is adopted as the contaminant removal performance metric for various pump-and-treat (P&T) schemes. The metric is defined as:

$$R_{residual} = \sum_{i=1}^{3} \frac{M_{C_i,residual}}{M_{C_i,initial}} = \frac{M_{C_1,residual}}{M_{C_1,initial}} + \frac{M_{C_2,residual}}{M_{C_2,initial}} + \frac{M_{C_3,residual}}{M_{C_3,initial}} \tag{5}$$

where $M_{Ci,residual}$ represents the residual mass of contaminant C_i after

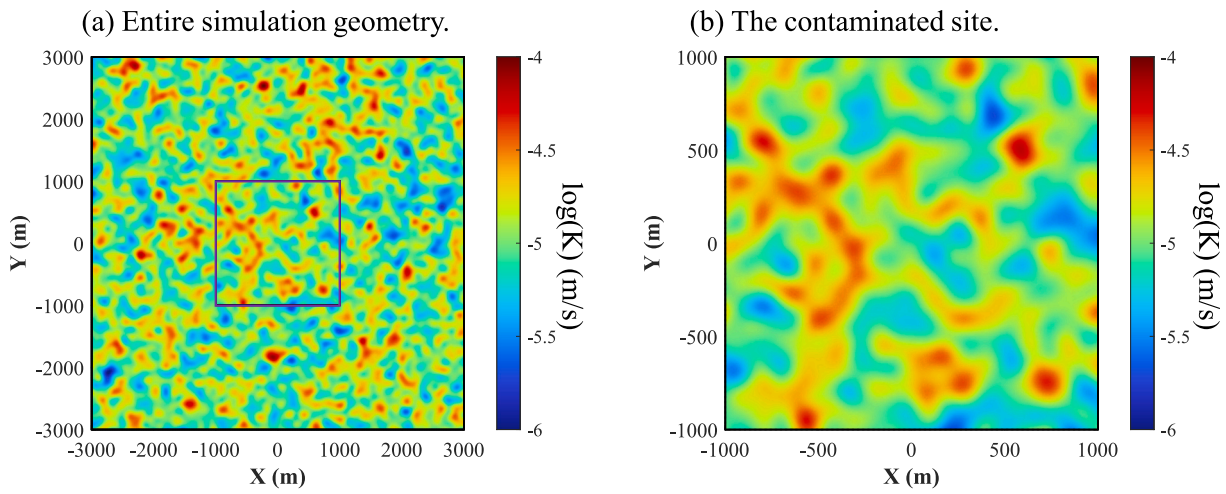


Fig. 2. Spatial distribution of hydraulic conductivity (K) field.

remediation; $M_{Ci,\text{initial}}$ denotes the initial mass of contaminant C_i. The main reason for choosing this metric is that there are three contaminants (C1, C2, C3) from five factories, each contaminant varies in concentration and may exhibit severe toxicity. This metric emphasizes the proportional reduction of each contaminant; thus, the low-concentration contaminants are not overlooked.

The alternative metric, total residual contaminant mass ($M_{\text{total}} = M_{C1} + M_{C2} + M_{C3}$), is not used. Because, when optimization relies on the total mass of three contaminants, the remediation would prioritize contaminants of larger quantity and may neglect the contaminants with lower masses.

2.2. Surrogate models

This subsection introduces five surrogate models aimed for pump-and-treat (P&T) optimization: Kriging, Polynomial Interpolation, Support Vector Regression (SVR), Random Forest (RF), and Deep Neural Network (DNN). The five surrogate models were selected from three distinct methodological families: (1) geostatistical interpolation (Kriging), (2) deterministic algebraic interpolation (Polynomial), and (3) machine learning approaches (SVR, RF, DNN). The mechanism and strengths of these models are outlined here in the following subsections. Note that Kriging, Polynomial, SVR, and RF models have been implemented on MATLAB R2022b, DNN has been implemented on Python 3.9 with PyTorch Deep Learning toolbox.

These models are trained and validated using the dataset generated by the numerical model, with input features comprising the 15 adjustable parameters: x and y coordinates and flow rate proportions for each of the five pumping wells. The output is the total residual ratio (R_{residual}, Eq. 4), which reflects the overall remediation efficiency. After training, the surrogate models can provide highly efficient predictions of remediation efficiency (R_{residual}) based on these inputs with specified values. So the surrogates can replace the numerical model for rapid P&T optimization.

2.2.1. Kriging

Kriging has been widely adopted in geostatistical interpolation and has recently gained prominence as an effective surrogate modeling technique for groundwater systems. The Kriging interpolation process is shown in Fig. 4, where values at unmeasured locations (estimated values) are predicted based on measured data points (measured values).

When it comes to mathematical formulation, Kriging is mainly composed of a polynomial trend and a random process. The polynomial trend (low-order linear or quadratic polynomial) is employed to capture the global mean response of the system; the random process accounts for local variations by quantifying spatial correlations that decrease with distance with covariance functions. Together, they enable Kriging to provide more accurate interpolation predictions. Kriging can be effective for approximating complex, spatially correlated systems with sparse data. The Kriging method assumes a constant mean and variance across the domain. This may oversimplify complex aquifer heterogeneities and lead to smoother predictions that underestimate the extreme or irregular distribution of parameters. More detailed explanations of the kriging can be referred to (Zhang et al., 2022).

In this study, Kriging surrogate models were implemented using Gaussian Process Regression (GPR) in MATLAB. The hyperparameters included a squared exponential kernel function to capture spatial correlations (akin to ordinary Kriging), feature standardization set to true for normalization. Training was performed on 1000 samples, with validation on 200 samples to evaluate performance metrics such as RMSE and R^2.

2.2.2. Polynomial interpolation

Polynomial Interpolation surrogate offers a mathematically straightforward approach to approximate numerical model responses. This deterministic method constructs a single algebraic polynomial that passes through all training data points. While lower-order polynomials ($n \leq 3$) are typically employed for practical applications to avoid Runge's phenomenon (Burden and Faires, 2011).

In this work, we assigned that the polynomial function in the surrogate model takes the form:

$$f(\mathbf{x}) = \beta_0 + \Sigma_{i=1}^{n}\beta_i x_i + \Sigma_{i=1}^{n}\beta_{i,i}x_i^2 + \Sigma_{i=1}^{n-1}\Sigma_{j=i+1}^{n}\beta_{i,j}x_i x_j \quad (6)$$

where $\mathbf{x} = (x_1, x_2, \ldots. x_n)$ represents the 15-dimensional input vector (e.g., well locations, pumping rates), and β_0, β_i, $\beta_{i,i}$, $\beta_{i,j}$, are coefficients for the constant, linear, squared, and interaction terms, respectively. These coefficients are determined by fitting a linear regression model to the training data, the least-squares error would be minimized to ensure accurate approximation of the numerical model outputs. Polynomial Interpolation assumes the system response can be approximated by a smooth, continuous polynomial function. Thus, it may make less accurate predictions for highly nonlinear or discontinuous relationships in complex groundwater systems.

In this study, polynomial regression surrogate models were implemented in MATLAB. The model incorporated second-order polynomial features, consisting of linear terms for each of the 15 inputs, squared terms for each input, and pairwise interaction terms between inputs. We employed the Iteratively Reweighted Least Squares (IRLS) with the default 'bisquare' weight function and a tuning constant of 4.685, so that observations with large residuals are downweighted iteratively until

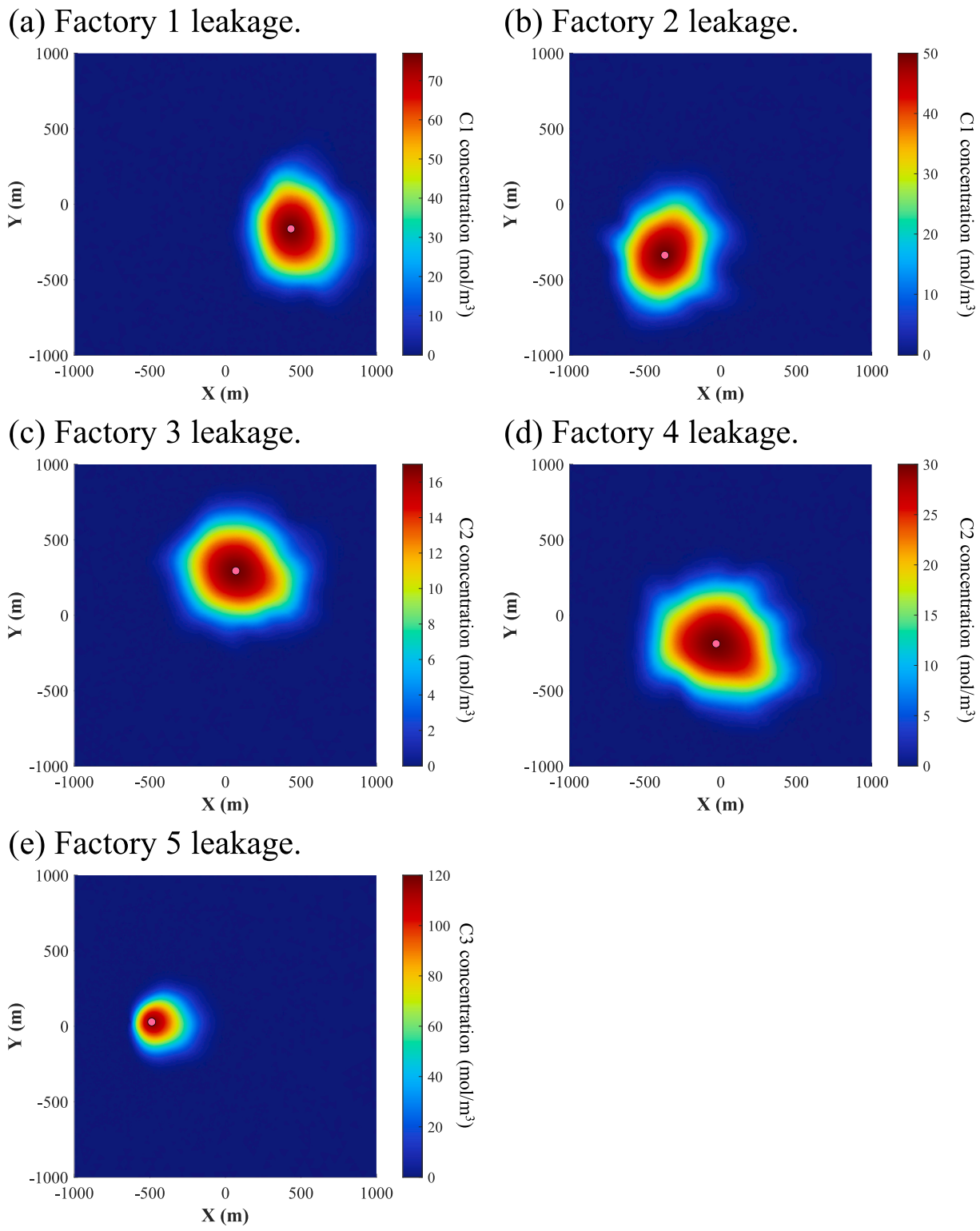


Fig. 3. The initial contaminant plume distribution caused by five factories (The pink points represents factory locations). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

convergence.

2.2.3. Support vector regression

Support Vector Machine (SVM) is a powerful supervised learning algorithm developed by (Cortes and Vapnik, 1995) based on statistical learning theory and structural risk minimization. This method is originally designed for binary classification. As shown in Fig. 5, it works by identifying a decision boundary (hyperplane) that maximizes the margin between classes. Fig. 5 exhibits a classification question where the bondary is linear. For nonlinear problems, kernel functions (e.g., Gaussian or polynomial) would be used to implicitly transform data into a higher-dimensional space; then complex nonlinear relationships would be transformed into linearly separable problems. This elegant mathematical framework guarantees a global optimum through convex

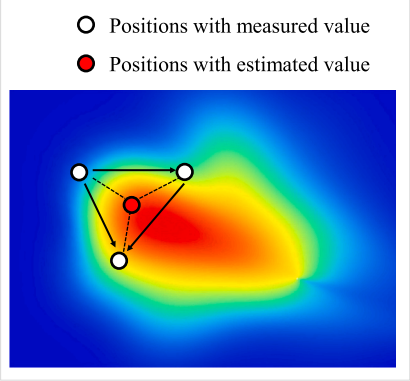


Fig. 4. Schematic of Kriging interpolation.

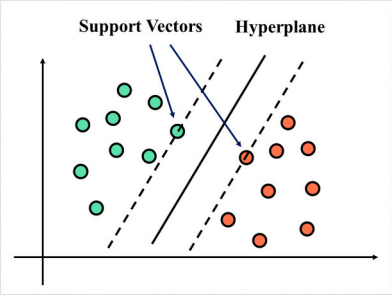


Fig. 5. Schematic of Support Vector Machine.

optimization (Boyd, 2004; Boyd et al., 2011). Thus, the SVM method has been recognized as both theoretically sound and computationally efficient (Bennett and Parrado-Hernández, 2006; Vapnik, 2000).

Support Vector Regression (SVR) is the extension of SVM's principles to regression tasks. Instead of maximizing class separation, SVR fits a hyperplane within a tolerance margin (*e*-insensitive tube), penalizing only deviations larger than *e* (as shown in Fig. 6). SVR is able to handle high-dimensional data and sparse samples. It has been recorded to be effective for predicting contaminant concentrations in groundwater systems (Ouyang et al., 2017). SVR assumes that complex relationships can be captured by mapping data into a higher-dimensional space via kernel functions. Thus it exhibits flexibility in modeling more nonlinear patterns. This flexibility is expected to better handle irregular or non-smooth parameter distributions, compared to Kriging's assumptions.

In this study, Support Vector Regression model employed a Gaussian (RBF) kernel function to handle nonlinear relationships. Following a trial-and-error test, the optimal parameters were determined as follows: C_{box} = 10.00 (the box constraint, a regularization parameter that balances low training error with model complexity by constraining the Lagrange multipliers), Epsilon = 0.01 (the epsilon-insensitive margin width, which establishes a tolerance band around predictions where deviations are not penalized; lower values heighten sensitivity to errors), and KernelScale = 1.00 (the scaling factor governs the kernel's sensitivity to input variations).

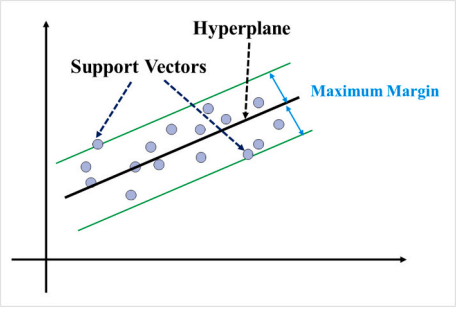


Fig. 6. Schematic of Support Vector Regression.

2.2.4. Random Forest

Random Forest (RF) is an ensemble machine learning method that constructs multiple decision trees during training and outputs their averaged predictions (Breiman, 2001). As shown in Fig. 7, each decision tree functions as a hierarchical predictor that recursively partitions the feature space through optimized binary splits, where the splitting criteria maximize information gain for classification tasks or minimize prediction error for regression. Predictions are generated by propagating input features through each tree's split rules until reaching terminal nodes containing the final output values. RF assumes that the averaging predictions from multiple decision trees can effectively capture complex and heterogeneous patterns. This adaptability may improve predictive accuracy for non-linear problems.

The model's robustness stems from two fundamental randomization techniques: bootstrap aggregating, where individual trees train on randomly sampled subsets of the original data, and feature subspace selection, where each split considers only a fraction of available features. Key hyperparameters requiring optimization include the number of constituent trees, maximum allowable tree depth, and minimum samples required for node splitting, typically tuned through cross-validation procedures. This ensemble approach enables RF to effectively model complex, high-dimensional relationships characteristic of groundwater systems (Z. Wang et al., 2024). In this study, each Random Forest surrogate model consisted of 1500 decision trees. They were configured for regression mode and all available predictors were used at each split for feature selection. This implemented bootstrap aggregating without random subspace sampling. A minimum leaf size of 5 is implemented.

2.2.5. Deep neural network

Artificial Neural Networks (ANNs) represent a foundational machine learning approach inspired by biological neurons. Artificial Neural Networks (ANNs) are biologically inspired computational systems that process information through interconnected layers of neurons, as shown in Fig. 8. In each neuro, there would be weighted summation of inputs, bias addition, and nonlinear transformation via activation functions.

The DNNs assume that complex relationships can be learned through layered transformations by the neurons and the nonlinear activation functions. The network's predictive output is refined through back-propagation, where prediction errors are propagated backward to adjust weights and biases via gradient descent, progressively minimizing the

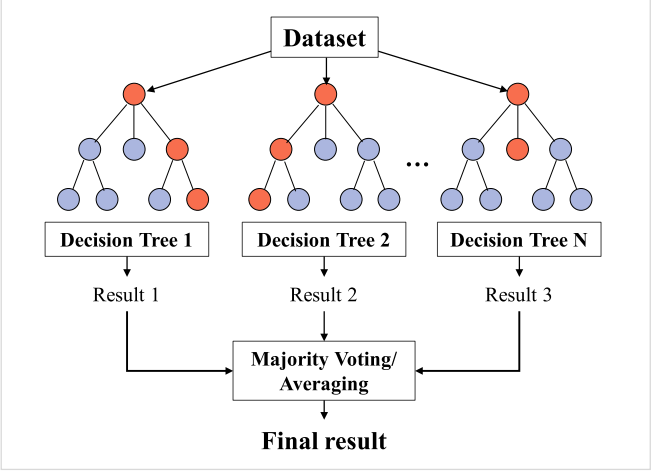


Fig. 7. Schematic of Random Forest.

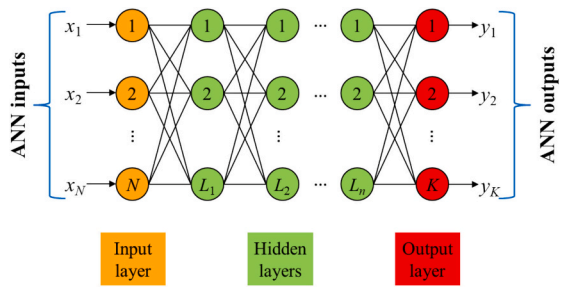


Fig. 8. Schematic of artificial neural network (modified after (Wang et al., 2024a, 2024b)).

discrepancy between simulated and predicted contaminant distributions. While shallow networks suffice for linear relationships, deeper architectures (more than 3 hidden layers) demonstrate superior performance by capturing complex nonlinear interactions (Zhou et al., 2021).

In this study, Deep Neural Network surrogate models were implemented in PyTorch. It included eight hidden layers with neuron counts of 4800, 2400, 2400, 1200, 1200, 600, 600, and 120, respectively. Leaky ReLU activation functions were used with a negative slope of 0.01 for nonlinearity. Training employed the Adam optimizer with a fixed learning rate of 0.00005. Mean squared error served as the loss function. The model with the lowest validation loss was saved.

2.2.6. *Comparative analysis of surrogate models*

The above surrogate methods have different working mechanisms, thus different suitable application scenarios. In the comparative study by (Villa-Vialaneix et al., 2012, the task is approximating N_2O fluxes and N leaching), Kriging is more accurate than other models for small datasets; and for large datasets, random Forest is more accurate than SVR and Kriging; but SVR handles noisy data well. In the review work by (Razavi et al., 2012), they noted that Kriging is more effective for low-dimensional problems; Polynomials can be used as global surrogates (fitting the entire input space) or in local optimization contexts; neuro network is highly effective for non-linear, complex response surfaces, it is the most used surrogate in the reviewed studies. According to the input dimension (D) and training data sample size (n), (Forrester and Keane, 2009) provided suggestions on the application of the different surrogates.

As noted by (Asher et al., 2015), such comparison literatures are numerous. We cannot list all their results. So, we provide Table 2 to summarize the studies that used these models (at least two papers), application areas, and suggested application conditions. Note that the suggestions are preliminary, and it still requires actual testing to determine the best model for a specific application scenario. Readers interested in further details can refer to the cited studies.

However, Forrester and Keane (2009) emphasized that no surrogate method universally outperforms others, results will depend on the application scenario, and factors such as the input dimension and the size of training set (Breiman, 2001). So Viana et al. (2009) suggested using multiple surrogates in a single framework. Matott and Rabideau (2008) have proved that using multiple surrogates improved the optimized objective function and reduced runtime.

For our problem of P&T system optimization, the input dimension is 15 (D = 15), the training dataset has 1000 samples ($n = 1000$), and the problem complexity/nonlinearity is somehow uncertain. Thus, all five models is potential to make effective predictions.

2.3. *Genetic algorithm for optimization*

In this study, a Genetic Algorithm (GA) was employed to optimize pump-and-treat schemes for groundwater remediation, targeting the minimization of residual pollutant mass across three distinct contaminants. GA is an evolutionary optimization technique inspired by natural selection processes. It could effectively address complex, non-linear problems in hydrogeology by navigating high-dimensional parameter spaces (Maier et al., 2014; Zheng et al., 1999) and escaping local optima (Singh and Datta, 2006).

Unlike gradient-based methods which rely on local gradient information and may converge to local optima in non-convex problems, GA's evolutionary approach could effectively handle the nonlinearity and multimodality of P&T optimization problem. Although GA requires more iterations of the forward model, the use of low-cost surrogate models significantly reduces computational expense. Gradient-based methods may struggle with the discontinuous relationships in our P&T problem. So the GA method is more suitable for this study.

In our implementation, the population of each generation comprises 1000 candidate solutions. Each candidate solution is a 15-dimensional vector of the parameters including well locations and pumping rates. The objective function to be minimized is the normalized residual contaminant (R_{residual}). Our implementation of the GA follows the following procedures:

(1) Initialization: The initial population is established by randomly generating 1000 individuals. Each individual represents a potential P&T operation scheme defined by the 15 parameters.
(2) Selection: The surrogate models calculate the total residual ratio (R_{residual}) to evaluate contaminant removal efficiency for all P&T schemes in the current population. Selection probabilities are

Table 2
Comparison of surrogate models and application suggestions. (D represents input dimension, n represent training data size).

Model	Successful Applications	Application areas.	Suggested application conditions
Kriging	(Garcet et al., 2006; Qiang et al., 2024; Zhang et al., 2022)	Nitrate Leaching Process Modeling, groundwater remediation, contaminant source identification, pump-and-treat optimization etc.	D < 20, $\boldsymbol{n}$ < 500. Applicable to complex non-linear problems;
Polynomial Interpolation	(Singh and Verma, 2019; Zaghiyan et al., 2021)	Groundwater level, water quality.	D ⟨20,n⟩ 500. For simple problem with smooth variations.
Support Vector Regression (SVR)	(Ly et al., 2013; Yoon et al., 2011)	Groundwater level prediction, Rainfall data analysis	D > 20, $\boldsymbol{n}$ > 500. Robust for nonlinear, high-D input. Noised data.
Random Forest (RF)	(Pham et al., 2020; Schoppa et al., 2020)	Stream and flood discharge forecast.	n > 500. Applicable to complex non-linear problems
Deep Neural Network (DNN)	(Chen et al., 2020; Dawson and Wilby, 1999; Deng et al., 2024; Secci et al., 2022b; Somogyvári et al., 2017; Yoon et al., 2007; Zhi et al., 2024)	Heat transfer in fractured media, contaminant transport.	Most widely adopted; For various scenarios with large datasets

then computed using a fitness-proportional formula; individuals with lower $R_{residual}$ values (indicating better remediation performance) have higher probabilities of being selected. Also note that the individuals with lower $R_{residual}$ values may be selected multiple times during sampling. This implements the survival-of-the-fittest principle.

(3) Reproduction: Using the selected individuals from the previous step, the new generation is created through four distinct yet complementary mechanisms: (i) Elitism preservation: 250 selected individuals are directly transferred to the next generation without modification; (ii) Mutation: 250 new individuals are generated by applying controlled Gaussian perturbations (σ = 0.1) to randomly selected parent solutions; (iii) Crossover: 250 offspring are produced through uniform crossover of parameter sets from parent pairs; (iv) Diversity injection: 250 completely new solutions are randomly generated within the defined parameter bounds to maintain population diversity. Together, these four steps produce a new generation comprising 1000 individuals.

(4) Evaluation: The $R_{residual}$ scores of the new population are calculated using surrogate models, with the lowest score recorded as an indicator of the GA's convergence progress.

(5) Iteration: Steps 2–4 are repeated until a convergence is achieved. The P&T strategy is optimized for minimizing the total residual ratio ($R_{residual}$) across the three contaminants.

The GA algorithm is run independently for each surrogate model (Kriging, PolyInterp, SVR, RF, DNN) to generate the optimized P&T parameters. To remove the effect of random initial population on optimization results, we performed 20 independent optimization runs (each with a different random initialization) for each surrogate. The resulting P&T parameters were evaluated in the high-fidelity simulation model, and the best-performing scheme (lowest true residual contaminant mass) from these runs was selected as the representative result for that surrogate. Finally, the overall best P&T scheme was identified by comparing the selected schemes across all surrogates.

3. Results

3.1. Numerical simulation model

We conducted 1200 simulations, generating a dataset with 1000 samples for training and 200 samples for validating the prediction accuracy of these surrogate models. These datasets were created using the synthetic simulation model (Section 2.1), with inputs consisting of 15 randomly generated adjustable parameters: the x, y coordinates and pumping rates for five wells. Outputs are the total residual contaminant percentage values of three contaminants, ranging from 20% to 200% (Fig. 9). A significant number of values distributed around 60–80%. The range is wide, thus random P&T configurations may yield poor performance. Note that from the training-validation dataset, the minimum residual percentage of the three contaminants is 20.278%. It will be interesting to check whether our optimization method could find a configuration with residuals lower than 20.278%.

One example of the simulated contaminant transport process by the numerical simulation model is shown in Fig. 10. The simulation tracks the migration of three contaminants—C1, C2, and C3—over a 10-year period. As pumping wells extract groundwater, the plumes are drawn toward the wells.

The shape of the contaminant plume changes over time. At the beginning (Fig. 10a, e, i), they are distributed in round shapes due to injection. As pumping starts (Fig. 10b, f, j), the plumes are stretched by the pumping wells. The contaminant plumes begin to move toward the pumping wells, while the change is minimal. By the 5th year (Fig. 10c, g, k), significant pollutant mass has been extracted into the wells. At the final frame at 10th years (Fig. 10d, h, l), the contaminant plume develops into a band-like pattern, distributed near the pumping wells. Contaminants C_1 and C_2 exhibited higher residual levels near the pumping wells, while C_3 exhibits comparatively less residual. Indicating more effective remediation. Have a basic knowledge about these plume evolutions may also help contaminant remediation efforts.

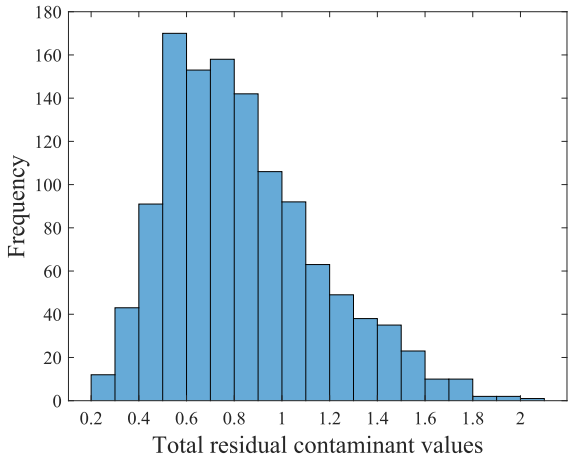


Fig. 9. Histogram of total residual contaminant percentages ($R_{residual}$) for various P&T configurations.

It is rather challenging to determine the optimal pump-and-treat (P&T) scheme via reasoning. On one hand, it is possible that using dispersed well locations may be advantageous, as they can cover a broader area of the plume. On the other hand, it is also possible that using concentrated well placement can be advantageous, as it could enhance extraction efficiency in high-concentration zones. To overcome this confusion, we need the simulation-optimization techniques to determine the most effective P&T strategy.

3.2. Surrogate model

3.2.1. Overall validation

As mentioned in Section 3.1, 1000 datasets are used to train the surrogate models (including Kriging, Polynomial Interpolation, Support Vector Regression, Random Forest, and Deep Neural Network). The performance of these five models was evaluated using 200 validation datasets. This evaluation has been realized via two key metrics: Root Mean Square Error (RMSE) and correlation coefficient (Corre). The validation results are shown in Fig. 11 and Table 3. RMSE measures the average error by comparing the predictions by the surrogate models and the numerical simulation outputs. A lower RMSE indicates more precise predictions. The correlation coefficient assesses the linear relationship between predictions and numerical simulation outputs. A high Corre ensures accurate representation of complex plume dynamics. Using both metrics together ensures models are precise (low RMSE) and capture key patterns (high Corre).

As shown in Table 3, among the various evaluated surrogate models, Deep Neural Network (DNN) performed the best, with the lowest RMSE value of 0.1503 and the highest correlation coefficient of 0.8761. The predictions are more approximate to the actual data, as shown in Fig. 11e. Kriging and Support Vector Regression showed comparable performance: the RMSE values (0.1604 and 0.1623) and Corre values (0.8597 and 0.8604) are not far from DNN.

Polynomial Interpolation is less accurate, with an RMSE of 0.1794 and Corre of 0.8220. The Random Forest model is the least effective, with the highest RMSE of 0.2125 and the lowest Corre of 0.7342. As shown in Fig. 11d, it is clear that the predictions show more substantial deviations from the 'y = x' line.

3.2.2. Scenario-specific performance analysis

Recognizing that "no surrogate modeling method universally

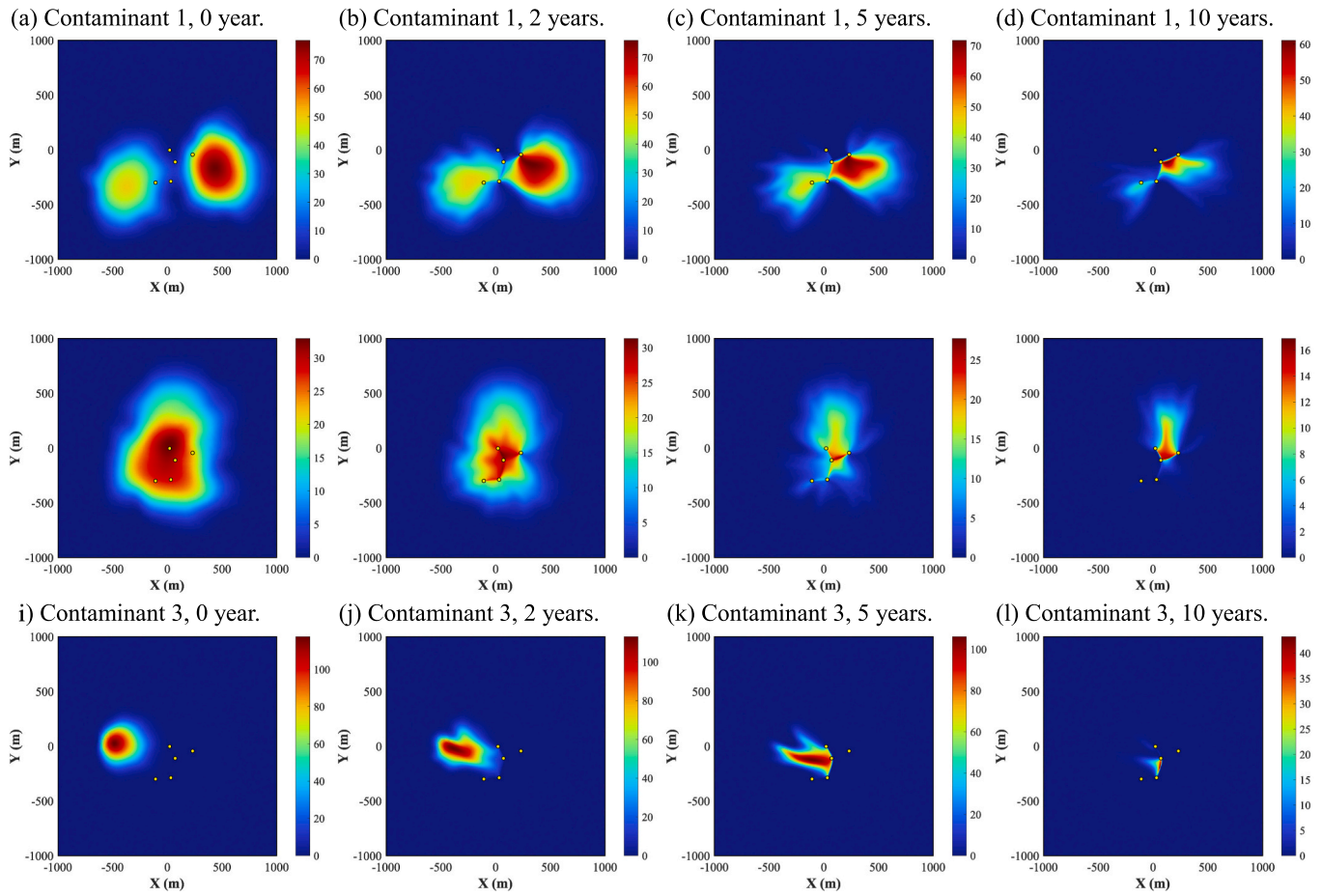


Fig. 10. Contaminant Transport process during the P&T process.

outperforms others, as each possesses unique strengths and weaknesses" (Forrester and Keane, 2009), we checked on all of the single instances of the validation set to reveal the scenario-dependent performance variations. Specifically, we recorded the frequency at which each model achieved top performance (lowest prediction error) for each individual validation instances. The results are shown in Fig. 12 and Table 4.

The results demonstrate three model, SVR, RF and DNN, demonstrated comparable effectiveness in achieving optimal predictions: 47, 47 and 46 good instances respectively. Polynomial Interpolation showed moderate performance, it has been the best for 34 validation cases. Kriging is the least frequent top-performer for only 26 instances.

Interestingly, despite DNN exhibited overall best validation performance in subsection 3.2.1, it performed best in only 46 instances. This means it is definitely not universally optimal. Notably, Kriging showed comparatively good accuracy in the overall validation, but it only performed best for just 26 instances; the Random Forest model, showed comparatively poor accuracy (Table 3), but it performed best for 47 instances, which is actually high. These findings reflect that each model possesses unique scenario-dependent strengths. This aligns with (Forrester and Keane, 2009).The results suggest that we should jointly use the multiple surrogate models, rather than relying on a single 'best' model" within the subsequent inversion framework.

It is possible that the residual contaminant mass exhibit highly nonlinear characteristics with the variation of the P&T configuration parameter space (15 adjustable parameters: well locations and pumping rates). For the three models (SVR, RF, and DNN) that excel in scenarios with complex, nonlinear relationships, so they have more top-performing instances (47, 47, 46). PolyInterp, assuming a smooth polynomial function, performs well in simpler, less nonlinear configurations, resulting in fewer top instances (34). Kriging's homogeneity assumption limits its ability to model complex, nonlinear P&T outcomes, explaining its lowest top instances (26).

It is possible that the residual contaminant mass exhibits highly nonlinear characteristics with variations in the P&T configuration parameter space (15 adjustable parameters: well locations and pumping rates. The three models SVR, RF, and DNN are better at treating complex and nonlinear scenarios because the assumption of these models. Thus they provided more top-performing instances (47, 47, 46). The PolyInterp model assumes a smooth polynomial function, thus it provided fewer top instances (34). Kriging's homogeneity assumption limits its ability to model complex, nonlinear P&T outcomes, explaining its lowest top instances (26).

3.3. Optimization results

Table 3 presents the inversion results of groundwater remediation schemes obtained from five surrogate models, including optimized well coordinates (X, Y) and corresponding pumping rates. The remediation effectiveness, total residual contaminant values R_{residual}, from five optimization processes are also provided. The spatial distribution of these well locations is visually represented in Fig. 13, where each dot indicates a well position, and the size of the dot corresponds to the pumping rate, with larger dots signifying higher pumping rates.

Via the total residual contaminant (R_{residual}) value, we check the cleanup effectiveness of each optimized scheme. Using Kriging, we obtained the highest total residual of 21.69%, which is least effective among the five. PolyInterp achieves the lowest total residual of 17.48%, which is the most successful pollution cleanup strategy. For the last three models, SVR, Random Forest and DNN, the performances are moderate, $R_{\text{residual}} = 19.31\%$, 19.41% and 19.24% respectively. Despite

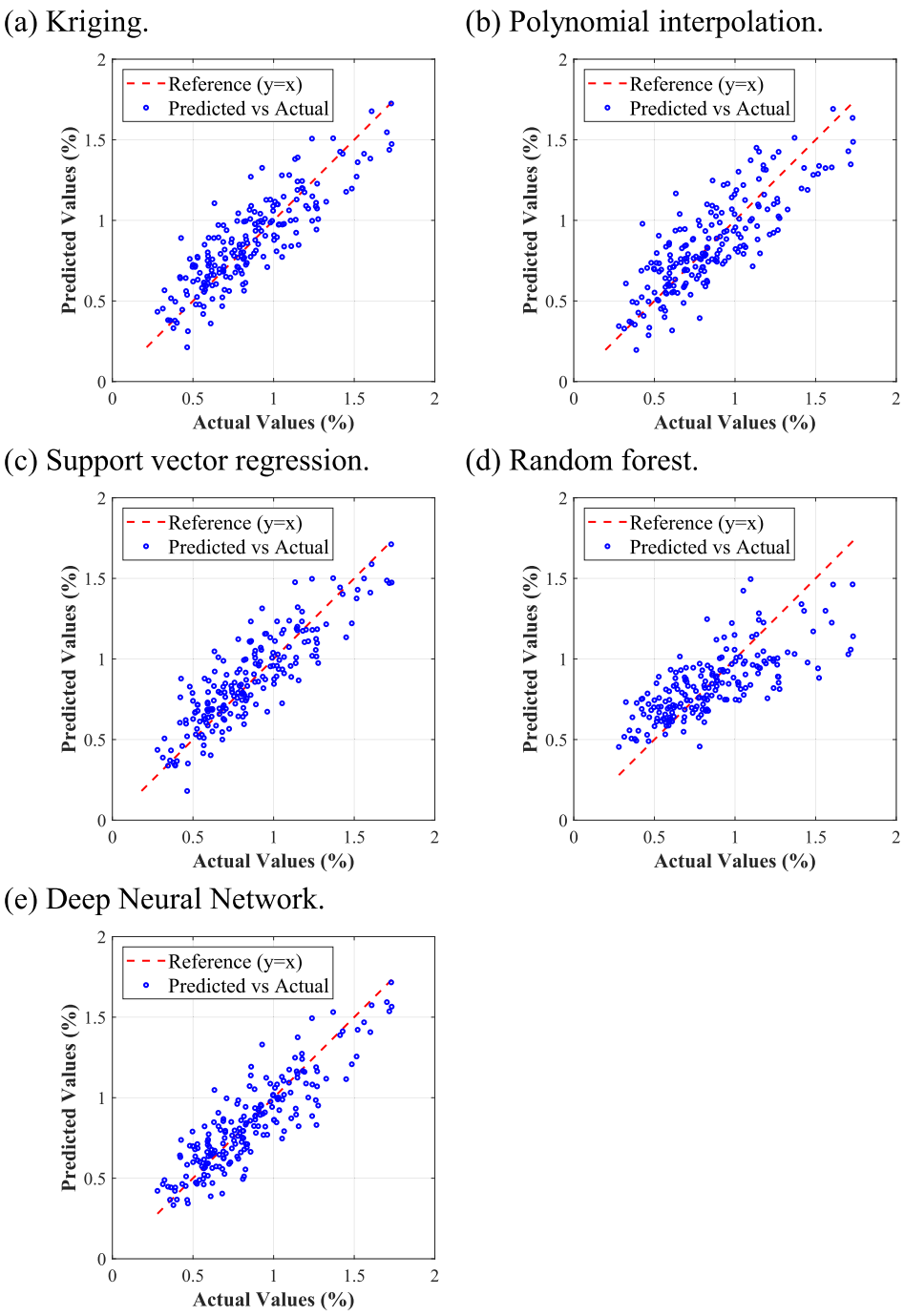


Fig. 11. The scatter plots of the validation results of five surrogates.

Table 3
The validation results of five surrogates.

Models	RMSE	Corre
Kriging	0.1604	0.8597
PolyInterp	0.1794	0.8220
SVR	0.1623	0.8604
RF	0.2125	0.7342
DNN	0.1503	0.8761

all schemes exhibits optimized results, PolyInterp provides the best pollution cleanup outcome, with the lowest residual.

The scheme offered by the PolyInterp model (Fig. 13b) is kind of special, it suggested the pumping flow rate of well 3 should be reduced to a near-negligible value (80 m^3/day). This value is merely 1% of total pumping rate. This result implies that it may be unnecessary to implement this well 3. If we can reduce the quantity of pumping wells, the cost for drilling and installing pumping equipment would be saved by 20%. The other surrogate model didn't offer such suggestions. Remarkably, this simplified configuration maintains superior remediation performance: the residual amount ($R_{residual}$) of 17.48% is lower than the best results from alternative models ($R_{residual}$ range: 19.2–21.7%). It is well recognized that P&T systems requires high operational costs due to long-term pumping and well maintenance. From an operational cost perspective, eliminating one pumping well significantly reduces ongoing expenses. Thus, it would be more cost-effective and practical for real-world P&T system implementation.

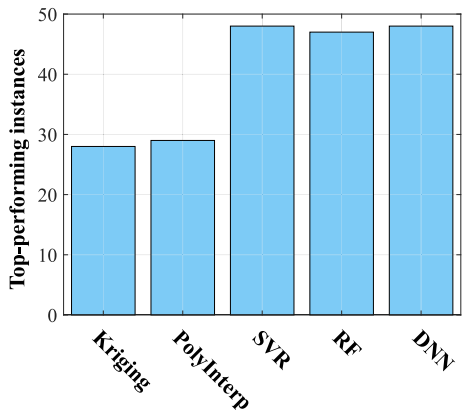


Fig. 12. Top-performing instances of five surrogates.

Table 4
Top-performing instances of five surrogates.

Models	Top-performing instances
Kriging	26
PolyInterp	34
SVR	47
RF	47
DNN	46

3.4. Additional validation of out-of-distribution performance

To further substantiate the benefits of the multi-surrogate ensemble in P&T optimization, we performed 200 additional independent genetic algorithm optimization runs for both the proposed multi-surrogate approach and a single-surrogate. For the baseline surrogate model, we employed the DNN model, who has the highest overall accuracy. All optimized P&T schemes are then evaluated on the numerical simulation model. The distributions of verified residual contaminant mass are presented in Fig. 14 as histograms. According to the Figure, the ensemble-surrogate inversion approach exhibits a tighter and more left-shifted distribution, which may suggest stronger robustness.

Recall that the training dataset for the surrogates exhibited a minimum residual contaminant mass of 20.278%. We adopt this minimum residual contaminant mass as a reference threshold: solutions with verified residual mass below this threshold are considered as one successful optimization, as they surpass the best outcome observed during surrogate construction. Out of the 200 runs, the single-surrogate (DNN) model produced 28 solutions that has a residual concentration below 20.278%, so the frequency is 14%; while the multi-surrogate ensemble achieved this in 64 runs, so the frequency is 32%. This demonstrates that the ensemble approach more frequently identifies high-performing remediation designs that exceed the reference performance level.

Table 6 summarizes key performance statistics from the 200 verified runs. The multi-surrogate ensemble clearly outperforms the strongest single-surrogate alternative, delivering a lower median (21.03% vs. 24.05%), lower mean (21.25% vs. 24.63%), and better optimal solution (15.97% vs. 16.91%).

Notably, the globally best design (15.97% residual mass) originated from the SVR surrogate rather than the DNN, highlighting the practical advantage of incorporating diverse surrogates. These results confirm that even a straightforward ensemble strategy provides measurable improvements in OOD extrapolation and overall remediation performance compared to relying on the single best surrogate.

It is interesting to note that the best optimization schemed is produced by the SVR surrogate of the multiple-surrogate optimization framework. It happened again that the best solution is not from the DNN, which was clearly the most accurate surrogate model in the overall validation test.

This result again illustrates the core practical value of the multi-surrogate approach. By incorporating this diversity, a simple ensemble strategy reliably delivers better and more robust remediation outcomes than relying exclusively on the single top-performing surrogate.

4. Discussion

4.1. Advantages of using multiple surrogates

In the field of hydrogeological inversion optimization, surrogate models have been widely adopted to approximate complex numerical simulations, to reduce the high computational cost. However, the most popular practice in previous studies is to adopt a single-model strategy: researchers typically construct multiple surrogate models, then validate their predictive accuracy and finally select the single most accurate model for inversion optimization. This approach rests on the assumption that a model excelling in overall validation will also yield optimal results in all of the scenarios and in the optimization phase. However, it is not the case (Fig. 12 and Table 4). Under our tested conditions, no single model consistently outperforms others across all scenarios. Instead, each model has the possibility to generate more accurate results than the others in the specific cases.

Concerned that this finding might be biased or limited to the specific field parameters in the paper, we conducted tests in Appendix A to verify its generalizability. Specifically, we performed 8 additional tests using K fields with different random patterns and parameters, and we compared the performance of the five surrogate models. The results remain consistent: no single model consistently outperforms others. Thus, this finding is likely to hold true across other new P&T conditions. It may also extend to the other area that employ surrogate models.

Previous studies in hydrogeological optimization often relied on single surrogate models for pump-and-treat (P&T) system design (Luo and Lu, 2014; Majumder and Eldho, 2020; Matott et al., 2006; Qiang et al., 2024; Song et al., 2025; Zhang et al., 2022). In contrast, our multiple-surrogate framework independently couples models like DNN and PolyInterp with the optimization algorithm, comparing their outcomes to identify superior solutions. For instance, while applying the conventional single-model strategy would have led us to select the DNN model for its higher validation accuracy, while PolyInterp achieved the lowest residual (17.48%, Table 5). Thus, it is advantageous to employ multiple models to uncover better optimization outcomes (Fig. 12).

We would suggest a multiple surrogate simulation-optimization framework: rather than selecting a single model based on validation accuracy, we propose constructing multiple surrogate models, and conducting separate inversion optimizations for each. By aggregating the results and performing a comparative analysis, this strategy has the potential to yield a superior outcome. The success of PolyInterp in achieving the lowest residual (17.48%) is the valid support for this strategy.

We acknowledge that previous researchers (Ouyang et al., 2017; Xing et al., 2019) have also developed the ensemble surrogate for integrating the strengths of various surrogate models. The ensemble surrogate has been achieved by assigning dynamic weights to each surrogate model based on their validation performance.

Goel et al. (2007) tested a series of ensemble surrogate method with various weighting strategies against individual surrogates and optimization results. They outlined three effective global weighting approaches. First, the weights are assigned based on the errors for each surrogate divided by the total errors, this method provides only a modest advantage to better models to avoid over-reliance on any one. Second, the best surrogate per experimental setup is selected and is given full weight. Third, it represents a tunable method balancing individual model confidence with averaging to mitigate outliers. Result demonstrated that ensemble surrogate model by the third method exhibited higher robustness and accuracy. Importantly, they discovered the “best”

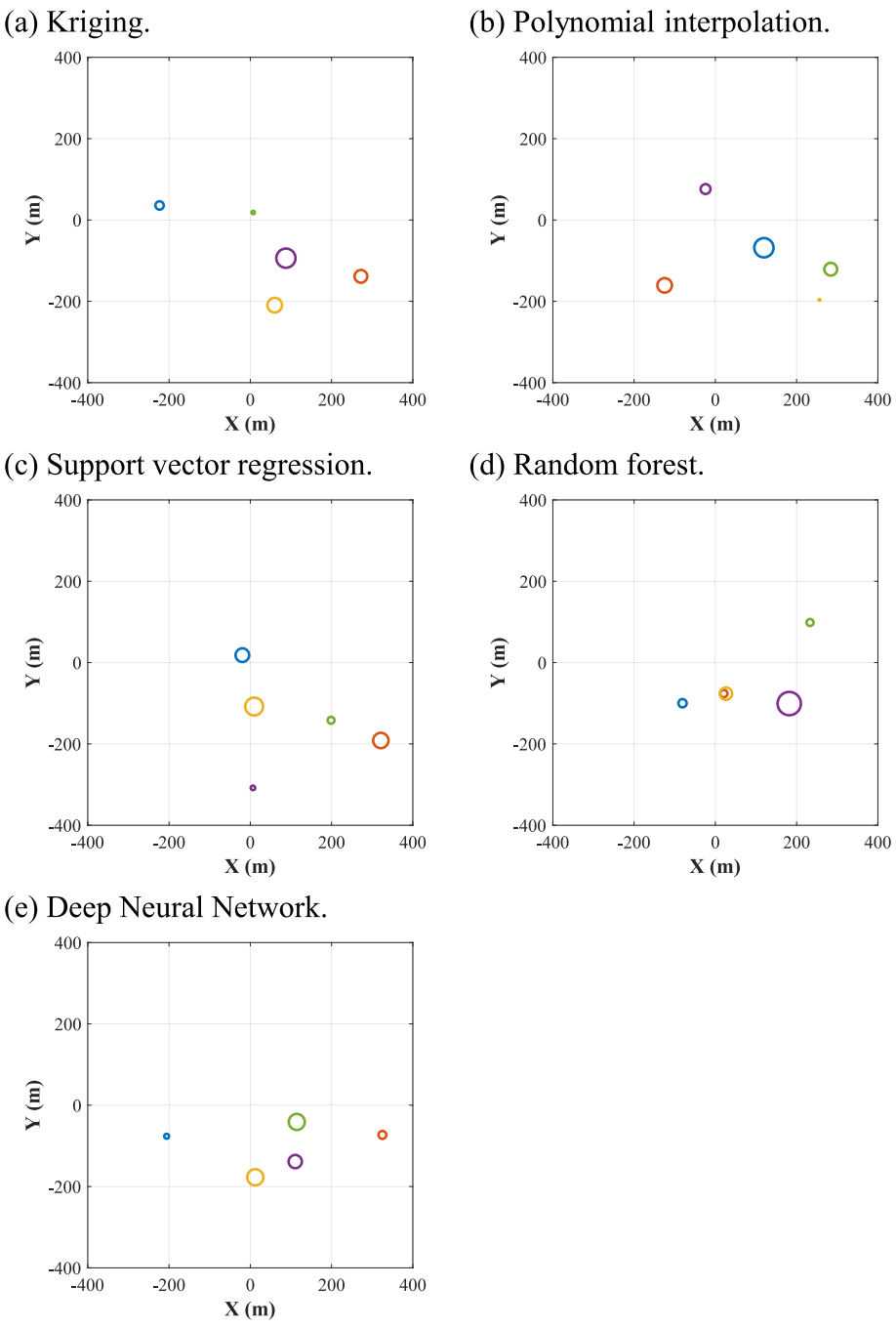


Fig. 13. Well locations and pumping rates inferred from optimization.

surrogate changed in over 40% of setups under different experiment design conditions. This inconsistency supports our method of doing independent optimizations for each surrogate to protect top performers that might get weakened in the "blending" process of weighting. (Christelis et al., 2019) also observed that the weighted ensemble surrogate model "did not consistently outperform single surrogates"; and using the ensemble surrogate for optimization has contributed limited improvement to the results obtained by using a single model. We are concerned that if we use the ensemble surrogate method, in scenarios where a single model excels, its predictive accuracy may be compromised by blending with the other inferior predictions. Thus, the multiple surrogate simulation optimization framework is favored in this paper. By coupling each surrogate independently with the optimization algorithm, unique solutions are retained rather than averaged. However, we recognize that both ensemble and multiple-model strategies represent meaningful innovations. Future work should systematically compare these two strategies across various application conditions.

Besides, some prior work have combined P&T with in-situ remediation to enhance contaminant treatment performance and optimize costs (Thornton et al., 2014). Our study focuses on P&T alone to assess the multiple-surrogate framework, but hybrid cost-saving approaches are a valuable future direction. Similarly, while some studies employed injection wells in the P&T site. The injection wells could stabilize groundwater level and enhance contaminant transport in the remediation process (Chang et al., 2007). Despite that we didn't employ injection wells, this omission does not affect our core findings, and injection wells could be explored in future work.

Despite its advantages, the multiple-surrogate framework requires

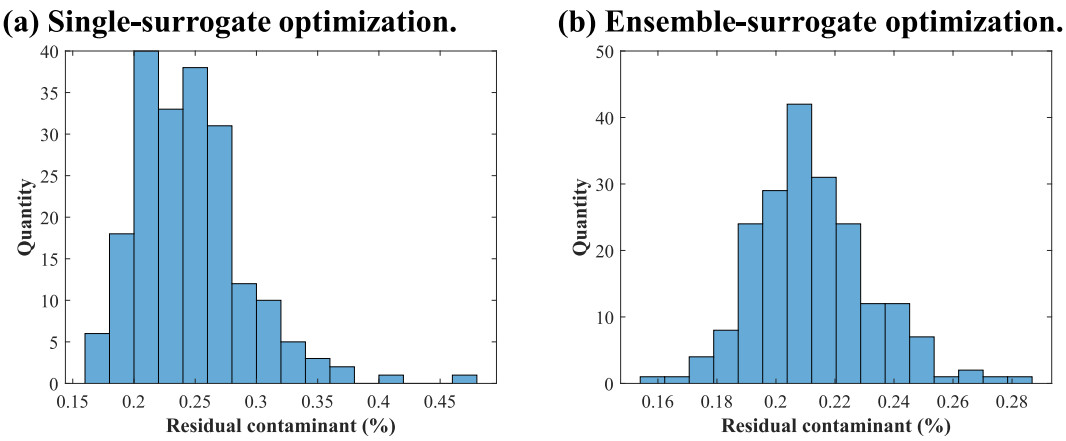


Fig. 14. Histograms of verified residual contaminant mass from multiple independent GA optimization runs.

Table 5
Optimized P&T parameters and $R_{residual}$ from five surrogate Models.

Surrogate models	Wells	X (m)	Y (m)	Pumping Rate (1000 m^3 / day)	$R_{residual}$ (%)
Kriging	1	−223.28	35.68	1.09	
	2	272.29	−138.59	1.77	
	3	60.08	−209.51	2.06	21.69
	4	87.70	−93.92	2.77	
	5	7.16	18.36	0.31	
PolyInterp	1	119.53	−68.27	2.78	
	2	−124.81	−160.80	2.05	
	3	256.26	−196.49	0.08	17.48
	4	−24.13	76.05	1.32	
	5	284.09	−121.05	1.78	
SVR	1	−19.49	18.28	1.91	
	2	321.30	−191.59	2.17	
	3	9.53	−108.02	2.56	19.31
	4	6.53	−308.15	0.49	
	5	198.95	−141.89	0.87	
RF	1	−80.82	−99.97	1.08	
	2	20.82	−76.25	0.93	
	3	26.07	−76.40	1.70	19.41
	4	181.89	−100.81	3.40	
	5	233.01	98.61	0.89	
DNN	1	−205.90	−76.39	0.55	
	2	325.18	−73.13	1.01	
	3	12.32	−177.28	2.31	19.24
	4	110.65	−138.63	1.85	
	5	114.46	−41.12	2.28	

Table 6
Median, mean, and best verified residual contaminant mass (%).

Method	Median	Best	Mean
Multi-surrogate	21.03	15.97	21.25
Single-surrogate DNN	24.05	16.91	24.63

more computational cost and human effort. Training and optimizing five surrogate models significantly increases computational demands compared to a single-model approach, as each model requires separate training. This also demands substantial human effort for model setup, tuning, and validation. However, given the severity of groundwater contamination and the high costs of P&T remediation, the improved accuracy from identifying superior solutions may justify these efforts.

To generalize our multiple-surrogate framework to other contamination sites with varying hydrogeological conditions, several data requirements must be met. Detailed aquifer information (structure, thickness, boundaries, porosity, heterogeneous permeability) needs site-specific characterization via borehole data or geophysical surveys. Accurate mapping of the contaminant field's spatial distribution is also essential to model plume dynamics. The P&T setup conditions, including well number, placement, and pumping rates, must be tailored to site realities.

Applying the multiple-surrogate framework to real-world scenarios requires addressing practical challenges beyond controlled simulations, including (1) limited data availability, as obtaining comprehensive aquifer and contaminant data is costly and site-specific; (2) regulatory and logistical constraints, such as the pumping implementation in the wells. These challenges may demand increased field efforts but ensure effective remediation.

4.2. Limitations and perspectives

This study primarily focuses on the development and application of multiple surrogate models, alongside the design of inversion strategies. It is a numerical study. If it is expected for real-world P&T applications, implementing the P&T remediation (using or not using the multiple-surrogate method) requires firstly, characterizing aquifer heterogeneity (via pumping tests or borehole data); and the exact contamination concentration distribution by methods like tracer tests or geophysical mapping. These are the important basics for P&T system design.

There are several limitations. First, we didn't consider the injection wells that reinject the treated-water into the groundwater. Although it is not considered in this work, it is entirely feasible to incorporate injection strategies, and we plan to explore this in future research. Second, the optimization assumes fixed pumping rates. The advantage of dynamic adjustments according to real-time conditions are not considered. We can implement dynamic adjustments of pumping rates to further enhance remediation effectiveness like (Wang and Zheng, 1997). Last but not least, no constraints have been incorporated into the optimization process. For example, it requires to implement a maximum allowable hydraulic drawdown to prevent aquifer overexploitation. But these should not avoid our findings related to multiple-surrogate framework.

In this study, the groundwater flow field induced by pump-and-treat operations was approximated as steady-state to improve computational efficiency. This approximation does not fully account for transient drawdown propagation and the evolving hydraulic gradients during the early pumping phase. Consequently, this could lead to modestly faster predicted plume migration and earlier capture of contaminant mass compared to fully transient simulations. If the proposed method will be applied to realistic field-scale problems, it is necessary to adopt fully transient flow simulations as the standard approach.

A notable limitation of the present study is the linear min-max rescaling applied to the hydraulic conductivity fields. After generation by a Gaussian model, each field was rescaled to confine values strictly within the range [−6, −4] m/s. This setup was implemented to force the logK values to desired ranges. However, this may reduce the occurrence

and intensity of extreme high- and low-permeability zones. Consequently, the rescaled fields may lead to somewhat weaker tailing effects in the simulated remediation process. In future work, we plan to avoid such artificial bounding by accepting the natural range of each realization, thereby more fully representing the inherent variability of randomly generated fields.

Future research aims to develop an adaptive method to address the challenge of insufficient field data in real-world groundwater remediation. Given that field data is often limited, we propose using data collected during the P&T process to iteratively characterize aquifer properties (e.g., heterogeneous permeability) and contaminant distributions (e.g., plume dynamics). This approach starts with a preliminary remediation scheme based on insufficient data and refines it as new data is gathered, optimizing well placement and pumping rates dynamically. Integrating this method with real-time monitoring systems, such as sensors for contaminant concentrations, can further enhance adaptive P&T strategies. This make our method more effective for practical applications.

In term of the surrogate models, we recognize the potential for integrating advanced lumped or effective upscaling models to enhance predictive capabilities. For instance, the multirate mass-transfer (MRMT) model effectively upscales anomalous solute transport in heterogeneous media under radial convergent flow by linking apparent capacity coefficients to aquifer anisotropy and connectivity (Pedretti et al., 2014). Thus, we plan to develop novel surrogate method that integrate with MRMT-like formulations to further improve the optimization framework.

This study employs a 2D synthetic aquifer to simulate horizontal flow and contaminant transport, providing a simplified system to efficiently evaluate the multiple-surrogate optimization framework. However, real-world aquifers are inherently three-dimensional, exhibiting vertical variations in hydraulic properties and flow dynamics. Transitioning to a 3D simulation model would likely introduce key differences in results across three main aspects: (1) increased difficulty in characterizing aquifer parameters and contaminant distributions, potentially leading to inaccuracies if data is limited; (2) the residual contaminant masses should be lower because as 3D models capture layered permeability or preferential flow paths that may cause some contaminants to remain unaffected; (3) the effect of vertical boundary conditions should be considered because the upcoming or leakage from overlying/underlying layers can significantly influence the fate of contaminant. While our 2D model effectively demonstrates the multiple-surrogate framework's advantages, adopting 3D modeling in future work could enhance realism for complex aquifer systems. This study is carried out on the idealized assumption that aquifer parameters and contaminant fields are perfectly known. In the future, we will develop probabilistic approaches to address the uncertainties of these information.

5. Conclusions

This study aimed to develop improved surrogate modeling approaches to optimize Pump-and-Treatment (P&T) system design for groundwater remediation. The parameters to be designed include the spatial coordinates of extraction wells and pumping rates for each well. The ultimate goal is to support effective environmental contamination management. We systematically assessed five state-of-the-art surrogate models – Kriging, Polynomial Interpolation (PolyInterp), Support Vector Regression (SVR), Random Forest (RF), and Deep Neural Network (DNN) - to determine their effectiveness in designing P&T schemes. The results demonstrate that the DNN model demonstrated statistically superior predictive accuracy. This reflects that the deep learning neuro network model is advantageous at learning nonlinear relationships in contaminant transport modeling.

While, our most striking finding is that no single model consistently outperforms others across all scenarios. Although DNN achieved the highest predictive accuracy, it only generated the best prediction results in 46 out of 200 cases. Conversely, some models show lower overall validation accuracy, they still produced optimal solutions in many cases (for example, Random Forest model has the lowest validation accuracy, but it generated 47 best predictions out of 200 validations). These findings demonstrate that we cannot identify a universally “best” surrogate model.

The above findings suggest that employing a multiple surrogate simulation-optimization framework could be beneficial. In this framework, each surrogate model (Kriging, PolyInterp, SVR, RF, and DNN) is independently trained and employed for optimization. Our experimental tests confirmed that this framework has contributed to better results: a P&T scheme that achieved higher contaminant remediation efficiency and lower final residual contaminant (17. 5% compared to 19.2–21.7%). It is also interesting to note that, this P&T scheme suggested that we can remove one of the 5 pumping wells. It could significantly reduce the implement cost for the P&T scheme while didn't harm the remediation results.

This study highlights that it is beneficial to move beyond conventional approach that relying on single “best-performing” surrogate model in simulation-optimization studies. While advanced models like DNN exhibited the highest predictive accuracy, their optimization performance does not necessarily arrive at the superior remediation solution. Instead, we can use a multiple-surrogate framework to let the advantages of all different surrogate models be fully used. Using the multiple-surrogate framework has been proved at least in this work to be more effective in identifying effective and cost-efficient P&T schemes.

CRediT authorship contribution statement

Chaoqi Wang: Writing – original draft, Software, Data curation. **Zhi Dou:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Ning Chen:** Methodology, Data curation. **Yan Zhu:** Validation, Investigation. **Zhihan Zou:** Formal analysis, Data curation, Conceptualization. **Jian Song:** Software, Methodology. **Shen-Huan Lyu:** Writing – review & editing, Supervision, Funding acquisition.

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.

Data availability

Data will be made available on request.

Acknowledgments

Zhi Dou reports financial support was provided by the National Natural Science Foundation of China (Grant No. W2511045 and 42272278) and the Natural Science Founda-tion of Jiangsu Province (Grant No. BK20240190). Chaoqi Wang reports financial support was provided by the National Natural Science Foundation of China (No. 42302295). Shen-Huan Lyu report financial support was provided by the National Natural Science Foundation of China (No. 62306104), Hong Kong Scholars Program (No. XJ2024010), Research Grants Council of the Hong Kong Special Administrative Region, China (GRF Project No. CityU11212524), Natural Science Foundation of Jiangsu Province (No. BK20230949). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Appendix

To assess the robustness of surrogate model performance against realization-specific biases and parameter variations, we generated 8 additional independent K fields. As shown in Fig. A1, the first 4 K fields are generated with the same parameters (correlation length 100 m; logK bounds −6 to −4 m/s), but each has a distinct random pattern. The K fields 5–8 have various parameters: 5 and 6 use larger correlation lengths of 200 m; 7 and 8 use smaller correlation lengths of 80 m. While K field 5 and K field 7 kept the logK bounds of −6 to −4 m/s, while 6 and 8 are assigned with larger bounds of logK: −6.5 to −3.5 m/s

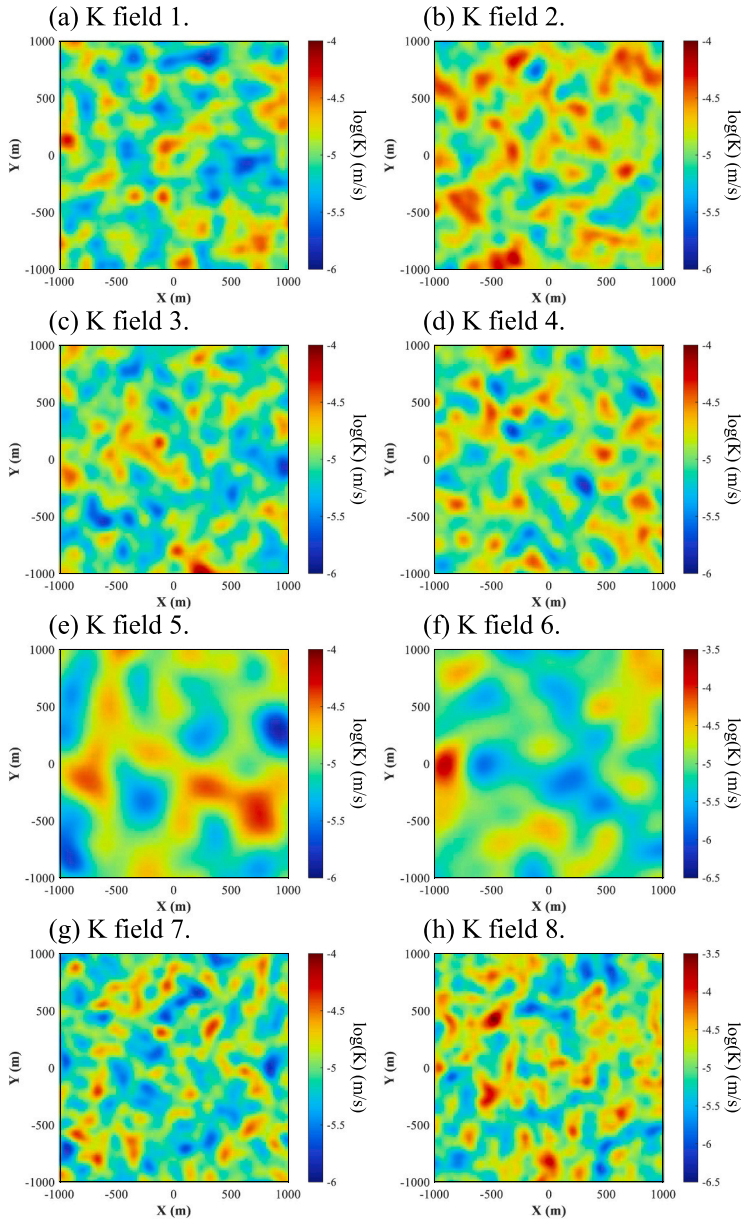


Fig. A1. The 8 new K fields.

For each K field, we used the same setup with the manuscript to build the surrogate models. Specifically, we first conducted 1200 numerical simulations of the pollutant extraction process, with randomly generated parameters (positions and pumping rates of the 5 wells) and the final residual contaminant mass is recorded. Second, the training dataset was formed by combining these random parameters and residual masses (1000 samples for training, 200 for validation). Third, all 5 surrogate models (Kriging, PolyInterp, SVR, RF, DNN) were built and trained for each field. Performances were evaluated on the validation set using RMSE (Root Mean Square Error) and the results are provided in Table A1.

Table A1
RMSE values of the supplement validation tests (The model with lowest error in each test is marked by underline).

K fields	Kriging	PolyInterp	SVR	RF	DNN
1	0.1923	0.2018	0.1882	0.2333	0.1889
2	0.1677	0.1882	0.1662	0.2182	0.1612
3	0.1574	0.1760	0.1577	0.2167	0.1539
4	0.1789	0.1947	0.1717	0.2337	0.1676
5	0.1878	0.1964	0.1793	0.2321	0.1768
6	0.1977	0.2051	0.1961	0.2346	0.1943
7	0.2051	0.2170	0.1974	0.2425	0.1981
8	0.1757	0.1952	0.1690	0.2305	0.1715

Since lower RMSE indicates better predictive accuracy, we can find that the SVR and DNN models are most accurate in all of the tests. The Random Forest model (RF) is less accurate in this overall validation stage.

We also counted how many times each surrogate model gave the lowest RMSE in the 200 validation samples per field. The results are shown in Table A2. Model RF exhibited a higher frequency for giving superior performance: usually more than 47 out of 200. The quantities are different. However, no model recorded zero instances. Despite that model Kriging exhibited much less top-performing instances, it still yielded the most accurate predictions in 10 to 40 cases per test.

Table A2
Top-performing instances of five surrogates in the supplement validation tests.

K fields	Kriging	PolyInterp	SVR	RF	DNN
1	15	38	48	58	41
2	33	37	40	58	32
3	21	44	48	52	35
4	23	38	50	47	42
5	22	39	46	52	41
6	22	38	48	54	38
7	26	35	45	56	38
8	23	33	51	47	46

These results can support “no single model consistently outperforms others across all scenarios.”. Although we did not extend to more fields, we still estimate that specific counts would shift with various K fields, but the pattern should persist: no model is always the most accurate, and no model is always the least accurate. These results should be supportive to our conclusion that no single surrogate model consistently outperforms the others across all scenarios.

References

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51, 5957–5973.

Bae, M.S., Kim, J.-H., Lee, S., 2024. Hydraulic containment of TCE contaminated groundwater using pulsed pump-and-treat: performance evaluation and vapor intrusion risk assessment. Environ. Pollut. 347, 123683.

Baskaran, P., Abraham, M., 2022. Evaluation of groundwater quality and heavy metal pollution index of the industrial area, Chennai. Phys. Chem. Earth, Parts A/B/C 128, 103259.

Bennett, K.P., Parrado-Hernández, E., 2006. The interplay of optimization and machine learning research. J. Mach. Learn. Res. 7, 1265–1281.

Boyd, S., 2004. Convex Optimization. Cambridge UP.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundat. Trends Mach. Learn. 3, 1–122.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burden, R.L., Faires, J.D., 2011. Interpolation & Polynomial Approximation Cubic Spline Interpolation I. Numerical Anal. 144–163.

Carroll, K.C., Brusseau, M.L., Tick, G.R., Soltanian, M.R., 2024. Rethinking pump-and-treat remediation as maximizing contaminated groundwater. Sci. Total 918, 170600.

Chang, L.-C., Chu, H.-J., Hsiao, C.-T., 2007. Optimal planning of a dynamic pump-treat-inject groundwater remediation system. J. Hydrol. (Amst) 342, 295–304.

Chen, Y., Song, L., Liu, Y., Yang, L., Li, D., 2020. A review of the artificial neural network models for water quality prediction. Appl. Sci. 10, 5776.

Christelis, V., Kopsiaftis, G., Mantoglou, A., 2019. Performance comparison of multiple and single surrogate models for pumping optimization of coastal aquifers. Hydrol. Sci. J. 64, 336–349.

Ciampi, P., Esposito, C., Bartsch, E., Alesi, E.J., Papini, M.P., 2023. Pump-and-treat (P&T) vs groundwater circulation wells (GCW): which approach delivers more sustainable and effective groundwater remediation? Environ. Res. 234, 116538.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Dawson, C.W., Wilby, R.L., 1999. A comparison of artificial neural networks used for river forecasting. Hydrol. Earth Syst. Sci. 3, 529–540.

Deng, D., Wang, Y., Zhong, Z., Wang, X., Yao, Y., 2025. A deep-learning-based proxy model for fast prediction of temperature during CO2 circulation in hydrothermal reservoir. Appl. Therm. Eng. 273, 126473.

Deng, Y., Kang, X., Ma, H., Qian, J., Ma, L., Luo, Q., 2024. Characterization of discrete fracture networks with deep-learning based hydrogeophysical inversion. J. Hydrol. (Amst). 631, 130819.

Elshall, A.S., Ye, M., Finkel, M., 2020. Evaluating two multi-model simulation–optimization approaches for managing groundwater contaminant plumes. J. Hydrol. (Amst) 590, 125427.

Fan, G., Zhang, D., Zhang, J., Li, Z., Sang, W., Zhao, L., Xu, M., 2022. Ecological environmental effects of Yellow River irrigation revealed by isotope and ion hydrochemistry in the Yinchuan plain, Northwest China. Ecol. Indic. 135, 108574.

Forrester, A.I.J., Keane, A.J., 2009. Recent advances in surrogate-based optimization. Prog. Aerosp. Sci. 45, 50–79.

Garcet, J.D.P., Ordonez, A., Roosen, J., Vanclooster, M., 2006. Metamodelling: theory, concepts and application to nitrate leaching modelling. Ecol. Model. 193, 629–644.

Giselle Fernández-Godino, M., Park, C., Kim, N.H., Haftka, R.T., 2019. Issues in deciding whether to use multifidelity surrogates. AIAA J. 57, 2039–2054.

Goel, T., Haftka, R.T., Shyy, W., Queipo, N.V., 2007. Ensemble of surrogates. Struct. Multidisciplinary Optimization 33, 199–216.

Hamutoko, J.T., Wanke, H., Voigt, H.J., 2016. Estimation of groundwater vulnerability to pollution based on DRASTIC in the Niipele sub-basin of the Cuvelai Etosha Basin, Namibia. Phys. Chem. Earth, Parts A/B/C 93, 46–54.

Harvey, C.F., Haggerty, R., Gorelick, S.M., 1994. Aquifer remediation: a method for estimating mass transfer rate coefficients and an evaluation of pulsed pumping. Water Resour. Res. 30, 1979–1991.

He, B., He, J., Wang, L., Zhang, X., Bi, E., 2019. Effect of hydrogeological conditions and surface loads on shallow groundwater nitrate pollution in the Shaying River basin: based on least squares surface fitting model. Water Res. 163, 114880.

He, Q., Li, P., Wang, Y., He, X., Fida, M., Elumalai, V., 2024. Hydrochemical characteristics of groundwater and their controlling mechanisms in irrigation and non-irrigation areas-a comparative study in the Guanzhong plain of China. Phys. Chem. Earth, Parts A/B/C 136, 103781.

He, X., Li, P., Wu, J., Wei, M., Ren, X., Wang, D., 2021. Poor groundwater quality and high potential health risks in the Datong Basin, northern China: research from published data. Environ. Geochem. Health 43, 791–812.

Hou, Z., Lu, W., Chen, M., 2016. Surrogate-based sensitivity analysis and uncertainty analysis for DNAPL-contaminated aquifer remediation. J. Water Resour. Plan. Manag. 142, 4016043.
Huang, C., Mayer, A.S., 1997. Pump-and-treat optimization using well locations and pumping rates as decision variables. Water Resour. Res. 33, 1001–1012.
Kuo, C.-H., Michel, A.N., Gray, W.G., 1992. Design of optimal pump-and-treat strategies for contaminated groundwater remediation using the simulated annealing algorithm. Adv. Water Resour. 15, 95–105.
Lee, M., Jung, Y., Hwang, C., Kim, Minjik, Kim, Minwoo, Lee, U., Lee, I., 2024. An efficient multi-fidelity design optimization framework for a thermoelectric generator system. Energ. Conver. Manage. 315, 118788.
Lee, M., Jeong, M.G., Lee, J., Lee, B.J., Lee, I., 2025. Efficient and robust thermal battery design optimization leveraging physically similar data. Appl. Therm. Eng. 269, 126009.
Lee, M., Lee, J., Choi, J.-H., Kim, N.H., Lee, I., 2026. A novel adaptive quality-based multi-fidelity surrogate framework for multiple low-fidelity data sources. Adv. Eng. Inform. 69, 103973.
Li, F., Liu, Y., Nazir, N., Ayyamperumal, R., 2024. Evaluating the influencing factors of groundwater evolution in rapidly urbanizing areas using long-term evidence. Phys. Chem. Earth, Parts A/B/C 136, 103728.
Li, J., Lu, W., Luo, J., 2021. Groundwater contamination sources identification based on the long-short term memory network. J. Hydrol. (Amst) 601, 126670.
Liang, H., Li, P., Elumalai, V., Tian, Y., Kou, X., 2025. Hydrochemical characteristics, groundwater nitrate sources and potential health risks in a typical alluvial plain of Northwest China. Phys. Chem. Earth, Parts A/B/C 139, 103903.
Luo, J., Lu, W., 2014. Comparison of surrogate models with different methods in groundwater remediation process. J. Earth Syst. Sci. 123, 1579–1589.
Luo, J., Ma, X., Ji, Y., Li, X., Song, Z., Lu, W., 2023. Review of machine learning-based surrogate models of groundwater contaminant modeling. Environ. Res. 238, 117268.
Ly, S., Charles, C., Degré, A., 2013. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: a review. Biotechnologie (agronomie, société et environnement) 17, 2.
Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., et al., 2014. Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. Environ. Model. Software 62, 271–299.
Majumder, P., Eldho, T.I., 2020. Artificial neural network and grey wolf optimizer based surrogate simulation-optimization model for groundwater remediation. Water Resour. Manag. 34, 763–783.
Masocha, M., Dube, T., Owen, R., 2020. Using an expert-based model to develop a groundwater pollution vulnerability assessment framework for Zimbabwe. Phys. Chem. Earth, Parts A/B/C 115, 102826.
Matott, L.S., Rabideau, A.J., 2008. Calibration of complex subsurface reaction models using a surrogate-model approach. Adv. Water Resour. 31, 1697–1707.
Matott, L.S., Rabideau, A.J., Craig, J.R., 2006. Pump-and-treat optimization using analytic element method flow models. Adv. Water Resour. 29, 760–775.
Mo, S., Zabaras, N., Shi, X., Wu, J., 2019. Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. Water Resour. Res. 55, 3856–3881.
Müller, S., Schüler, L., Zech, A., Heße, F., 2022. GSTools v1. 3: a toolbox for geostatistical modelling in Python. Geosci. Model Dev. 15, 3161–3182.
Multiphysics, C., 1998. Introduction to Comsol Multiphysics®. COMSOL Multiphysics, Burlington, MA (accessed Feb 9, 32).
Ning, J., Li, P., He, X., Ren, X., Li, F., 2024. Impacts of land use changes on the spatiotemporal evolution of groundwater quality in the Yinchuan area, China, based on long-term monitoring data. Phys. Chem. Earth, Parts A/B/C 136, 103722.
Ouyang, Q., Lu, W., Miao, T., Deng, W., Jiang, C., Luo, J., 2017. Application of ensemble surrogates and adaptive sequential sampling to optimal groundwater remediation design at DNAPLs-contaminated sites. J. Contam. Hydrol. 207, 31–38.
Pedretti, D., Fernàndez-Garcia, D., Sanchez-Vila, X., Bolster, D., Benson, D.A., 2014. Apparent directional mass-transfer capacity coefficients in three-dimensional anisotropic heterogeneous aquifers under radial convergent transport. Water Resour. Res. 50, 1205–1224.
Pham, L.T., Luo, L., Finley, A.O., 2020. Evaluation of random Forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds. Hydrol. Earth Syst. Sci. Discuss. 2020, 1–33.
Qiang, J., Zhang, S., Liu, H., Zhu, X., Zhou, J., 2024. A construction strategy of kriging surrogate model based on Rosenblatt transformation of associated random variables and its application in groundwater remediation. J. Environ. Manage. 349, 119555.
Rabbani, O., Ali, W., Khattak, G.A., Muhammad, S., Nafees, M., Iqbal, S., Din, I.U., Ahmad, A., Farooq, U., 2025. Spatial distribution, potential health risks, and sources of groundwater contamination in the semi-arid region. Phys. Chem. Earth, Parts A/B/C. 139, 103952.
Rao, N.S., Sunitha, B., Das, R., Kumar, B.A., 2022. Monitoring the causes of pollution using groundwater quality and chemistry before and after the monsoon. Phys. Chem. Earth, Parts a/b/c 128, 103228.
Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48.
Robinson, T.D., Eldred, M.S., Willcox, K.E., Haimes, R., 2008. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. AIAA J. 46, 2814–2822.
Rudiyanto, B., Birri, M.S., Widjonarko, Avian C., Kamal, D.M., Hijriawan, M., 2023. A genetic algorithm approach for optimization of geothermal power plant production: case studies of direct steam cycle in Kamojang. S Afr. J. Chem. Eng. 45, 1–9.
Schoppa, L., Disse, M., Bachmair, S., 2020. Evaluating the performance of random forest for large-scale flood discharge simulation. J. Hydrol. (Amst) 590, 125531.
Secci, D., Molino, L., Zanini, A., 2022a. Contaminant source identification in groundwater by means of artificial neural network. J. Hydrol. (Amst) 611, 128003.
Secci, D., Molino, L., Zanini, A., 2022b. Contaminant source identification in groundwater by means of artificial neural network. J. Hydrol. (Amst) 611, 128003.
Singh, P., Verma, P., 2019. A comparative study of spatial interpolation technique (IDW and kriging) for determining groundwater quality. GIS Geostatistical Techniques Groundwater Sci. 43–56.
Singh, R.M., Datta, B., 2006. Identification of groundwater pollution sources using GA-based linked simulation optimization model. J. Hydrol. Eng. 11, 101–109.
Somogyvári, M., Jalali, M., Jimenez Parras, S., Bayer, P., 2017. Synthetic fracture network characterization with transdimensional inversion. Water Resour. Res. 53, 5104–5123.
Song, X., Demirkanli, I., Hou, Z., Lin, X., Karanovic, M., Tonkin, M., Appriou, D., Mackley, R., 2025. Integrating analytical solutions and U-net model for predicting groundwater contaminant plumes in pump-and-treat systems. Adv. Water Resour. 105002.
Thornton, S.F., Baker, K.M., Bottrell, S.H., Rolfe, S.A., McNamee, P., Forrest, F., Duffield, P., Wilson, R.D., Fairburn, A.W., Cieslak, L.A., 2014. Enhancement of in situ biodegradation of organic compounds in groundwater by targeted pump and treat intervention. Appl. Geochem. 48, 28–40.
Truex, M., Johnson, C., Macbeth, T., Becker, D., Lynch, K., Giaudrone, D., Frantz, A., Lee, H., 2017. Performance assessment of pump-and-treat systems. Groundwater Monitor. Remediation 37, 28–44.
Truex, M.J., Johnson, C.D., 2017. Incorporating Pump-and-Treat Performance Assessment into Hanford Remedy Documents.
Truex, M.J., Johnson, C.D., Becker, D.J., Lee, M.H., Nimmons, M.J., 2015. Performance Assessment for Pump-and-Treat Closure or Transition.
Uugwanga, M.N., Kgabi, N.A., 2021. Heavy metal pollution index of surface and groundwater from around an abandoned mine site, Klein Aub. Phys. Chem. Earth, Parts a/b/c 124, 103067.
Vapnik, V., 2000. The Nature of Statistical Learning Theory. SpringerVerlag New York. Inc., New York, NY, USA.
Viana, F.A.C., Haftka, R.T., Steffen Jr., V., 2009. Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct. Multidiscip. Optim. 39, 439–457.
Villa-Vialaneix, N., Follador, M., Ratto, M., Leip, A., 2012. A comparison of eight metamodeling techniques for the simulation of N2O fluxes and N leaching from corn crops. Environ. Model. Software 34, 51–66.
Wang, C., Dou, Z., Zhu, Y., Yang, Z., Zou, Z., 2024a. Breaking the mold of simulation-optimization: direct forward machine learning methods for groundwater contaminant source identification. J. Hydrol. (Amst) 642, 131759.
Wang, J., Mao, Y., Chen, Y., Li, L., Qi, S., 2025. Enhanced toluene remediation in low-permeability zone by injecting rhamnolipid-coated ozone micro-nano bubble water combined with groundwater pumping. J. Hydrol. (Amst) 660, 133509.
Wang, M., Zheng, C., 1997. Optimal remediation policy selection under general conditions. Groundwater 35, 757–764.
Wang, Z., Le, T., Tian, K., van Phong, T., Bien, T.X., Pham, B.T., 2024b. Novel ensemble models based on the Split-point sampling and node attribute subsampling classifier for groundwater potential mapping. Earth and space. Science 11, e2023EA003338.
Xing, Z., Qu, R., Zhao, Y., Fu, Q., Ji, Y., Lu, W., 2019. Identifying the release history of a groundwater contaminant source based on an ensemble surrogate model. J. Hydrol. (Amst) 572, 501–516.
Xu, H., Zhang, H., Qin, C., Li, X., Xu, D., Zhao, Y., 2024. Groundwater Cr (VI) contamination and remediation: a review from 1999 to 2022. Chemosphere 142395.
Yoon, H., Hyun, Y., Lee, K.-K., 2007. Forecasting solute breakthrough curves through the unsaturated zone using artificial neural networks. J. Hydrol. (Amst) 335, 68–77.
Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol. (Amst) 396, 128–138.
Zaghiyan, M.R., Eslamian, S., Gohari, A., Ebrahimi, M.S., 2021. Temporal correction of irregular observed intervals of groundwater level series using interpolation techniques. Theor. Appl. Climatol. 145, 1027–1037.
Zha, Y., Yeh, T.-C.J., Illman, W.A., Mok, C.M.W., Tso, C.-H.M., Carrera, B.A., Wang, Y.-L., 2019. Exploitation of pump-and-treat remediation systems for characterization of hydraulic heterogeneity. J. Hydrol. (Amst) 573, 324–340.
Zhang, S., Qiang, J., Liu, H., Zhu, X., Lv, H., 2022. A construction strategy for conservative adaptive kriging surrogate model with application in the optimal design of contaminated groundwater extraction-treatment. Environ. Sci. Pollut. Res. 29, 42792–42808.
Zhang, Z., Ran, B., Gong, C., Yan, N., Yang, J., Shen, C., Wang, Y.-L., 2025. Enhancing groundwater remediation efficiency through integrating pump-and-treat system and groundwater circulation well. Process Saf. Environ. Prot. 194, 1454–1464.
Zheng, C., Wang, P.P., 1999. An integrated global and local optimization approach for remediation system design. Water Resour. Res. 35, 137–148.
Zheng, C., Wang, P.P., et al., 1999. MT3DMS: a Modular Three-dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide.
Zhi, W., Appling, A.P., Golden, H.E., Podgorski, J., Li, L., 2024. Deep learning for water quality. Nat. Water 1–14.

Zhou, Z., Tartakovsky, D.M., 2021. Markov chain Monte Carlo with neural network surrogates: application to contaminant source identification. Stoch. Env. Res. Risk A. 35, 639–651.
Zhou, Z., Roubinet, D., Tartakovsky, D.M., 2021. Thermal experiments for fractured rock characterization: theoretical analysis and inverse modeling. Water Resour. Res. 57, e2021WR030608.
Zhu, Q., Wen, Z., Zhan, H., Yuan, S., 2020. Optimization strategies for in situ groundwater remediation by a vertical circulation well based on particle-tracking and node-dependent finite difference methods. Water Resour. Res. 56, e2020WR027396.