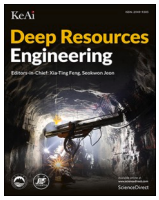
Deep Resources Engineering 1 (2024) 100009

Contents lists available at ScienceDirect

Deep Resources Engineering

journal homepage: www.keaipublishing.com/en/journals/deepre/



Review

Metaheuristic algorithms for groundwater model parameter inversion: Advances and prospects

Junjun Chen [a,b], Zhenxue Dai [b,*]

[a] National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou 221116, China
[b] College of Construction Engineering, Jilin University, Changchun 130026, China

ARTICLE INFO

Keywords:
Groundwater
Inverse modeling
Metaheuristic algorithms
Genetic algorithm
Particle swarm optimization
Simulated annealing
Differential evolution

ABSTRACT

Groundwater inverse modeling is a vital technique for estimating unmeasurable model parameters and enhancing numerical simulation accuracy. This paper comprehensively reviews the current advances and future prospects of metaheuristic algorithm-based groundwater model parameter inversion. Initially, the simulation-optimization parameter estimation framework is introduced, which involves the integration of simulation models with metaheuristic algorithms. The subsequent sections explore the fundamental principles of four widely employed metaheuristic algorithms—genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), and differential evolution (DE)—highlighting their recent applications in water resources research and related areas. Then, a solute transport model is designed to illustrate how to apply and evaluate these four optimization algorithms in addressing challenges related to model parameter inversion. Finally, three noteworthy directions are presented to address the common challenges among current studies, including balancing the diverse exploration and centralized exploitation within metaheuristic algorithms, local approximate error of the surrogate model, and the curse of dimensionality in spatial variational heterogeneous parameters. In summary, this review paper provides theoretical insights and practical guidance for further advancements in groundwater inverse modeling studies.

1. Introduction

Groundwater numerical simulation models are vital tools for subsurface system quantitative analyses. They reveal complex processes in subsurface systems across various spatial and temporal scales [1]. As an active component in the critical zone of earth, groundwater is widely distributed across the crust and lithosphere. An accurate representation of groundwater distribution contributes significantly to geoscience-, energy-, and environmental-related areas, including land subsidence[2,3], contaminant transports[4,5], landslides[6,7], nuclear waste disposal [8,9], and geological carbon storage[10,11].

A significant challenge in a specific numerical simulation process is estimating model parameters (e.g., hydraulic conductivity and release concentrations of solute) accurately[12,13]. However, the complexity of the geological formation in lithology results in heterogeneity and spatial variability in aquifers[14]. Thus, it is difficult to directly describe the model parameters for an entire aquifer using sparse and limited observational data[15,16]. Currently, inverse modeling stands as a crucial approach in estimating model parameters. It is a reverse solution procedure according to available aquifer system conditions and observed information (e.g., hydraulic heads and solute concentrations) [17–19]. The theories of a majority of inverse modeling approaches can be explained by the Bayesian theorem. In this theorem, the prior distribution represents the uncertainty of model parameters according to the initial beliefs or assumptions. In contrast, the posterior distribution is the conditional distribution of model parameters given additional observation data. It represents the updated beliefs about the parameters after incorporating the new data[20]. Thus, a feasible way to solve parameter inversion is to use model parameters with the maximum posterior probability as the best solution. To realize this purpose, the simulation-optimization framework is commonly used, where model parameters with maximum posterior probability are searched out by integrating simulation models and optimization algorithms[21–23]. The key stages of simulation-optimization framework implementation are optimization model establishment and solution. Specifically, when the constraints of the constructed optimization model satisfy convex optimization regularization conditions, gradient-based optimization algorithms can be used to obtain inversion results with high efficiency.

* Corresponding author.
E-mail address: dzx@jlu.edu.cn (Z. Dai).

https://doi.org/10.1016/j.deepre.2024.100009
Received 5 December 2023; Received in revised form 2 April 2024; Accepted 2 April 2024
Available online 7 April 2024

Representative algorithms include the Gaussian-Newton algorithm[24,25], Levenberg-Marquardt algorithm[26,27], and the conjugate gradient method[28,29] et al. Nevertheless, establishing convex optimization regularization constraints is commonly difficult due to the significant nonlinearity inherent in groundwater systems and limited available data. In such cases, the model parameters estimated by gradient-based optimization algorithms might be local optimums. Thus, the gradient-based optimization algorithms only perform well when the prior region or initial gauss is close to the true model parameter values [30,31]. Metaheuristic algorithms are another category of optimization algorithms inspired by natural laws and human intelligence. Compared to gradient-based optimization algorithms, metaheuristic algorithms are outstanding for their robustness and capability on global optimization. Metaheuristic algorithms can also be used for non-differentiable or discontinuous optimization problems where gradient-based algorithms may not be practical[23]. Owing to these advantages, metaheuristic algorithms have been used in various disciplines, such as energy power [32,33], civil engineering[34,35], agriculture[36,37], transportation [38,39], and earth science [40,41], etc.

This review paper focuses on the application of metaheuristic algorithms on groundwater model parameter inversion. The contents are organized as follows: Section 2 introduces the basic mathematical principles of the simulation-optimization inversion framework; Section 3 briefly presents the four most popular metaheuristic algorithms and some of their representative application cases; Section 4 designed a synthetic solute transport numerical model to illustrate the implementation of metaheuristic algorithm-based groundwater model parameter estimation; Section 5 gives out the current challenges and noteworthy research directions; Finally, the conclusion of this work are presented in Section 6.

2. Simulation-optimization inversion framework

For inversion studies, an optimization problem can be summarized as follows: investigating how to search for a set of model parameters $\boldsymbol{X}^*$, which minimizes or maximizes an objective function within certain constraints.

The mathematical model for solving optimization problems is called the inversion optimization model. It typically includes three elements: decision variables, an objective function, and constraints. The mathematical representation is generally formulated as follows[42] :

$$\min f(\mathrm{X}) \quad s.t. \begin{cases} g_i(\mathrm{X}) \le 0, (i = 1, 2, \ldots, m) \\ h_j(\mathrm{X}) = 0, (j = 1, 2, \ldots, t) \end{cases} \tag{1}$$

Where $\mathrm{X} = (x_1, x_2, \ldots, x_n)^T$ represents decision variables of the inversion optimization model;

n is the number of decision variables; $f(\mathrm{X})$ is the objective function; $g_i(\mathrm{X}) \le 0, i = 1, 2, \ldots, m$ represents inequality constraints; and $h_j(\mathrm{X}) = 0, j = 1, 2, \ldots, t$ are equality constraints. Specifically, if the objective function and constraints in Eq. (1) involve nonlinear functions, the optimization problem is referred to as a nonlinear programming optimization problem; conversely, it is called a linear programming optimization problem. Most of the groundwater inverse modeling studies are nonlinear programming optimization problems.

According to the basic theory of the above inversion optimization model, The detailed information for the three essential elements of decision variables, objective function, and constraints are introduced as follows:

Step 1 : Determine decision variables for the inversion optimization model.

Decision variables are parameters that decision-makers can modify to achieve desired model outputs. In groundwater numerical models, decision variables refer to model parameters to be estimated. Adjusting of these parameters influences the calibration fit results and prediction accuracy. Specifically, decision variables may be arbitrary model conditions, such as physical properties of aquifers, source-sink terms, and boundary conditions. In this paper, the decision variables of a numerical model with N_m model parameters are represented as a vector of $\boldsymbol{m} \in \mathbb{R}^{N_m \times 1}$.

Step 2: Constructing the objective function.

The objective function for model parameter inversion is typically derived using the generalized least squares objective function obtained through the Bayesian theorem. The detailed derivation process is as follows:

Assuming $\mathrm{y} \in \mathbb{R}^{N_{obs} \times 1}$ is the simulation results of model parameter $\boldsymbol{m}$; $\widetilde{\mathrm{y}}_{obs} \in \mathbb{R}^{N_{obs} \times 1}$ is the true observation data of model responses; $\boldsymbol{\varepsilon} \in \mathbb{R}^{N_{obs} \times 1}$ is the vector of observation noise. The mathematical relationships among these variables can be expressed as:

$$\begin{aligned} \widetilde{\mathrm{y}}_{obs} &= \mathrm{y} + \varepsilon \\ &= \mathrm{F}_{HF}(\mathrm{m}) + \varepsilon \end{aligned} \tag{2}$$

where $\boldsymbol{F}_{HF}(\cdot)$ is the operator of the high-fidelity groundwater numerical model.

According to the Bayesian theorem, the inversion problem is equivalent to an optimization problem of maximizing the posterior distribution probability function:

$$\mathrm{m} = \arg\max p(\mathrm{m}|\widetilde{\mathrm{y}}_{obs}) \tag{3}$$

$$p(\mathrm{m}|\widetilde{\mathrm{y}}_{obs}) = \frac{p(\mathrm{m})p(\widetilde{\mathrm{y}}_{obs}|\mathrm{m})}{\int p(\widetilde{\mathrm{y}}_{obs}|\mathrm{m})p(\mathrm{m})d\mathrm{m}} \propto p(\mathrm{m})\mathscr{L}(\mathrm{m}|\widetilde{\mathrm{y}}_{obs}) \tag{4}$$

where $p(\boldsymbol{m})$ is the prior distribution of model parameter $\boldsymbol{m}$; $\mathscr{L}(\mathrm{m}|\widetilde{\mathrm{y}}_{obs}) \equiv p(\widetilde{\mathrm{y}}_{obs}|\mathrm{m})$ is the likelihood function; $p(\widetilde{\mathrm{y}}_{obs}) = \int p(\widetilde{\mathrm{y}}_{obs}|\mathrm{m})p(\mathrm{m})d\mathrm{m}$ is the marginal likelihood (i.e., a constant value). The likelihood function is determined by the probability density of the observational noise $\boldsymbol{\varepsilon}$. When $\boldsymbol{\varepsilon}$ is assumed to follow a Gaussian distribution, the likelihood function can be expressed as:

$$\mathscr{L}(\mathrm{m}|\widetilde{\mathrm{y}}_{obs}) = \prod_{i=1}^{N_{obs}} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\widetilde{\mathrm{y}}_{obs}[i] - \mathrm{F}^i_{HF}(\mathrm{m})}{\sigma_i} \right)^2 \right] \tag{5}$$

where σ_i is the standard deviation of the ith observation; $\widetilde{\mathrm{y}}_{obs}[i]$ represents the ith element in the observation vector $\widetilde{\boldsymbol{y}}_{obs}$; $\mathrm{F}^i_{HF}(\mathrm{m})$ is the ith output value of the high-fidelity numerical model $\boldsymbol{F}_{HF}(\boldsymbol{m})$. Specifically, if the prior distribution of model parameter m is uniform in an interval form, the inverse problem is transferred to the L2-norm-based optimization problem:

$$\begin{aligned} m &= \arg\min \sum_{i=1}^{N_{obs}} \frac{1}{\sigma_i} \left[\widetilde{\mathrm{y}}_{obs}[i] - \mathrm{F}^i_{HF}(\mathrm{m}) \right]^2 \\ &= \arg\min \sum_{i=1}^{N_{obs}} \frac{1}{\sigma_i} [\widetilde{\mathrm{y}}_{obs}[i] - \mathrm{y}[i]]^2 \end{aligned} \tag{6}$$

The mathematical expression in Eq. (6) represents the objective function for the groundwater inversion optimization model, applicable to most groundwater inversion scenarios.

Step 3 : Determination of constraint conditions

The constraints for establishing an inversion optimization model include equality and inequality constraints.

Equality constraints refer to conservation relations among model input parameters and simulated outputs. Typically, these relations can be represented using the forward numerical model $\boldsymbol{F}_{HF}(\boldsymbol{m})$.

Inequality constraints refer to the upper and lower bounds of the prior range for model parameters:

$$\mathrm{m}^L \le \mathrm{m} \le \mathrm{m}^U \tag{7}$$

where, $\boldsymbol{m}^L$ and $\boldsymbol{m}^U$ represent vectors formed by the upper and lower

bounds of the value ranges for each dimension of the model parameters $\boldsymbol{m}$, respectively.

According to the above described decision variables, objective function, and constraints, a general mathematical expression of the inversion optimization model for groundwater model parameters is as follows:

$$\begin{cases} \min \sum_{i=1}^{N_{obs}} \frac{1}{\sigma_i} [\tilde{y}_{obs}[i] - y[i]]^2 \\ y = F_{Forward}(m, \theta_{Forward}) \\ m^L \le m \le m^U \end{cases} \tag{8}$$

3. Optimization algorithms

After establishing an inversion optimization model, the inversion results can be obtained through optimization algorithms according to the constraints and objective function. Metaheuristic algorithms are effective for solving parameter optimization problems because they can perform global searches. According to incomplete statistics, there are currently more than 500 types of metaheuristic algorithms. These algorithms can be broadly classified into two categories based on the differences in their inspiration mechanisms. One category of methods simulates biological processes, while the others are based on principles from physics. This section will introduce four representative metaheuristic algorithms and discuss their applications in water resources. These methods are genetic algorithm, particle swarm optimization, differential evolution, and simulated annealing. Among them, the simulated annealing algorithm is a physics-inspired method, while the other three are bio-inspired. According to the review by Rajwar et al. [43], they are also the four most cited methods on the end of 2022.

3.1. Genetic algorithm

The genetic algorithm (GA) was proposed by Holland John [44] and can be considered as one of the earliest and most widely applied metaheuristic algorithms. The design inspiration for the GA comes from the natural selection theory and genetic principles. The fundamental idea is to simulate the process of biological evolution, evolving individuals within a population over generations through genetic operations. This iterative evolution aims to produce individuals increasingly adapted to the environment, seeking the optimal solution. Decision variables are regarded as chromosomes when the GA is employed to address optimization problems. These chromosomes are updated according to the genetic principles of genes and undergo operations such as crossover and mutation. Subsequently, based on their objective function values, the best chromosomes are retained through survival of the fittest. After multiple iterations, the chromosome that satisfies the termination condition is considered as the result of the model parameter inversion.

Han et al. [45] proposed to combine the advection-dispersion equation of contaminants in groundwater with GA for contamination source identification, including the source location, release history, and release intensity. Tegegne and Kim [46] used genetic algorithms to develop and evaluate the reservoir operation rules by reflecting the uncertainty of reservoir inflows. Rajesh et al. [47] adopted genetic algorithms to identify unconfined aquifer parameters using pumping test data. Ushijima and Yeh [48] developed an optimal experimental design method by combining a genetic algorithm and a reduced order model.

3.2. Particle swarm optimization

Particle Swarm Optimization (PSO) was initially proposed by Eberhart and Kennedy [49]. Its inspiration primarily comes from the cooperative behavior of bird flocks during foraging. During a foraging process, if a bird is unaware of the exact location of food, it will fly towards the region around its nearest bird. In the PSO algorithm, the decision variables are abstracted as particles. The position of each particle represents a feasible solution to the problem, and the food corresponds to the objective function. During each iteration, particles move toward their recorded optimal positions, and the global optimal position will be obtained according to the entire swarm. Eventually, all particles converge around the optimal value of the objective function. Key PSO implementing steps are parameter initialization, objective function evaluation, velocity update, and position update.

Gill et al. [22] extended the PSO to deal with multi-objective-based parameter estimation problems in hydrology. Abdelaziz and Zambrano [50] used the PSO for inverse modeling of a coupled flow and transport groundwater model in a fractured gneiss aquifer. Qu and Lou [51] applied PSO in dealing with the optimal allocation model of regional water resources to maximize the comprehensive economic, social and environmental benefits of regional water consumption. Mozaffari et al. [52] adopted the PSO algorithm to optimize the parameters of support vector regression to realize groundwater level prediction.

3.3. Differential Evolution

Differential Evolution (DE) was proposed by Storn and Price [53], building upon evolutionary concepts similar to genetic algorithms. The DE and GA methods have in common that they both require generating an initial population randomly, followed by the fitness values of each individual as the selection criterion. The main processes of DE also include three steps: mutation, crossover, and selection. The difference lies in their methods for population updating. In GA, the probability of parent crossover and the selection of offspring are determined after mutation based on their fitness values. In DE, a mutation vector is generated from the parent differential vector. This mutation vector and the parent individual vector are combined through crossover to create a new individual vector. Subsequently, this new vector is directly selected compared to its parent individual.

ElçI and Ayvaz [54] proposed a hybrid optimization approach for a regional-scale groundwater flow model, where the global optimization is realized by the DE algorithm. Wang et al. [55] coupled the DE algorithm with a Soil and Water Assessment Tool to realize daily automatic calibration for water environmental capacity. Chiu [56] applied the DE algorithm to identify the parameter structure in a groundwater model of the Pingtung Plain in Taiwan using groundwater level data. Gurarslan and Karahan [57] adopted the DE algorithm to solve problems of groundwater pollution source identification through parameter optimization processes.

3.4. Simulated annealing

The simulated annealing (SA) algorithm originated from the ideas proposed by Metropolis et al. [58]. After that, Kirkpatrick et al. [59] incorporated the annealing concept into combinatorial optimization. It is a stochastic optimization algorithm based on the Monte Carlo iterative strategy, drawing inspiration from the annealing process in solid-state physics and general combinatorial optimization problems. In each iteration of SA, a neighboring solution will be randomly generated according to the current one. Then, the acceptance of the new solution will be determined according to an acceptance probability. The uniqueness of the SA lies in its ability to accept solutions that are worse than the current one. This enables a jump-like parameter optimization across the entire parameter space, ensuring the global effectiveness of the final optimization results.

Day-Lewis et al. [60] used the SA algorithm to identify fracture-zone geometry according to hydraulic data. Jha and Datta [61] determined the characteristics of the groundwater contaminant source through an adaptive SA algorithm. Cunha and Marques [62] developed an SA algorithm-based multiobjective optimization framework for conducting water distribution network design. Tsai et al. [63] estimated the model

parameters of the transient storage model for hyporheic exchange modeling through the SA method.

4. Numerical experiment

A synthetic solute transport model is designed to demonstrate the introduced metaheuristic algorithms. Fig. 1 presents the flow domain of the designed solute transport model. The domain is a saturated aquifer with a size of 10 m×20 m. The left and right boundaries have constant hydraulic heads of 1.0 m and 0 m, respectively. The other two boundaries are both no-flow boundaries. The left boundary is also a contaminant source area, with a constant concentration of 1.0 mol/L. The permeability parameters of this model are divided into four zones K1∼K4, and their true values are 1.510×10^{-12} m^2, 4.035×10^{-12} m^2, 2.797×10^{-12} m^2, and 2.401×10^{-12} m^2, respectively.

The distribution of hydraulic heads and solute concentrations are simulated using TOUGHREACT, a verified and well-documented numerical simulation program for reactive non-isothermal flows of multiphase fluids in porous and fractured media[64,65]. Here, all these four permeability parameters are considered as unknown parameters to be estimated. The observation data are generated according to the numerical simulation results, including hydraulic heads at 24 observation well positions (black prisms) and solute concentrations of 15 time points. The 15 observed times are shown in Table 1. Thus, a total of 384 (24×15 +24) observation data are obtained. The prior information of the four identified parameters are all defined as 5.0×10^{-14} m^2∼5.0×10^{-12} m^2.

When employing metaheuristic algorithms to address model parameter inversion problems, repeated forward simulations can significantly reduce computational efficiency if the numerical simulation model is CPU intensive. Currently, most inversion modeling studies involve surrogate models to mitigate the high computational burden associated with repeated calls to the high-fidelity forward numerical model, thereby enhancing the efficiency of the inversion computations. A surrogate model is an approximation of the relationship between model inputs and outputs at a lower computational cost. One commonly used approach is known as the data-driven-based machine learning methods. In these methods, a training dataset composed of model input parameters and corresponding model response results are generated using the high-fidelity numerical model. Then, a surrogate model is constructed based on this dataset through supervised learning to realize nonlinear regression [66]. In this study, a deep residual convolutional neural network (ResNet) is utilized for building the surrogate model. ResNet, a deep convolutional neural network incorporating residual blocks, has been demonstrated in some previous studies to exhibit excellent robustness in characterizing complex nonlinear relationships

Table 1
Time setting for numerical simulation output.

Time serial	Output time (days)	Time serial	Output time (days)	Time serial	Output time (days)
1	2	6	12	11	22
2	4	7	14	12	24
3	6	8	16	13	26
4	8	9	18	14	28
5	10	10	20	15	30

[67]. To obtain a ResNet based surrogate model, a training dataset with 500 samples and another testing dataset with 100 samples are prepared. The predictive accuracy of ResNet on the testing dataset after model training is shown in Fig. 2. Specifically, the horizontal axis represents the high-fidelity numerical model outputs, while the vertical axis represents the surrogate model outputs. Points closer to the diagonal line indicate that the predicted results of the surrogate model closely match the predictions of the numerical model. According to Fig. 2, most of the pair-wise points are distributed along the diagonal line, indicating that the performance of this surrogate model satisfies the inversion procedure.

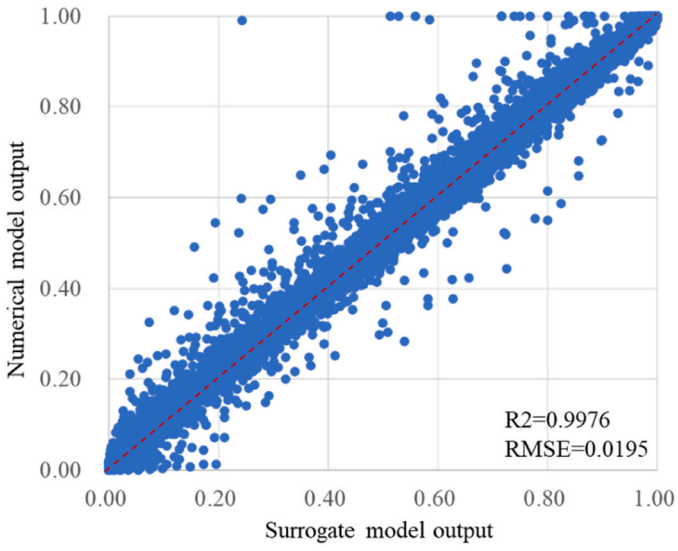


Fig. 2. Pair-wise comparison between surrogate model and numerical model output values.

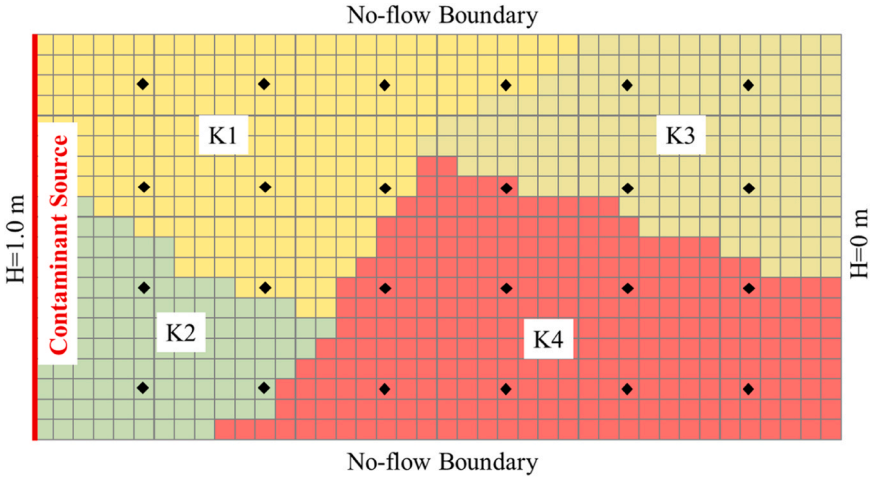


Fig. 1. Flow domain of the solute transport model.

The ResNet-based surrogate model is integrated into the simulation-optimization framework. The hyperparameters of four metaheuristic algorithms are set as shown in Table 2. Fig. 3 presents the convergence of the objective function during the optimization process for the four algorithms. In Fig. 3(d), the iterations displayed for SA range from 2 to 50. This is because the objective function value of the initial iteration is 0.487, a relatively significant value that influences the presentation of objective function values in subsequent iterations. Referring to Fig. 3(a) to (d), all of these four metaheuristic algorithms ultimately reach the inversion optimization constraints (i.e., their objective functions all approach zero) despite variations in convergence speed among the four metaheuristic algorithms.

The model parameter inversion results of these four metaheuristic algorithms are presented in Table 3. To assess their accuracy, the relative errors are calculated as shown in Table 4. According to Table 4, model parameter estimation results by these four metaheuristic algorithms are highly accurate, and all of their relative errors are below 0.05. Thus, the estimated parameter results can be considered near their global optimum. Additionally, the inversion results are assessed according to the convergence of the calibrated numerical model. The model parameter values estimated by these four metaheuristic algorithms are respectively input into the original numerical model, and their corresponding numerical model outputs are obtained. Then, the RMSE and R^2 values are calculated, as shown in Table 5. The RMSE values of the four methods are all less than 0.01, and the R^2 values are all beyond 0.999. Hence, the inversion results from the four metaheuristic algorithms closely match the predicted outcomes following the calibration of the numerical model, accurately reflecting the observed data.

The case study above briefly illustrates using these metaheuristic algorithms for groundwater inversion, which involves numerical model building, surrogate model construction, and parameter optimization. The accuracy of parameter estimation results and the convergence of calibrated numerical models demonstrate reliable inversion performances of these four metaheuristic algorithms.

5. Outlook

Despite the successful applications of metaheuristic algorithms, what should be acknowledged is that there is currently no single optimization algorithm capable of addressing all optimization problems. Therefore, most current research focuses on improving the original algorithms to address their shortcomings in specific applications. There are currently three main research directions worthy of attention, namely the improvement of metaheuristic algorithms, the improvement of surrogate models, and the application of parameter dimensionality reduction methods. An in-depth exploration of these directions is expected to advance the application and development of metaheuristic algorithms in practical scenarios.

5.1. Improvement of metaheuristic algorithms

The purpose of metaheuristic algorithms is to update an initial set of model parameters to satisfy specific constraints. The performance of a metaheuristic algorithm lies in balancing its relationship between diverse exploration and centralized exploitation. By employing diverse exploration, the algorithm is capable of identifying global optimal solutions throughout viable regions, preventing the optimization process from being trapped in local optima. However, the drawback is that this process is time-consuming and may result in lower precision. On the other hand, centralized exploitation involves an exact exploration of a specific parameter region, leading to a faster and more accurate determination of the optimal solution. The disadvantage of this process is the potential to get stuck in local optima. The essence of improving metaheuristic algorithms is to balance the diverse exploration and the centralized exploitation. Therefore, current efforts are directed towards improving the search space and convergence speed. For example, Li et al. [68] developed an improved initialization method named diagonal linear uniform initialization, which is able to adopt the diagonal subspace sampling instead of the whole space. Deng et al. [69] proposed a multi-strategy and hybrid algorithm to improve the whale optimization algorithm by integrating the chaotic mapping, the black widow algorithm, the opposition-based learning strategy, and the adaptive coefficients. Kazemzadeh-Parsi et al. [70] modified the Firefly algorithm in three ways: adding memory, preventing premature convergence to local optima, and proposing a newly updated formula. Then, it was used optimal remediation design of unconfined contaminated aquifers. Too and Abdullah [71] proposed an improved GA using a competition strategy and a dynamic mutation rate, enhancing the global search capability and the search behavior of mutation processes.

Table 2
hyperparameter setting for four metaheuristic algorithms.

GA		PSO	
Size of population	300	Size of population	300
Maximum iteration	100	Maximum iteration	100
Probability of mutation	0.003	inertia weight	0.8
		cognitive parameter	0.5
		social parameter	0.5
DE		**SA**	
Size population	100	Maximum temperature	1
Maximum iteration	100	Minimum temperature	1×10^{-9}
Probability of mutation	0.001	Long of chain	300
Coefficient of mutation	0.5	Cooldown time	300

5.2. Enhancement of surrogate models

The accuracy of inversion results obtained by a simulation-optimization framework depends largely on the local prediction accuracy of surrogate models around the true model parameter values. However, in some complex groundwater model problems, there may be strong nonlinearity or multi-processes involved in forward simulation. For instance, some study cases might include multi-component reactive solute transport models with robust nonlinear features[72]. Numerical simulation for these models typically requires a large number of computational iterations. Although surrogate models can help significantly reduce the computational cost of simulation-optimization processes, there are inevitably some approximate errors between surrogate models and numerical simulations [73]. Thus, enhancing the local prediction accuracy of surrogate models is inevitable in inversion studies. Firstly, it is generally acknowledged that increasing the number of training samples can enhance the prediction accuracy of surrogate models on the whole training dataset[17]. However, if the high-fidelity numerical simulation model is complex, generating more training samples requires extra high-fidelity numerical simulations, and the computational burden is still heavy. Therefore, a common approach is to improve local prediction accuracy[74]. One strategy is using a local adaptive update policy for surrogate models, gradually enhancing their prediction accuracy near true parameter values by adding local training samples iteratively[17,75,76]. Another strategy is to implement multi-fidelity surrogate models. Surrogate models are treated as low-fidelity models in this method. The local prediction accuracy is enhanced by establishing a nonlinear mapping that takes both low-fidelity model output and model parameters as inputs and outputs high-fidelity numerical model results[77–79].

5.3. Parameter dimensionality reduction

In some cases, it is necessary to recognize the heterogeneity and spatial variability of aquifers for accurately predicting contaminant transport. In such scenarios, the dimensionality of model parameters is determined by the number of discrete grids within the model, often extending to thousands or beyond. In high-dimensional inversion

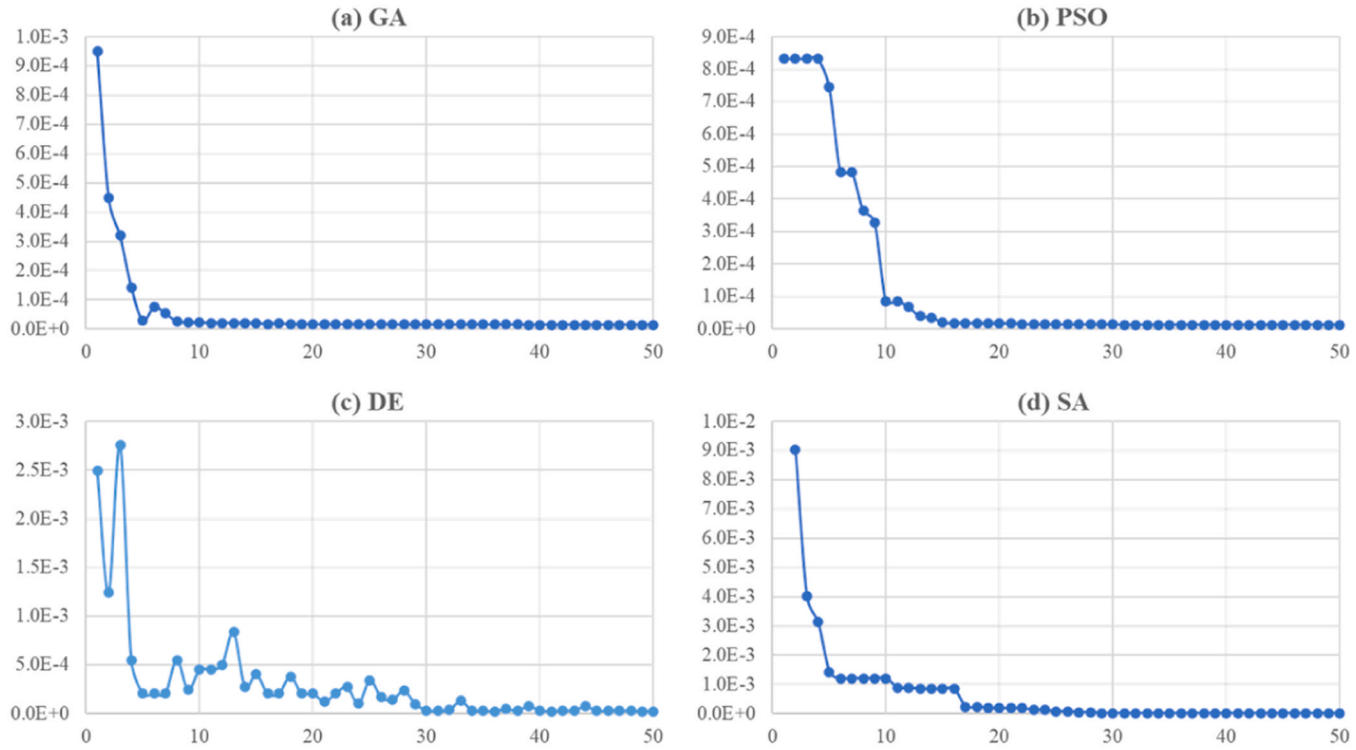


Fig. 3. The convergence process of four metaheuristic algorithms.

Table 3
Permeability estimated results by four metaheuristic algorithms.

	GA	PSO	DE	SA
K1(m^2)	1.524×10^{-12}	1.516×10^{-12}	1.510×10^{-12}	1.516×10^{-12}
K2(m^2)	3.908×10^{-12}	3.864×10^{-12}	3.865×10^{-12}	3.861×10^{-12}
K3(m^2)	2.732×10^{-12}	2.755×10^{-12}	2.761×10^{-12}	2.756×10^{-12}
K4(m^2)	2.343×10^{-12}	2.353×10^{-12}	2.362×10^{-12}	2.354×10^{-12}

Table 4
Relative error of parameter estimation results.

	GA	PSO	DE	SA
K1(m^2)	0.0094	0.0040	0.0005	0.0038
K2(m^2)	0.0313	0.0424	0.0422	0.0430
K3(m^2)	0.0233	0.0151	0.0131	0.0148
K4(m^2)	0.0242	0.0201	0.0165	0.0199

Table 5
RMSE and R2 values of calibrated numerical model outputs.

	GA	PSO	DE	SA
RMSE	0.00861	0.00856	0.00780	0.00856
R^2	0.99945	0.99946	0.99952	0.99946

problems, both the construction of surrogate models and the implementation of simulation-optimization frameworks suffer from the challenge known as the "curse of dimensionality" [80,81]. Currently, one common solution is to represent high-dimensional model parameters using low-dimensional vectors. Then, the high-dimensional inversion problem will be solved using conventional inversion methods by estimating low-dimensional vectors. If the model parameter follows a Gaussian random field, the Karhunen-Loève expansion (KLE) method can be used, because each of these high-dimensional model parameters can be effectively captured using a constrained number of KL terms[82]. For instance, in the work of Zhang et al. [83], a 3200-dimensional random field was represented using 200 KL terms, preserving approximately 98.8% of the total field variance. However, groundwater model parameters may not conform to Gaussian random fields in some scenarios where the Karhunen-Loève expansion (KLE) method is unsuitable. Another method is generative deep learning approaches for representing high-dimensional parameters with low-dimensional vectors. These methods are applicable to both Gaussian and non-Gaussian random fields. Represent methods include variational autoencoders (VAEs)[84–86], generative adversarial networks (GANs)[87–89], and other strategies integrating VAEs and GANs[90,91]. These approaches construct relationships between low-dimensional standard distributions (e.g., uniform distribution) and high-dimensional distributions through unsupervised learning. Then, high-dimensional heterogeneous parameters are represented by low-dimensional latent vectors (i.e., parameters after dimensionality reduction). For example, Chen et al. [92] represented a heterogenous permeability random field with 3200 dimensions using 10-dimensional vector through deep convolutional GAN. Zhan et al. [80] constructed a regression from 100 random vectors to heterogeneous aquifer structures with 24000 dimensions using a single-sample GAN method. Mo et al. [93] proposed a convolutional adversarial autoencoder method to realize dimensionality reduction and implemented in both 2D and 3D domain cases.

6. Conclusion

In recent decades, numerical models have been widely employed for dynamic simulation studies of subsurface systems. However, accurately estimate groundwater model input parameters is still a significant challenge. This paper provides a comprehensive review of metaheuristic algorithms-based inversion modeling frameworks. We emphasize four popular classical metaheuristic algorithms and their applications. A synthetic solute transport model is introduced to illustrate general processes for implementing these four metaheuristic algorithms. The inversion results of this case further demonstrate the feasibility of using metaheuristic algorithms to address groundwater model parameter

inversion problems. Finally, the outlook section presents some noteworthy research topics according to the drawbacks within the current simulation-optimization inversion framework. These topics include the improvements of heuristic algorithms, enhancements of local prediction accuracy by surrogate models, and the exploration of generative deep learning methods for the low-dimensional representation of high-dimensional parameters.

Declaration of Competing Interest

All authors declare that no conflict of interest exists.

Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities (XJ2023005201), the National Natural Science Foundation of China (NSFC: U2267217, 42141011, and 42002254).

References

[1] T. Xu, J. Samper, C. Ayora, et al., Modeling of non-isothermal multi-component reactive transport in field scale porous media flow systems, J. Hydrol. 214 (1-4) (1999) 144–164.
[2] L. Zhu, H. Gong, Y. Chen, et al., Effects of Water Diversion Project on groundwater system and land subsidence in Beijing, China, Eng. Geol. 276 (2020) 105763.
[3] M. Bagheri-Gavkosh, S.M. Hosseini, B. Ataie-Ashtiani, et al., Land subsidence: a global challenge, Sci. Total Environ. 778 (2021) 146193.
[4] B. Arora, S.S. Şengör, N.F. Spycher, et al., A reactive transport benchmark on heavy metal cycling in lake sediments, Comput. Geosci. 19 (3) (2014) 613–633.
[5] D.B. Das, V. Nassehi, Modeling of contaminants mobility in underground domains with multiple free/porous interfaces, Water Resour. Res. 39 (3) (2003).
[6] P. Marino, G.F. Santonastaso, X. Fan, et al., Prediction of shallow landslides in pyroclastic-covered slopes by coupled modeling of unsaturated and saturated groundwater flow, Landslides 18 (1) (2021) 31–41.
[7] Y. Xue, F. Miao, Y. Wu, et al., Combing soil spatial variation and weakening of the groundwater fluctuation zone for the probabilistic stability analysis of a riverside landslide in the Three Gorges Reservoir area, Landslides 20 (5) (2023) 1013–1029.
[8] L. De Windt, N.F. Spycher, Reactive transport modeling: a key performance assessment tool for the geologic disposal of nuclear waste, Elements 15 (2) (2019) 99–102.
[9] S. Jia, Z. Dai, Z. Yang, et al., Uncertainty quantification of radionuclide migration in fractured granite, J. Clean. Prod. 366 (2022).
[10] Z. Dai, L. Xu, T. Xiao, et al., Reactive chemical transport simulations of geologic carbon sequestration: methods and applications, Earth-Sci. Rev. 208 (2020) 103265.
[11] M.A. Celia, Geological storage of captured carbon dioxide as a large-scale carbon mitigation option, Water Resour. Res. 53 (5) (2017) 3527–3533.
[12] F.T.C. Tsai, N. Sun, W.W.G. Yeh, Global-local optimization for parameter structure identification in three-dimensional groundwater modeling, Water Resour. Res. 39 (2) (2003).
[13] W.W.G. Yeh, Review of parameter identification procedures in groundwater hydrology: the inverse problem, Water Resour. Res. 22 (2) (1986) 95–108.
[14] Z. Dai, R.W. Ritzi, C. Huang, et al., Transport in heterogeneous sediments with multimodal conductivity and hierarchical organization across scales, J. Hydrol. 294 (1-3) (2004) 68–86.
[15] M.R. Soltanian, R.W. Ritzi, C.C. Huang, et al., Relating reactive solute transport to hierarchical and multiscale sedimentary architecture in a Lagrangian-based transport model: 2, Part. Displac. Var. 51 (3) (2015) 1601–1618.
[16] M.R. Soltanian, R.W. Ritzi, C.C. Huang, et al., Relating reactive solute transport to hierarchical and multiscale sedimentary architecture in a Lagrangian-based transport model: 1, Time-Depend. Eff. Retard. Factor 51 (3) (2015) 1586–1600.
[17] J. Chen, Z. Dai, Z. Yang, et al., An improved tandem neural network architecture for inverse modeling of multicomponent reactive transport in porous media, Water Resour. Res. 57 (12) (2021).
[18] G. Kourakos, A. Mantoglou, Inverse groundwater modeling with emphasis on model parameterization, Water Resour. Res. 48 (5) (2012).
[19] C. Krapu, M. Borsuk, M. Kumar, Gradient-based inverse estimation for a rainfall-runoff model, Water Resour. Res. 55 (8) (2019) 6625–6639.
[20] J.A. Vrugt, Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation, Environ. Model. Softw. 75 (2016) 273–316.
[21] J. Sreekanth, B. Datta, Review: simulation-optimization models for the management and monitoring of coastal aquifers, Hydrogeol. J. 23 (6) (2015) 1155–1166.
[22] M.K. Gill, Y.H. Kaheil, A. Khalil, et al., Multiobjective particle swarm optimization for parameter estimation in hydrology, Water Resour. Res. 42 (7) (2006).
[23] O.B. Haddad, M.M.R. Tabari, E. Fallah-Mehdipour, et al., Groundwater model calibration by meta-heuristic algorithms, Water Resour. Manag. 27 (7) (2013) 2515–2529.
[24] J. Li, D. Elsworth, A Modified gauss-newton method for aquifer parameter identification, Groundwater 33 (4) (1995) 662–668.
[25] E. Laloy, N. Linde, C. Ruffino, et al., Gradient-based deterministic inversion of geophysical data with generative adversarial networks: Is it feasible? Comput. Geosci. (2019) 133.
[26] A. Sciortino, T.C. Harmon, W.W.G. Yeh, Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions, Water Resour. Res. 36 (7) (2000) 1723–1735.
[27] V.H. Nhu, Levenberg-Marquardt method for ill-posed inverse problems with possibly non-smooth forward mappings between Banach spaces, Inverse Probl. 38 (1) (2022).
[28] C.H. Huang, J.X. Li, S. Kim, An inverse problem in estimating the strength of contaminant source for groundwater systems, Appl. Math. Model. 32 (4) (2008) 417–431.
[29] M.C. Hill, Solving groundwater flow problems by conjugate-gradient methods and the strongly implicit procedure, Water Resour. Res. 26 (9) (1990) 1961–1969.
[30] Z. Dai, J. Samper, Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications, Water Resour. Res. 40 (7) (2004) W07407.
[31] Z. Dai, J. Samper, Inverse modeling of water flow and multicomponent reactive transport in coastal aquifer systems, J. Hydrol. 327 (2006) 447–461.
[32] A.M. Nassef, M.A. Abdelkareem, H.M. Maghrabie, et al., Review of metaheuristic optimization algorithms for power systems problems, Sustainability 15 (12) (2023).
[33] Q. Wang, G. Chen, M. Khishe, et al., Multi-objective optimization of IoT-based green building energy system using binary metaheuristic algorithms, J. Build. Eng. 68 (2023).
[34] R. Ma, M. Karimzadeh, A. Ghabussi, et al., Assessment of composite beam performance using GWO-ELM metaheuristic algorithm, Eng. Comput. 38 (SUPPL 3) (2022) 2083–2099.
[35] Y. Yalcin, M. Orhon, O. Pekcan, An automated approach for the design of mechanically stabilized earth walls incorporating metaheuristic optimization algorithms, Appl. Soft Comput. 74 (2019) 547–566.
[36] S. Sabzi, Y. Abbaspour-Gilandeh, G. Garcia-Mateos, A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms, Comput. Ind. 98 (2018) 80–89.
[37] H. Shayanfar, F.S. Gharehchopogh, Farmland fertility: A new metaheuristic algorithm for solving continuous optimization problems, Appl. Soft Comput. 71 (2018) 728–746.
[38] I. Abril Martinez-Salazar, J. Molina, F. Angel-Bello, et al., Solving a bi-objective Transportation Location Routing Problem by metaheuristic algorithms. Eur. J. Oper. Res. 234 (1) (2014) 25–36.
[39] G. Chaharmahali, D. Ghandalipour, M. Jasemi, et al., Modified metaheuristic algorithms to design a closed-loop supply chain network considering quantity discount and fixed-charge transportation. Expert Syst. Appl. 202 (2022).
[40] A. Lindsay, J. McCloskey, M.N. Bhloscaidh, Using a genetic algorithm to estimate the details of earthquake slip distributions from point surface displacements, J. Geophys. Res. -Solid Earth 121 (3) (2016) 1796–1820.
[41] S.D.R. Maurice, D.A. Wiens, K.D. Koper, et al., Crustal and upper mantle structure of southernmost South America inferred from regional waveform inversion, J. Geophys. Res. -Solid Earth 108 (B1) (2003).
[42] J. Luo, X. Li, Y. Xiong, et al., Groundwater pollution source identification using Metropolis-Hasting algorithm combined with Kalman filter algorithm, J. Hydrol. 626 (2023).
[43] K. Rajwar, K. Deep, S. Das, An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges, Artif. Intell. Rev. (2023) 1–71.
[44] H. Holland John, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
[45] K. Han, R. Zuo, P. Ni, et al., Application of a genetic algorithm to groundwater pollution source identification, J. Hydrol. 589 (2020).
[46] G. Tegegne, Y.-O. Kim, Representing inflow uncertainty for the development of monthly reservoir operations using genetic algorithms, J. Hydrol. 586 (2020).
[47] M. Rajesh, D. Kashyap, K.S. Hari Prasad, Estimation of unconfined aquifer parameters by genetic algorithms, Hydrol. Sci. J. 55 (3) (2010) 403–413.
[48] T.T. Ushijima, W.W.G. Yeh, Experimental design for estimating unknown groundwater pumping using genetic algorithm and reduced order model, Water Resour. Res. 49 (10) (2013) 6688–6699.
[49] R. Eberhart, J. Kennedy, Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Citeseer, F, 1995.
[50] R. Abdelaziz, M. Zambrano, Particle swarm optimization for inverse modeling of solute transport in fractured gneiss aquifer, J. Contam. Hydrol. 164 (2014) 285–298.
[51] G.-d Qu, Z.-h Lou, Application of particle swarm algorithm in the optimal allocation of regional water resources based on immune evolutionary algorithm, J. Shanghai Jiaotong Univ. (Sci.) 18 (5) (2013) 634–640.
[52] S. Mozaffari, S. Javadi, H.K. Moghaddam, et al., Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization, Water Resour. Manag. 36 (6) (2022) 1955–1972.
[53] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, J. Glob. Optim. 11 (4) (1997) 341–359.
[54] A. Elçi, M.T. Ayvaz, Differential-Evolution algorithm based optimization for the site selection of groundwater production wells with the consideration of the vulnerability concept, J. Hydrol. 511 (2014) 736–749.

[55] Q. Wang, R. Liu, C. Men, et al., Temporal-spatial analysis of water environmental capacity based on the couple of SWAT model and differential evolution algorithm, J. Hydrol. 569 (2019) 155–166.
[56] Y.-C. Chiu, Application of differential evolutionary optimization methodology for parameter structure identification in groundwater modeling, Hydrogeol. J. 22 (8) (2014) 1731–1748.
[57] G. Gurarslan, H. Karahan, Solving inverse problems of groundwater-pollution-source identification using a differential evolution algorithm, Hydrogeol. J. 23 (6) (2015) 1109–1119.
[58] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, et al., Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092.
[59] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.
[60] F.D. Day-Lewis, P.A. Hsieh, S.M. Gorelick, Identifying fracture-zone geometry using simulated annealing and hydraulic-connection data, Water Resour. Res. 36 (7) (2000) 1707–1721.
[61] M. Jha, B. Datta, Three-dimensional groundwater contamination source identification using adaptive simulated annealing, J. Hydrol. Eng. 18 (3) (2013) 307–317.
[62] M. Cunha, J. Marques, A new multiobjective simulated annealing algorithm—MOSA-GR: application to the optimal design of water distribution networks, Water Resour. Res. 56 (3) (2020).
[63] C.H. Tsai, D.F. Rucker, S.C. Brooks, et al., Transient storage model parameter optimization using the simulated annealing method, Water Resour. Res. 58 (7) (2022).
[64] T. Xu, E. Sonnenthal, N. Spycher, et al., TOUGHREACT—A simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: applications to geothermal injectivity and CO2 geological sequestration, Comput. Geosci. 32 (2) (2006) 145–165.
[65] T. Xu, N. Spycher, E. Sonnenthal, et al., TOUGHREACT Version 2.0: A simulator for subsurface reactive transport under non-isothermal multiphase flow conditions, Comput. Geosci. 37 (6) (2011) 763–774.
[66] M.J. Asher, B.F.W. Croke, A.J. Jakeman, et al., A review of surrogate models and their application to groundwater modeling, Water Resour. Res. 51 (8) (2015) 5957–5973.
[67] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. F. (2016).
[68] Q. Li, Y.G. Bai, W.F. Gao, Improved initialization method for metaheuristic algorithms: a novel search space view, IEEE Access 9 (2021) 121366–121384.
[69] H. Deng, L. Liu, J. Fang, et al., A novel improved whale optimization algorithm for optimization problems with multi-strategy and hybrid algorithm, Math. Comput. Simul. 205 (2023) 794–817.
[70] M.J. Kazemzadeh-Parsi, F. Daneshmand, M.A. Ahmadfard, et al., Optimal remediation design of unconfined contaminated aquifers based on the finite element method and a modified firefly algorithm, Water Resour. Manag. 29 (8) (2015) 2895–2912.
[71] J. Too, A.R. Abdullah, A new and fast rival genetic algorithm for feature selection, J. Supercomput. 77 (3) (2020) 2844–2874.
[72] E. Abarca, A. Idiart, F. Grandia, et al., 3D reactive transport modeling of porosity evolution in a carbonate reservoir through dolomitization, Chem. Geol. 513 (2019) 184–199.
[73] L. Yan, T. Zhou, Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems, J. Comput. Phys. 381 (2019) 110–128.
[74] J. Zhang, Q. Zheng, D. Chen, et al., Surrogate-based Bayesian inverse modeling of the hydrological system: an adaptive approach considering surrogate approximation error, Water Resour. Res. 56 (1) (2020).
[75] W. Gong, Q. Duan, J. Li, et al., Multi-objective parameter optimization of common land model using adaptive surrogate modeling, Hydrol. Earth Syst. Sci. 19 (5) (2015) 2409–2425.
[76] W. Gong, Q. Duan, J. Li, et al., Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models, Water Resour. Res. 52 (3) (2016) 1984–2008.
[77] S. Chakraborty, T. Chatterjee, R. Chowdhury, et al., A surrogate based multi-fidelity approach for robust design optimization, Appl. Math. Model. 47 (2017) 726–744.
[78] P. Conti, M. Guo, A. Manzoni, et al., Multi-fidelity surrogate modeling using long short-term memory networks, Comput. Methods Appl. Mech. Eng. 404 (2023).
[79] D. Liu, Y. Wang, Multi-fidelity physics-constrained neural network and its application in materials modeling, J. Mech. Des. 141 (12) (2019).
[80] C. Zhan, Z. Dai, J. Samper, et al., An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network, J. Hydrol. 610 (2022) 127844.
[81] Y. Zhao, Q. Guo, C. Lu, et al., High-dimensional groundwater flow inverse modeling by upscaled effective model on principal components, Water Resour. Res. 58 (7) (2022).
[82] D.X. Zhang, Z.M. Lu, An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loeve and polynomial expansions, J. Comput. Phys. 194 (2) (2004) 773–794.
[83] J. Zhang, W. Li, G. Lin, et al., Efficient evaluation of small failure probability in high-dimensional groundwater contaminant transport modeling via a two-stage Monte Carlo method, Water Resour. Res. 53 (3) (2017) 1948–1962.
[84] E. Laloy, R. Hérault, J. Lee, et al., Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, Adv. Water Resour. 110 (2017) 387–405.
[85] X.Y. Kang, A. Kokkinaki, P.K. Kitanidis, et al., Hydrogeophysical characterization of nonstationary DNAPL source zones by integrating a convolutional variational autoencoder and ensemble smoother, Water Resour. Res. 57 (2) (2021).
[86] S.W.A. Canchumuni, A.A. Emerick, M.A.C. Pacheco, Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother, Comput. Geosci. 128 (2019) 87–102.
[87] E. Laloy, R. Hérault, D. Jacques, et al., Training-image based geostatistical inversion using a spatial generative adversarial neural network, Water Resour. Res. 54 (1) (2018) 381–406.
[88] S. Chan, A.H. Elsheikh, Parametric generation of conditional geological realizations using generative neural networks, Comput. Geosci. 23 (5) (2019) 925–952.
[89] J.C. Bao, L.P. Li, F. Redoloza, Coupling ensemble smoother and deep learning with generative adversarial networks to deal with non-Gaussianity in flow and transport data assimilation, J. Hydrol. 590 (2020).
[90] S.X. Mo, N. Zabaras, X.Q. Shi, et al., Integration of adversarial autoencoders with residual dense convolutional networks for estimation of non-gaussian hydraulic conductivities, Water Resour. Res. 56 (2) (2020).
[91] M. Cheng, F. Fang, C.C. Pain, et al., An advanced hybrid deep adversarial autoencoder for parameterized nonlinear fluid flow modelling, Comput. Methods Appl. Mech. Eng. 372 (2020).
[92] J. Chen, Z. Dai, S. Dong, et al., Integration of deep learning and information theory for designing monitoring networks in heterogeneous aquifer systems, Water Resour. Res. 58 (10) (2022).
[93] S. Mo, N. Zabaras, X. Shi, et al., Integration of adversarial autoencoders with residual dense convolutional networks for estimation of non-Gaussian hydraulic conductivities, Water Resour. Res. 56 (2) (2020).

Zhenxue Dai received his PhD of Civil Engineering in 2000 from the University of La Corunia, Spain. Since 2010, he has been a Fellow of the Geological Society of America. Currently he mainly conducts research on AI-based across-scale monitoring and modeling of reactive chemical transport in heterogeneous subsurface systems. He has published more than 200 papers in the mainstream academic journals across environmental, geoscience and hydrology. He was recognized as one of the World's Top 2% Scientists by Stanford University for three consecutive years (2021-2023).