Journal of Environmental Chemical Engineering 14 (2026) 122193

Contents lists available at ScienceDirect

Journal of Environmental Chemical Engineering

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

Advances and challenges in machine learning-based identification of organic pollutant sources in heterogeneous aquifers

Yue Zhang [a], Mingxu Cao [a], Zhenxue Dai [a,b,*], Hao Wang [c,**], Sida Jia [d], Lulu Xu [a], Xiaoying Zhang [a], Mohamad Reza Soltanian [e], Javier Samper Calvete [f], Huichao Yin [g], Kenneth C. Carroll [g]

[a] State Key Laboratory of Deep Earth Exploration and Imaging, College of Construction Engineering, Jilin University, Changchun 130026, China
[b] School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao 266033, China
[c] Water Resources Research Institute of Shandong Province, Jinan 250013, China
[d] School of Resources and Safety Engineering, Chongqing University, Chongqing 400044, China
[e] Departments of Geosciences and Environmental Engineering, University of Cincinnati, Cincinnati, OH 45220, USA
[f] Civil Engineering School and Department, University of A Coruña, Campus de Elviña, A Coruña 15071, Spain
[g] Plant & Environmental Sciences Department, New Mexico State University, Las Cruces, NM, USA

ARTICLE INFO

Keywords:
Machine learning
Organic pollutant source identification
Heterogeneous aquifers
Surrogate model
Uncertainty quantification

ABSTRACT

Accurate identification of toxic organic pollutant sources (OPS) is essential for sustainable water management and aquifer remediation. Traditional methods struggle with increasing complexity and uncertainty of monitoring data collected from different sources. Machine learning (ML) enhances data fusion and feature extraction for source identification and migration path characterization. Distinct from previous reviews on general hydrogeological ML applications, this study provides the first synthesis specifically dedicated to organic pollutant source identification (OPSI). It highlights significant advancements in the integration of multiple-source monitoring data with recent modeling techniques. The review emphasizes the diversity of organic pollutants (OP) and the complexity of their transport. It also examines ML applications for inverting high-dimensional, non-Gaussian hydrogeological parameters and stresses how surrogate models boost computational efficiency and summarizes popular ML algorithms used in this field. However, ML algorithms for source identification face challenges as their "black-box" nature limits interpretability, and high computational demands. In addition, effective ML applications rely on a robust monitoring network, and the quantification of uncertainty during the identification process remains challenging. To advance ML-based OPSI in heterogeneous aquifers, future research should prioritize: (i) Improving integration of multi-source heterogeneous data; (ii) Optimizing monitoring methods and networks to utilize data more comprehensively; (iii) Employing physics-informed and explainable deep learning (DL) to enhance model interpretability; and (iv) Developing or exploring new computational paradigms to improve identification accuracy and reduce uncertainty. Overcoming these challenges with emerging ML technologies will enable real-time OPSI and source tracking for cleaning up of the contaminated heterogeneous soils and aquifers.

1. Introduction

Hazardous chemicals are now extensively used in agriculture, daily life, and industry with global economic development. This trend is intensified by population growth driving rising demand for plastics, pharmaceuticals, pesticides, and petrochemicals [1]. OP (e.g., total petroleum hydrocarbon (TPH), benzene, toluene, ethylbenzene, and xylenes (BTEX), polycyclic aromatic hydrocarbons (PAHs), polyfluoroalkyl substances (PFAS), polychlorinated biphenyls (PCBs), phthalates (PAEs)) from these sources are released into the environment [2,3]. Due to their chemical stability, bioaccumulation potential, and toxicity (including mutagenic and carcinogenic effects), they

* Corresponding author at: State Key Laboratory of Deep Earth Exploration and Imaging, College of Construction Engineering, Jilin University, Changchun 130026, China.
** Corresponding author.
E-mail addresses: dzx@jlu.edu.cn (Z. Dai), sdsslkxyjywh@163.com (H. Wang).

https://doi.org/10.1016/j.jece.2026.122193
Received 29 November 2025; Received in revised form 11 February 2026; Accepted 12 March 2026
Available online 13 March 2026

persistently threaten ecosystems and contaminate groundwater [4,5]. This contamination requires urgent attention since groundwater constitutes ~97% of global freshwater and provides a primary drinking water source [6].

Once released into the subsurface, these OP exhibit diverse physical and chemical behaviors depending on their solubility and density [7]. For highly water soluble OP represented by certain pesticides and emerging contaminants such as PFAS, transport is primarily governed by advection, dispersion, and diffusion, while sorption remains a critical factor in both vadose and saturated zones [8]. Similarly, BTEX, due to their high solubility, readily migrate as dissolved plumes [9], exhibit strong mobility, which lead to the rapid expansion of the contamination range [10]. Conversely, hydrophobic OP such as PAHs exhibit high stable and low solubility [11]. Due to their persistence, PAHs act as long-term contamination sources, continuously releasing into groundwater despite their slower migration rates [12]. However, a critical challenge overlooked by traditional OPSI is that these distinct behaviors rarely occur in isolation. In field-scale applications, the spatial superposition of advectively mobile soluble OP and persistent hydrophobic OP gives rise to heterogeneous density-stratified source zones, among these, non-aqueous phase liquids (NAPLs) (including light non-aqueous phase liquids (LNAPLs) and dense non-aqueous phase liquids (DNAPLs)) pose the most significant environmental hazards, ultimately leading to the formation of highly complex, nonlinear subsurface contamination signatures [13,14]. Adding to this complexity, environmental factors dynamically influence these distributions. Factors such as groundwater level fluctuations [15], temperature [16], microbial communities [17], pH, electron acceptors, and salinity [18] influence organic contaminant distribution [19–21]. Taking BTEX as an example, these compounds partition from the NAPL into dissolved and gaseous phases. Gaseous-phase BTEX volatilize and ascend through soil pores, posing inhalation risks, while aqueous-phase BTEX migrate with groundwater flow, potentially contaminating drinking water sources [22]. This equilibrium is not static, significant water table changes and elevated temperatures can enhance biodegradation rates. Pollutants with lower adsorption affinity, higher solubility, and volatility readily diffuse, releasing into the atmosphere and groundwater [23]. Meanwhile, microbial communities influence the conditions under which biochemical reactions occur. Under anaerobic and slow microbial metabolic conditions, many aromatic hydrocarbons degrade very slowly or not at all [24]. Despite strict prioritization by international frameworks like the Stockholm Convention and the U.S. environmental protection agency (2025) [25], effective source control remains elusive. The distribution of these contaminants in heterogeneous aquifers (HA) is characterized by high concealment, extreme spatial variability, and non-linear dynamic evolution [26,27].

These theoretical complexities translate into substantial practical hurdles for OPSI (see Fig. 1). First, the observation gap: aquifer heterogeneity (e.g., preferential flow paths, low-permeability barriers) complicates direct observation of pollution extent [28–30]. Monitoring methods such as borehole sampling yield sparse data at discrete locations, hindering the capture of continuous spatiotemporal plume evolution [31]. This sparsity increases uncertainty in pollutant transport inverse modeling within heterogeneous settings [32]. Second, the curse of dimensionality: hydrogeological parameters frequently exhibit high-dimensional, non-Gaussian characteristics. Traditional models struggle to represent these complexities, increasing uncertainty in source-zone reconstruction [33,34]. Third, the ill-posedness: while source terms and geological parameters are key inversion variables, the impacts of boundary and initial conditions are often neglected [34–36]. Uncertainty in both initial conditions and boundary conditions can obscure parameter correlations and distort pollution history [37]. Finally, the computational bottleneck: physical model-based inversions (e.g., MODFLOW, MT3DMS) require extensive iterations, resulting in prohibitive computational costs [38,39]. While surrogate models offer computational efficiency gains, they present challenges such as high dimensionality and the need for novel algorithms [40,41]. Similarly, data-driven models risk bias if physical mechanisms are inadequately embedded [40,42].

To bridge these observational gaps and mitigate the ill-posedness of source inversion, traditional OPSI frameworks have seen the development of many specialized techniques, without effective integration among them. Geochemical fingerprinting (e.g., chromatography and mass spectrometry) combined with multivariate statistics (e.g., Principal Component Analysis (PCA) [43]) to distinguish source types and their degradation products [44–46]. However, relying solely on discrete sampling points often fails to capture the spatial continuity of deep, concealed source zones in heterogeneous aquifers. Geophysical techniques (e.g., resistivity, electromagnetic, ground-penetrating radar, seismic) provide non-invasive imaging by detecting subsurface physical property anomalies [47–50], thereby establishing correlations between shallow and deep pollutants and identifying OP in heterogeneous environments [51]. However, the indirect nature of these measurements and the inherent non-uniqueness of geophysical inversion often introduce significant uncertainty into source identification. Consequently, conventional hydrogeological models and traditional investigation approaches, despite the development of various specialized techniques, often struggle to decouple these intricate dependencies. Given the focus on computational inversion, this review does not elaborate on the specific technical implementation of these geochemical and geophysical investigation methods. To address these bottlenecks and quantify the associated uncertainties, researchers have developed specialized probabilistic and geostatistical inversion frameworks. These approaches include response matrix optimization [52], stochastic random walk-based backward tracking models [53], nonlinear maximum likelihood estimation [54], and geostatistical inversion methods [55]. While

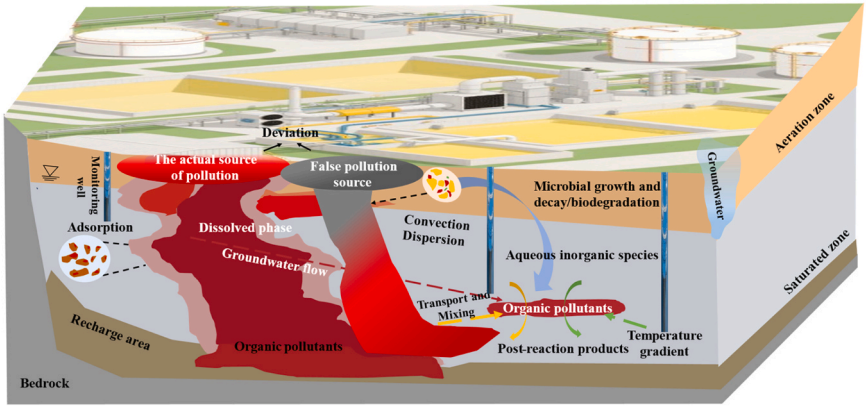

Fig. 1. Schematic diagram of challenges in identifying organic pollutants in heterogeneous aquifers (Modified from [12]).

these methods effectively quantify data correlations and improve source identification to some extent, they remain fundamentally limited. They exhibit strong dependency on prior assumptions and often struggle to achieve high-precision uncertainty quantification in complex heterogeneous systems characterized by sparse data and highly nonlinear processes.

To transcend the limitations of these conventional investigation and inversion frameworks, ML offers a robust alternative by excelling at nonlinear fitting, high-dimensional data processing, and adaptive optimization [56]. Consequently, ML has gained significant traction in hydrology for enhancing the accuracy and efficiency of OPSI [57]. However, previous reviews have predominantly focused on forward problems, groundwater level forecasting [58,59], groundwater quality prediction [60,61], parameter estimation for hydrogeological modeling [62,63]. A comprehensive synthesis specifically addressing the inverse problem of OPSI remains notably absent. OPSI constitutes a highly ill-posed inverse problem that requires reconstructing unknown source characteristics (e.g., location, intensity, and release history) from sparse monitoring data. This task is complicated by the non-uniqueness of solutions and the high nonlinearity inherent in heterogeneous aquifers. Initial applications used basic models such as Artificial Neural Networks (ANN), Generative Adversarial Networks (GAN), and Support Vector Regression (SVR) to map monitoring data to pollutant source release histories [64,65]. Furthermore, integrating multimodal data (time series, text, images [66]) improves data quality and hidden source identification [33,67]. Advances in artificial intelligence (AI) introduced DL models such as Transformer and ResNet are just starting to be used in pollutant prediction. These handle complex inversion tasks, including hydraulic parameters [40,56,68], and serve as efficient surrogate models within iterative frameworks, reducing computational costs. Recently, physics-informed approaches like Physics-Informed Neural Networks (PINN) have emerged as a fusion of process-based and data-driven modeling. PINN incorporate physical laws (e.g., partial differential equations (PDE), boundary conditions) as loss function regularization [69], improving generalization in data-sparse regions and mitigating inversion ill-posedness, equifinality, or non-identifiability [70,71]. Collectively, these ML advancements drive progress in OPSI.

Although ML has significantly advanced OPSI, practical applications face dual challenges: the intricate physicochemical behaviors of contaminants (e.g., migration, biodegradation, and adsorption) and the fundamental bottleneck of sparse monitoring data. Unlike previous reviews, this study is uniquely structured around the critical stages of the source identification process, systematically evaluating how ML strategies address specific challenges from data acquisition to mechanism integration. The main contributions of this paper are:

- Addressing the bottleneck of sparse monitoring data requires integrating multi-source heterogeneous datasets. A key future direction lies in developing dynamically optimized monitoring networks, which will significantly enhance data spatial representativeness and cost efficiency to support intelligent pollution source identification.
- Emphasizes the critical importance of accurately characterizing non-Gaussian parameter fields and systematically examines recent advances in intelligent characterization techniques (i.e., intelligent algorithms used to characterize the features of hydrogeological parameter fields). Furthermore, it demonstrates the necessity of synergistically inverting pollution sources, geological parameters, and boundary conditions.
- Analyzes how ML-based parameterization methods efficiently represent non-Gaussian fields by mapping high-dimensional parameters to low-dimensional latent features, thereby facilitating rapid surrogate modeling.
- Clarifies the three main limitations of ML-based OPSI in heterogeneous aquifers: (i) absence of physical mechanisms in algorithmic models and low computational efficiency, (ii) data sparsity due to monitoring network constraints, and (iii) lack of a systematic uncertainty quantification framework.

This review is divided into six sections. Section 2 overviews the mathematical models used to describe the spatiotemporal evolution of OP in heterogeneous aquifers, the challenges of sparse monitoring data, and the need for optimized monitoring networks. Section 3 highlights key technologies for OPSI. Section 4 summarizes common ML algorithms and models. Section 5 outlines the limitations of ML-based identification approaches. Section 6 explores future directions in OPSI.

2. Mathematical modeling and data challenges in heterogeneous aquifers

2.1. Mathematical models

In HA, the migration-transformation of OP is governed by two dominant trends [72]. The first trend is the migration trend driven mainly by the hydraulic field. The second is the attenuation, primarily influenced by heat, chemistry, diffusion, dispersion, and microbial processes [12,73,74]. Therefore, to quantify the multi-process coupled migration of OP in heterogeneous structures, a convection-dispersion-reaction transport equation can be established. The transport equation includes different transport mechanisms and reaction processes (Fig. 2). Hydrodynamics serve as the primary driving force for pollutant migration, typically determined by Darcy's law. The velocity of each phase in Darcy's scale is calculated using the multiphase Darcy's law equation (Eq. 1) [75], The reactive migration equation for groundwater pollutants (including both mobile and immobile species) is regulated by Eq. 2 [76].

$$v_\beta = -k\frac{k_\beta}{\mu_\beta}(\nabla P_\beta - \rho_\beta g) \tag{1}$$

$$\frac{\partial C_n}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ii}\frac{\partial C_n}{\partial x_i}\right) - \frac{\partial}{\partial x_i}(v_i C_n) + R_{reac,n} + \frac{q_s}{\theta}C_n^s \tag{2}$$

The soil (vadose zone) and aquifer media contain substantial organic matter, minerals, colloids, and microorganisms, providing favorable conditions for biochemical processes such as adsorption, degradation, and dissolution of OP. Adsorption-desorption governs contaminant storage capacity and effective diffusion efficiency [77], determining the partitioning behavior of OP between the vadose and saturated zones. OP in aquifers frequently exhibit desorption hysteresis or irreversibility in natural environments [78], necessitating investigation through various kinetic and thermodynamic models. However, standard models often

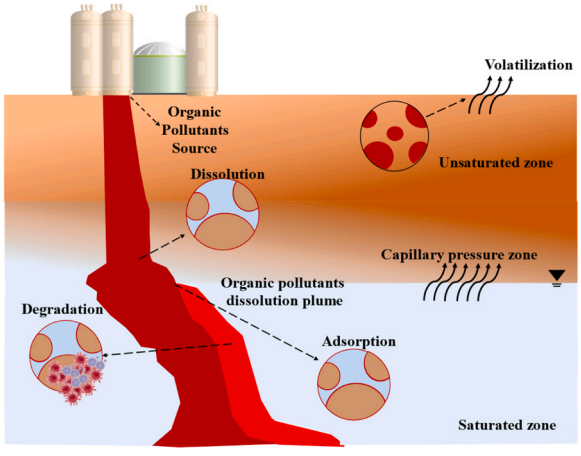

Fig. 2. Schematic diagram of the migration mechanism of organic pollutants in heterogeneous aquifers.

oversimplify complex transport behaviors. First, regarding kinetics, rate-limited sorption and desorption often drive non-equilibrium behaviors in aquifers, resulting in early contaminant breakthrough and extended elution tails. Predictive models that neglect these kinetic limitations may amplify uncertainty in source identification [79]. Simultaneously, recent modeling demonstrates that non-linear Freundlich isotherms are essential to capture transport behaviors across wide concentration ranges, where simplified linear assumptions fail. Neglecting these water-solid adsorption factors, which are identified as dominant determinants of retardation, can significantly compromise the accuracy of source identification and management strategies [80]. Furthermore, in multi-contaminant scenarios, mixture effects may exacerbate isotherm non-linearity and promote organic pollutant leaching once aqueous concentrations exceed critical thresholds. This competitive facilitated transport implies that single-solute models likely underestimate plume extent in high-concentration source zones [81].

To quantify these transport mechanisms, representative mathematical models are established. The Langmuir isotherm (Eq. 3) represents the fundamental adsorption model, describing monolayer surface adsorption with identical, non-interacting binding sites [82].

$$Q_e = Q_{max} K_l C_e / (1 + K_l C_e) \tag{3}$$

While Langmuir provides a theoretical baseline, the Freundlich isotherm (Table 1) is often more applicable for organic pollutants in aquifers, as it accounts for surface heterogeneity and non-linear sorption capacities [83]. Regarding adsorption kinetics, the pseudo-first-order kinetic model is a widely used method for analyzing data obtained from the adsorption of adsorbates from gas and solution phases. It generally indicates that the adsorption process is primarily driven by physical adsorption or weak interactions of the solute on the adsorption sites, rather than chemical adsorption. The expression is as follows:

$$\frac{d\theta}{dt} = k_a C_0 (1 - \varphi) - k_d \varphi \tag{4}$$

In contrast, the pseudo-second-order kinetic model is based on the adsorption capacity of the solid phase and is widely used to describe the time evolution of adsorption under non-equilibrium conditions. It should be noted that an important fundamental assumption of the pseudo-second-order kinetic model is that the adsorption reaction on the surface of the adsorbent is the rate-controlling step:

Table 1
Commonly used sorption and desorption models [77,83].

Model	Model Formula	Assumptions
Freundlich [84]	$Q_e = K_f C_e^n$	Assumes nonlinear sorption with no maximum sorption capacity.
Extended Freundlich [85]	$Q_e = \frac{K_i^{1/n} C_i}{\left(\sum_{j=1}^{N} K_j^{1/n} C_j\right)^{1-n}}$	Assumes a homogeneous surface with equal sites, identical maximum loading for all adsorbates, and no interactions among adsorbed species.
Distributed Reactivity Model [86]	$Q_e = x_1 K_{Dr} C_e + \sum_{i=1}^{m} (x_{nl})_i K_{Fi} C_e^{ni}$	Posits that the overall sorption isotherm for a natural solid is the sum of active component isotherms, exhibiting either linear or nonlinear behavior.
Dual Model [87]	$Q_T = K_{om} C_e + Q_{max,D}$	Combination of linear isotherm model and Langmuir model. Diffusion coefficients for two populations are very different.
Three-Domain Model [88]	$\frac{S_t}{S_0} = F_r \exp(-k_r t) + F_s \exp(-k_s t) + F_{vs} \exp(-k_{vs} t)$	Neglects back adsorption and internal diffusion between different adsorption sites.
Elovich's Model [89]	$q_t = \frac{1}{b_1} \ln(ab_1 t) = \frac{1}{b_1} \ln(ab_1) + \frac{1}{b_1} \ln(t)$	Posits that activation energy increases with adsorption time on a heterogeneous adsorbent surface.

$$\frac{dq_t}{dt} = k_2 (q_e - q_t)^2 \tag{5}$$

Permeability heterogeneity and pore structure within aquifers governs OP migration pathways. High-permeability zones form primary contaminant plumes through rapid advection, while low-permeability regions act as long-term contaminant reservoirs due to restricted diffusive transport [72]. Crucially, biodegradation is not uniform but is strictly controlled by thermodynamic hierarchies. Spatial variations of electron acceptors—including dissolved oxygen, nitrate, Fe(III), and sulfate establish redox gradients that regulate spatiotemporal patterns of microbial metabolism [90,91]. Geochemical analyses further validate these pathways; for instance, statistical tools like Mantel tests have confirmed a significant correlation between the depletion of electron acceptors (e.g., nitrate, sulfate) and the accumulation of metabolic byproducts (bicarbonate), providing direct hydrogeochemical signatures of anaerobic biodegradation [92]. Field investigations reveal that biodegradation is strictly redox-dependent, with optimal rates in iron-reducing zones consistently exceeding those in methanogenic and sulfate-reducing zones. Given the lower rate uncertainty in iron-reducing zones (e.g., for benzene), these areas offer more predictable attenuation. Neglecting this zonal heterogeneity significantly compromises the accuracy of source identification and plume modeling [93]. At the field scale, multivariate statistical frameworks have further quantified these drivers, revealing that hydrochemistry acts as the dominant factor controlling attenuation, with nitrate specifically identified as the primary determinant for benzene degradation [94]. Beyond electron acceptors, physicochemical conditions impose strict kinetic limits. Quantitatively, recent studies established functional relationships linking contaminant concentration to pH, dissolved oxygen, and oxidation-reduction potential, revealing distinct temporal phases where attenuation shifts from adsorption-dominance to biodegradation-dominance [95]. These biogeochemical interactions are modulated by temperature, salinity, pH, and other environmental factors [96]. While low temperatures often inactivate mesophiles, psychrotolerant microorganisms retain metabolic function, sustaining degradation processes that alter contaminant distribution and isotopic signatures even in cold aquifers [97]. Conversely, even a 5°C temperature increase significantly enhances petroleum hydrocarbon degradation and polycyclic aromatic hydrocarbons and alkane mixtures remain degradable under extreme temperatures [98,99]. Optimal biodegradation rates occur at specific salinity and pH thresholds [100]. Regarding the mathematical description of these rates, biodegradation typically exhibits two thermodynamic states: equilibrium conditions in which a dynamic balance exists between substrates and products, with rates governed by concentration distributions; and kinetic regimes influenced by environmental conditions and substrate concentrations, as well as microbial population growth and life-cycle dynamics. Single-substrate systems often follow first-order or zero-order kinetics, whereas multi-substrate environments often involve inhibition or competition (Table 2) [101,102]. The Monod model (Eqs. 6–7) remains fundamental for simulating zero-order, first-order, and mixed kinetic rates [103,104], with derivatives including Dual-Monod and Haldane models [105]. Substrate degradation kinetics are described by zero-order (Eq. 8), first-order (Eq. 9), second-order (Eq. 10) [106], where zero-order kinetics assume concentration-independent transformation (Eq. 6) and first-order kinetics exhibit linear concentration dependence (Eq. 7). While widely applicable, these models cannot distinguish aerobic/anaerobic conditions despite their utility in experimental contexts. The mixed-order model effectively combines first and second-order kinetics. Recent advances include a mathematical framework for multi-substrate interactions that accurately describes microbial growth and biodegradation of BTEX co-contaminants with chlorinated ethenes [107], though their broader applicability remains uncertain.

Table 2
Biodegradation kinetics models [108].

Model name Equation	Model name Equation
Monod [103]	$\mu = \frac{\mu_{max} S}{K_s + S}$
Andrews [109]	$\mu = \frac{\mu_{max} S}{K_s + S + S^2/K_i}$
Andrews and Noack [110]	$\mu = \frac{\mu_{max} S}{(K_s + S)(1 + \frac{S}{K_i})}$
Han–Levenspiel [110]	$\mu = \frac{\mu_{max}[1 - \frac{S}{K_i}]^n}{K_s + S - [1 - \frac{S}{K_i}]^m}$
Michaelis - Menten: two substrate reaction, competitive inhibition [111]	$\mu = \frac{\mu_{max} S}{S + K_s(1 + \frac{I}{K_i})}$
Two substrate, non - competitive inhibition [112]	$\mu = \frac{\mu_{max} S}{(S + K_s)(1 + \frac{I}{K_i})}$
Mixture of substrate, competitive inhibition [113]	$\mu = \frac{\mu_{max} S_i}{K_{si} + S_i + \sum_{j \neq i} S_j (\frac{K_{si}}{K_{sj}})}$
Mixture of substrate, non - competitive inhibition [113]	$\mu = \frac{\mu_{max} S_i}{K_{si} + S_i + \sum_{j \neq i} [S_j (\frac{K_{si}}{K_{sj}}) + \frac{S_i S_j}{K_{sj}}]}$
Mixture of substrate, uncompetitive inhibition [114]	$\mu = \frac{\mu_{max} S_i}{K_{si} + S_i + \sum_{j \neq i} \frac{S_i S_j}{K_{sj}}}$
SKIP, unspecific interaction [115]	$\mu = \frac{\mu_{max} S_i}{K_{si} + S_i + \sum_{j \neq i} S_i I_{ij}}$

$$R_s = -\mu_{max} \frac{X}{I_b} \left(\frac{S}{K_s + S}\right) \left(\frac{E}{K_E + E}\right) \tag{6}$$

$$\frac{dX}{dt} = -Yr_s - bX \tag{7}$$

When the rate of decrease in reactant concentration is proportional to its current concentration, and as time progresses, the reaction rate gradually decreases until the reactant is completely consumed or equilibrium is reached, first-order kinetics can be applied:

$$S = S_0 - k_0 t \tag{8}$$

$$S = S_0 \exp(-k_1 t) \tag{9}$$

The second-order kinetic equation indicates that the reaction rate decreases with the reduction of reactant concentration. Its rate constant is generally smaller than that of the first-order reaction, and it is often used to describe chemical reactions between two reactants [116]:

$$\frac{1}{C} = k \times t + \frac{1}{C_0} \tag{10}$$

2.2. Monitoring network optimization and high-quality data acquisition

Establishing a reliable monitoring scheme is the prerequisite for distinguishing anthropogenic sources from natural geochemical backgrounds [117]. To achieve this, integrated approaches combining multivariate statistics with isotopic tracers and hydrogeochemical inverse modeling have proven highly effective in defining regional natural background levels and resolving overlapping contamination sources (e.g., manure vs. fertilizers) [118,119]. The APCS-MLR receptor model, enhanced by entropy-weighted principal component analysis, has been successfully employed to quantitatively characterize source contributions (e.g., distinguishing salinity factors from industrial inputs), effectively overcoming the subjectivity of traditional feature extraction methods [43]. For complex OP, integrating high-resolution mass spectrometry (gas chromatography coupled to a high resolution mass spectrometry) with machine learning (e.g., partial least squares discriminant analysis) has established robust environmental forensic workflows. This approach successfully pinpointed diagnostic chemical indicators to distinguish between diverse contamination sources, achieving high predictive accuracy [45].

Once the target contaminants are defined, the physical layout of the monitoring network must be optimized. Sparse or poorly designed monitoring networks pose a critical challenge, as insufficient spatial density fails to capture the inherent heterogeneity of complex organic pollutant plumes. Therefore, network optimization is essential to maximize spatiotemporal information gain and minimize both data redundancy and source parameter uncertainty; methodologies such as information entropy theory are commonly used to quantify this information gain [120]. In practice, statistical approaches like time series clustering have proven effective in filtering duplicative data, with one study revealing that 30 out of 59 sensors were redundant, allowing for a streamlined design that prioritizes long historical records [121]. Moving beyond redundancy reduction, optimization strategies explicitly targeting high-concentration zones have demonstrated superior cost-effectiveness. For instance, by prioritizing areas with higher pollution source density or downstream of leachate pools, surrogate-assisted models achieved a 90.8% detection rate with only 15 wells [122]. Similarly, a quantitative optimization framework integrating GIS, Information Entropy, and principal component analysis was applied to select priority monitoring sites based on pollution source density and hydrogeological conditions. These results demonstrate that monitoring resources should be prioritized in high-density zones to maximize the cost-benefit ratio of data acquisition [123]. Furthermore, evolutionary algorithms have been employed to quantify the trade-off between performance and cost. For example, a Genetic Algorithm (GA) framework optimized network configuration by prioritizing areas with higher pollution levels to maximize Nash-Sutcliffe efficiency, though its reliability is constrained in karst aquifers affected by preferential flow [124]. While a multi-objective framework integrating Bayesian Maximum Entropy (BME) and NSGA-II selected just 5 out of 45 stations while retaining 76% of basin-wide information, demonstrating that significant cost reductions can be achieved without compromising data integrity [125]. However, trade-offs exist, for example, while stochastic simulation methods (incorporating K-means and modified Relevance Vector machine (RVM)) optimized the monitoring network to 25 wells (15 fewer than standard RVM) and reduced calibration time by 50–55%, they are often computationally demanding, limited to steady-state applications, and reliant on extensive historical water level data [126].

To further address spatial stochasticity and bridge data gaps in complex environments, integrating geophysical data has emerged as a robust solution. A transition probability and entropy-based framework was proposed, which utilizes a frequency-based statistical filter to prioritize stations with consistently high data worth, thereby minimizing sensitivity to model noise. Integrating Electrical Resistivity Tomography (ERT) into this network design significantly enhances system robustness, as ERT data exhibit strong noise resistance and reduce the uncertainty of solute concentration inversion in data fusion scenarios. However, the network configuration must carefully account for the depth-dependent sensitivity of geophysical data to ensure that the assimilated multi-source information provides sufficient physical constraints on deep subsurface structures [127]. Complementing this, to resolve vertical data gaps in complex coastal aquifers, an integrated approach combining geochemical sampling with geophysical well-logging was employed. This synergy allowed for the derivation of high-resolution vertical salinity profiles from formation resistivity, effectively distinguishing shallow saline intrusion from deep fresh fossil water and providing a robust basis for three-dimensional (3D) monitoring network configuration. [128].

Accurate source identification relies fundamentally on capturing the temporal variability of contaminant plumes, necessitating rigorous sampling frequency design and the integration of advanced neural network-based real-time monitoring systems [129]. While early

statistical approaches utilized confidence intervals to ensure baseline representativeness, rigorous optimization is required to balance cost and data density [130]. Addressing this, a novel spatiotemporal redesign of the groundwater level monitoring network in the Dehlran Plain employed a data fusion approach combining Spatiotemporal Kriging and ANN. By utilizing the value of information metric to determine non-uniform sampling frequencies, the optimized design differentiated between high-priority (20-day interval) and low-priority (32-day interval) sub-regions. This strategy successfully reduced the number of monitoring wells from 52 to 42 (a reduction of 10 wells) while maintaining the same estimation variance, thereby significantly lowering maintenance and operational costs. However, the current framework relies on independent simulation models, and future work should integrate rigorous physical flow models (e.g., MODFLOW) and advanced fusion techniques (e.g., Bayesian data fusion) to better account for estimation uncertainties [131]. However, traditional low-frequency sampling may still miss rapid contaminant pulses. Consequently, high-frequency sensors (5–15 min intervals) are increasingly necessary to provide the granular data required for precise source apportionment in dynamic systems [132]. Furthermore, in the context of emergency response, a Bayesian-Markov chain Monte Carlo (MCMC) framework identified a "crucial time" phenomenon, defined as the point where source identification error and uncertainty converge to a stable minimum. Theoretical analysis revealed that crucial time is primarily controlled by dispersion effects, establishing a linear quantitative relationship between the relative crucial time and spatial information entropy. This entropy-based metric serves as a guideline for designing emergency monitoring networks by determining when data acquisition yields sufficient accuracy. However, the current framework is limited to one-dimensional (1D) steady-state models, necessitating future validation in complex, real-world transport systems [133].

In addition to the methods discussed above, DL now addresses the persistent challenges of parametric uncertainty and data scarcity in HA (Fig. 3). To address parametric uncertainty and data scarcity, DL approaches are emerging as robust alternatives. Addressing the "black box" nature of ML, an attention-based Graph Neural Network (aGNN) was developed to model contaminant transport and quantify source-receptor causality. By integrating attention mechanisms with spatiotemporal embedding layers, the aGNN effectively captured highly nonlinear dependencies. Crucially, the model demonstrated superior computational efficiency, accelerating processing speeds by approximately 300 times for large-scale sites with sparse wells. Furthermore, the study enhanced interpretability by employing SHAP values to quantify the contribution of each pollution source, aligning statistical correlations with physical transport principles. Despite these strengths, the model's accuracy remains sensitive to the volume of available training data, emphasizing the need for sufficient historical records to maximize transferability [134]. To address high-dimensional parameter uncertainty in heterogeneous aquifers, a framework integrating Deep Convolutional Generative Adversarial Network (DCGAN) and information theory was proposed. By employing DCGAN for parameterization, the study utilized the Maximum Information Minimum Redundancy (MIMR) criterion to identify "hotspot maps" based on selection probability. This approach effectively optimizes monitoring locations in non-Gaussian random fields, offering a robust alternative to traditional methods. However, the current framework relies on deterministic subsurface processes, and future work must incorporate model process uncertainty to fully represent complex natural systems [135]. Regarding robustness, DL-based "ensemble MIMR" methods have demonstrated superior performance in handling sparse and noisy monitoring data, preventing the loss of key information due to stochastic randomness. To minimize uncertainties in high-dimensional permeability estimation, a data assimilation (DA) framework was developed by integrating DL-based surrogates with the MIMR criterion, and this ensemble MIMR-optimized method significantly outperforms conventional approaches in characterizing permeability fields. However, training globally accurate surrogates remains data-intensive; future research proposes adaptive update strategies and the integration of GAN to better characterize complex non-Gaussian features with fewer training samples [136]. Complementing this, a multivariate network design framework utilizing joint entropy demonstrated strong capability in estimating non-Gaussian permeability fields even under high-noise observations [137]. Moreover, ignoring uncertainties in hydraulic stresses (e.g., varying extraction rates) can lead to severe under-design, a framework integrating deep neural network and MIMR entropy criteria was proposed to optimize the detection of solute transport dynamics. The study quantified the critical impact of stress uncertainty, revealing that when groundwater extraction rates are uncertain, the required number of monitoring stations doubles, and optimal sensor locations shift from the aquifer bottom to the middle layer to capture the most informative concentration gradients. These findings underscore that ignoring boundary condition uncertainties (e.g., pumping variability) can lead to severe under-design of monitoring networks, a conclusion that is broadly applicable to contaminant source identification in complex flow fields [29].

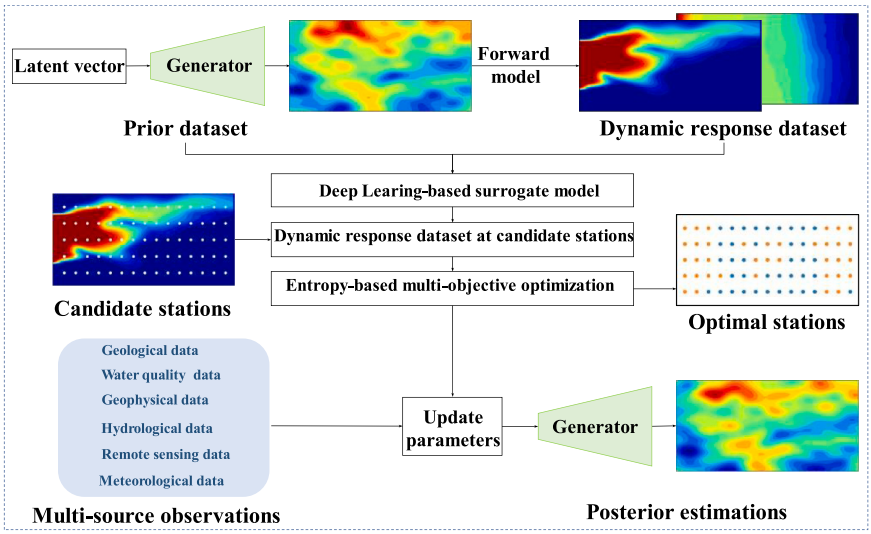

Fig. 3. Optimization of monitoring network and identification of organic pollutants based on deep learning.

3. Key technological innovations and breakthroughs

Building upon the mathematical foundations and data acquisition strategies outlined in Section 2, particularly the fact that even optimized monitoring networks often yield spatially sparse data for complex organic plumes, this section presents three interconnected technological innovations (Fig. 4). First, to maximize information extraction from these limited observation points, intelligent characterization, defined as generative deep learning strategies that reconstruct high-order topological features from sparse data, is introduced to resolve non-Gaussian heterogeneity. Second, addressing the coupling effects of organic contaminants, a unified synergistic inversion framework is proposed, this approach refers to the simultaneous estimation of contaminant source terms, geological parameters, and boundary conditions to mitigate solution non-uniqueness. Finally, to overcome the computational burden of iterating these high-fidelity models, surrogate model-driven acceleration is implemented to enable efficient inversion.

3.1. Intelligent characterization of non-gaussian parameter fields

Characterizing spatially heterogeneous formations like fractured aquifers is challenging due to prevalent non-Gaussian parameter distributions (e.g., hydraulic conductivity, source zone architecture). This challenge is particularly acute for organic contaminants, whose migration is strongly governed by these complex structures. Integrating hydrological and geophysical data, such as time-lapse ERT and tracer concentrations, decreases non-uniqueness but faces limitations from petrophysical uncertainties linking hydraulic conductivity and resistivity [138]. To mitigate this, a MCMC methodology is employed to explicitly quantify this uncertain relationship between electrical resistivity and concentration, thereby generating posterior realizations consistent with multi-physics data [139,140]. Furthermore, recent studies have advanced this integration by utilizing 3D electrical resistivity imagery to identify the heterogeneity of a contaminated aquifer, which serves as a robust parameterization basis for reconstructing hydraulic conductivity fields in contaminated aquifers under complex conditions (e.g., tidal influence) [141]. In contrast, the integration of hydraulic-head and self-potential (SP) data improves the characterization of non-Gaussian hydraulic conductivity [142,143]. Hydraulic tomography (HT) is also effective for the highly parameterized estimation of heterogeneous hydraulic properties. Notably, Guo et al. [144] pioneered this progress through HT-INV-NN, which trains inverse mappings using principal component analysis for Gaussian fields and GAN generators for non-Gaussian fields, demonstrating robust performance against input perturbations. Separately, Ji et al. [40] focused on sequential HT data processing, demonstrating that sequence models (Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Transformer) outperform Convolutional Neural Networks (CNN)-based encoders in training efficiency. While these advanced inversion architectures significantly enhance parameter estimation efficiency, a fundamental challenge remains in reconciling the scale disparity between microscopic observations and macroscopic predictions. To solve this problem, recent research treats the mass transfer coefficient as a lumped spatial random variable to bridge the gap between laboratory/measurement scales and the field scale. By integrating microscopic heterogeneities (e.g., matrix porosity, tortuosity, and mineral facies), an effective sorption coefficient was derived and validated via MC simulations to accurately represent field-scale transport [145]. In a broader context, upscaling has been proven as a valid approach to estimate large-scale parameters using small-scale data, effectively bridging the gap from micro-pores to regional reservoirs. Existing methodologies are generally categorized into deterministic approaches (e.g., volume averaging) and stochastic methods. While numerical solutions are gaining popularity, future frameworks aim to integrate uncertainty quantification and artificial intelligence to validate these models against multi-scale observations [146]. Building on these foundations, a critical frontier in current research is bridging the immense scale gap between pore-scale micro-structures (e.g., Micro-CT images) and field-scale tomography. DL technology offers a solid potential to bridge this gap by extracting pore geometry and estimating physical properties directly from CT images [147]. Various architectures have been successfully applied at the pore scale, GAN effectively reconstruct porous media characteristics [148], while 3D CNN have been used to predict flow velocity fields and permeability directly from pore structures [149]. Looking forward, addressing this issue remains a critical challenge in the domain of contaminant source identification.

Characterizing these heterogeneous formations inevitably confronts the dual challenges of high dimensionality and non-Gaussian data distributions. Gaussian assumption operates within distinct conditions, primarily applicable to homogeneous or weakly heterogeneous formations characterized by continuous parameters and low spatial correlation of extremes [150]. However, their reliance on the Gaussian assumption limits robustness under non-Gaussian conditions. Consequently, variable transformation techniques have been widely adopted to convert non-Gaussian fields into Gaussian-distributed equivalents, enabling DA in complex formations. Common approaches include normal score transformation [151], Gaussian anamorphosis [152], and discrete cosine transform [153]. These transformations enable DA in complex formations. Fundamentally, this Gaussian assumption implies both unimodality and full support [154]. While attempts have been made to relax these limitations, the reparameterization and update process may result in the loss of some key features, with the Gaussian assumption still intact [155]. Recent advances in DL significantly enhance the characterization of non-Gaussian parameter fields within inverse problems, though each architecture presents distinct limitations. Variational Autoencoders (VAE) effectively represent non-Gaussian spatial distributions of rock properties as low-dimensional geological models with minimal fidelity loss [156]. Integrating VAE with other DL models substantially improves traditional parameterization methods. Coupling VAE with the Ensemble Smoother with Multiple Data Assimilation (ES-MDA) enables reconstructing well-defined channelized facies [157]. However, VAE are renowned for their training stability and speed, progressively improving image quality during learning. However, their reliance on Gaussian latent space assumptions often forces a

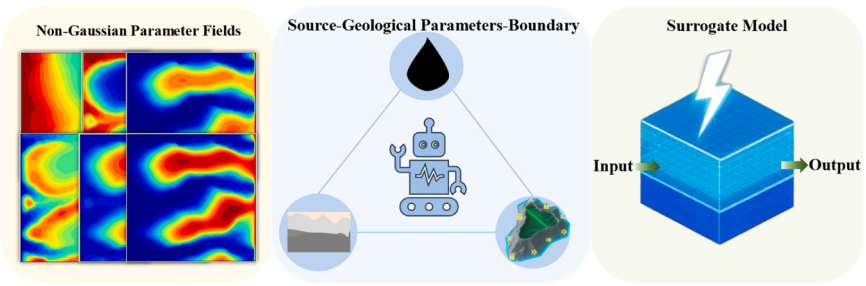

Fig. 4. Schematic diagram of three key technologies.

trade-off between model complexity and realism, frequently resulting in blurred or over-smoothed geological features [158]. Conversely, GAN are recognized for producing high-fidelity images, facilitating visualization and calibration of geological models [159,160], and can generate more accurate structural features than VAE [161]. Furthermore, utilizing enhanced-resolution CT images as inputs improves fluid flow simulation results and yields more realistic pore-scale distributions from GAN [162]. Nevertheless, standard GAN face inherent challenges, including non-convergence and mode collapse, where limited output diversity leads to the omission of critical geological variations. Linearity in the generator may lead to inaccurate inversion results, particularly when paired with stochastic gradient-based inversion techniques [163]. To address these defects, targeted architectural designs have been proposed. The State-Parameter Identification GAN (SPID-GAN) incorporates dual generator-discriminator pairs, demonstrating strong performance in identifying bidirectional mappings of state parameter [164]. Similarly, SGAN utilizes a fully unsupervised DC-GAN architecture to map spatial noise arrays via fractionally strided CNN, enabling arbitrary large-scale texture synthesis (e.g., fractures) to enhance diversity. An inversion framework combining Spatial GAN (SGAN) with MCMC can recover complex geological features closely approximating true models [159], and Style-Based Generative Adversarial Networks (StyleGAN) generate high-fidelity images through iteratively refined generators that precisely control image nuances [165]. Critically, however, limitations persist: SPID-GAN's dependence on high-quality paired data restricts generalization in unpaired scenarios, whereas SGAN lacks constraints for nonlinear and non-stationary characteristics (e.g., irregular faults). Consequently, SGAN's single forward-pass mechanism may struggle with texture morphing continuity and the detailed expression of rare geological variations [166]. A critical comparative analysis shows clear trade-offs between VAE and GAN in characterizing subsurface heterogeneity, prompting the proposal of Adversarial Autoencoders (AAE) to efficiently and stably generate subsurface sedimentary structures [167].

Beyond standalone architectures, recent research has shifted toward unified inversion frameworks, the ES_{DL} framework synergizes ensemble smoother with DL by directly learning update mechanisms rather than approximating forward models, effectively overcoming challenges of high dimensionality, non-Gaussian distributions, and feature loss in DA [34,155]. The GeoSinGAN-DOCRN-ILUES framework integrates Geological Single-Image Generative Adversarial Networks (GeoSinGAN), Deep Octave Convolutional Residual Dense Networks (DOCRN), and the Iterative Local Updating Ensemble Smoother (ILUES) to effectively characterize heterogeneous aquifer structures in both Gaussian and non-Gaussian geological environments [168]. For contaminant source estimation, GAN-ILUES and its optimized variants (e.g., GAN-Optimized AR-Net-WL(OANW)-ILUES)) enable joint inversion of plume intensities and hydraulic conductivity fields in non-Gaussian groundwater systems [169]. Most recently, diffusion-based architectures have set a new benchmark. The AEdiffusion-based framework combines diffusion denoising probabilistic model with VAE via a generator-refiner strategy to reconstruct high-dimensional hydraulic conductivity fields from low-dimensional latent representations. Although this approach exhibits superior feature extraction and stability, it incurs substantially higher computational costs than Wasserstein generative adversarial network with gradient penalty implementations [170]. Subsequent efficiency breakthroughs emerged through AEdiffusion-ILUES-ARNW, which jointly estimates contaminant source parameters and channelized hydraulic conductivity fields while reducing inversion time by over 55% without compromising accuracy [171].

3.2. Synergistic inversion of source-geological parameters-boundary conditions

Inverse identification of groundwater contaminant sources requires simultaneous determination of three core elements: contaminant source, heterogeneous geological parameters, and dynamic boundary conditions [172]. Boundary condition variations govern system forcing and stress conditions that influence the direction of contaminant diffusion and transport [173]. Geological parameters such as hydraulic conductivity fields govern preferential flow paths through heterogeneous media [174]. Contaminant source terms exhibit transient multi-source characteristics. Their nonlinear couplings with geological parameters and spatiotemporally varying boundary conditions significantly increase inversion uncertainty in heterogeneous aquifers. While conventional methods focus on sources and hydrogeological parameters, they often ignore boundary conditions [175]. This mismatch propagates substantial errors into inversion outputs, especially given the transient multi-source characteristics and nonlinear couplings with heterogeneous media. However, ML applications for rapid coupled inversion remain scarce.

Synergistic inversion of contaminant sources, hydrogeological parameters, and boundary conditions represents a critical frontier in groundwater contaminant source identification, where methodological innovations focus on quantifying their high-dimensional nonlinear couplings and associated uncertainty. Several recent studies demonstrate this progress: Wang et al. [176] proposed a novel ES-MDA approach using a "wheel battle" strategy. This method sequentially identifies and updates source information, model parameters, and boundary conditions within each assimilation cycle. As a critically important benefit, it maintains accuracy across different boundary condition modes (e.g., constant, sinusoidal, and single-peaked). Xu et al. [177] employed the Tree-based Pipeline Optimization Tool (TPOT), an AutoML tool, to concurrently construct five surrogate models (Extreme Gradient Boosting (XGBoost), Random Forest (RF), Extra Trees Regressor (ETR), and elasticnet method) for the simultaneous identification of source information, model parameters, and boundary conditions. This approach significantly enhanced computational efficiency and accuracy. Luo et al. [65] combined the DL method (Multilayer Perceptron (MLP)) with traditional ML (SVR), achieving rapid and collaborative identification of pollution source information, hydrogeological parameters, and boundary conditions. Among them, MLP had a higher recognition accuracy for pollution source information, while SVR had better recognition effects for hydrogeological parameters and boundary conditions.

3.3. Surrogate model-driven computational acceleration

Surrogate models enable rapid forward modeling by providing computationally efficient approximations of complex simulation input-response relationships. The systematic development of surrogate models progresses through three interdependent stages: data sampling, model construction, and validation. This workflow has seen steady evolution, driven by ongoing methodological innovations that refine its implementation. In the sampling phase, methodologies have transitioned from static MC designs to adaptive space-filling strategies (e.g., Latin hypercube sampling [178]; adaptive schemes [179]). Specifically, random sampling often suffers from poor space-filling characteristics, whereas adaptive methods, despite their high accuracy, incur prohibitive computational costs due to substantial sample requirements. To reconcile this, a hybrid strategy is widely adopted, utilizing optimal latin hypercube sampling to generate a limited initial dataset, followed by adaptive sampling that sequentially augments critical regions, thereby enhancing surrogate model validity [180]. Subsequently, modeling has advanced from single surrogates to ensemble frameworks [181], which achieve superior accuracy by aggregating multiple base learners. By leveraging the complementary strengths of different algorithms, this methodology effectively addresses source identification for both conservative and reactive contaminants—a significant advantage given the prevalence of chemical reactions in real-world scenarios. However, this enhanced capability introduces an inevitable trade-off between physical fidelity and computational efficiency: as decision variables increase, the

computational burden of high-fidelity data generation grows substantially [182]. Finally, addressing overfitting risks is critical to ensure robust generalization, necessitating interventions across model structure, training, and validation. Regarding model structure, innovations such as residual connections allow for deeper DL architectures without overfitting [62], while Proper Orthogonal Decomposition effectively constrains model complexity [183]. During the training process, algorithmic interventions including early stopping serve as essential safeguards [184]. Ultimately, validation protocols have matured from basic holdout tests to robust cross-validation schemes protocols [180]. Notably, while the dependence on rigorous validation varies as it is less critical for low-complexity models like Kriging, it remains an indispensable step for developing robust Support Vector Machine (SVM) and ANN models [185]. These advancements collectively address computational bottlenecks in OPSI.

Traditional surrogate paradigms exhibit critical limitations. Hierarchical approaches simplify physics or numerical resolution [186], whereas projection-based techniques compress governing equations into reduced-order subspaces. However, the latter are often highly code-intrusive and cannot efficiently treat strong nonlinearity [187]. Specifically, addressing limitations related to nonlinear model dynamics and complex boundary conditions often requires substantial modifications to the corresponding equations or solver codes. Although such modifications can improve accuracy, they inevitably increase computational time, thereby negating the efficiency benefits of model reduction [188]. Consequently, both paradigms often fail to deliver adequate results under strong nonlinearity [189]. This collective deficiency explains the ascendancy of data-driven intelligent surrogates [180]. For low-to-moderate dimensional problems, ML models (e.g., kriging [180], polynomial regression [68], radial basis functions [190], kernel extreme learning machines [191], SVR [192]) are widely applied due to their efficient implementation. For example, Chang et al. [193] constructed computationally efficient surrogates using kriging, ANN, and MLP to replace demanding forward models in the inversion of contaminant transport parameters.

Breakthrough in High-Dimensional Nonlinear Mapping For complex feature learning in high-dimensional spaces, DL has demonstrated powerful nonlinear mapping capabilities, effectively overcoming the curse of dimensionality that limits traditional surrogates. Forward simulation can be regarded as an image-to-image translation process. This image-to-image mapping surrogate model is one of the most widely used DL-based surrogate models in OPSI. Early foundational frameworks, a Bayesian deep convolutional encoder-decoder surrogate model based on Stein's method-based variational gradient descent, adapts the image-to-image regression approach from the field of computer vision to problems driven by stochastic partial differential equations, realizes high-precision uncertainty quantification and propagation for high-dimensional stochastic inputs in flow problems in heterogeneous media via approximate Bayesian inference on millions of network parameters, and achieves superior predictive accuracy and uncertainty quantification performance to competing techniques such as Gaussian processes and ensemble methods even with limited training data, with the Bayesian statistics of its predictive outputs matching the results obtained from MC estimates [194]. To further capture dynamic system behaviors, subsequent research advanced to Deep Convolutional Encoder-Decoder networks. In this architecture, the encoder extracts high-level coarse features from high-dimensional permeability fields, while the decoder refines these features to output pressure or saturation fields, thereby better representing the high-dimensional time-varying outputs of dynamic systems [195], this surrogate model integrated with a deep autoregressive neural network also enables high-dimensional parameter estimation for groundwater contaminant sources [196]. Handling non-Gaussian heterogeneity remains a specific challenge. Mo et al. [197] proposed the Deep Residual Dense Convolutional Network (DRDCN). This model accurately approximates high-dimensional and highly complex 2D and 3D forward models using limited training data. Furthermore, it operates in conjunction with the Convolutional Adversarial Autoencoder and ILUES (CAAE-ILUES) to successfully invert non-Gaussian hydraulic conductivity fields under heterogeneous patterns. Kang et al. [198] developed an Convolutional Variational Autoencoder (CVAE) as a surrogate for source zone structure inversion, and CVAE with an Iterative Ensemble Smoother (ES). This approach demonstrated effective prediction of DNAPLs depletion behavior and dissolved concentrations. Beyond structural fidelity, improving physical consistency is also critical. He et al. [199] coupled a Theory-guided Fully Convolutional Neural Network (TgFCNN) surrogate with ES. The proposed TgFCNN explicitly incorporates physical constraints (composed of residuals of contaminant transport governing equations) into its loss function. This physics-informed approach enhances network performance, achieving higher accuracy than conventional CNN models in identifying contamination sources and conductivity fields under various scenarios. In parallel, for multi-task DA, Xia et al. [200] resolved multi-task data assimilation (DA) (hydraulic head, contaminant concentration) by coupling Residual Dense Convolutional Networks (RDCNN) with ILUES, outperforming standard deep convolutional neural networks surrogates (lacking residual learning) and original models in convergence, mapping fidelity, and computational efficiency to achieve state-of-the-art inversion performance.

Alternatively, the second category of surrogate models is based on Point-to-Point mapping. This architecture is essentially a regression model, where various algorithms can be employed to construct the mapping. As noted by [193], hyperparameter optimization has been successfully applied to adjust shallow learning surrogates (e.g., Kriging, MLP, and ANN), demonstrating the feasibility of these models which benefit from shorter training times. However, the potential of DL surrogates cannot be ignored. Through more extensive training, DL models can achieve better effects. Consequently, if one aims for higher precision in surrogate modeling, DL-based methods should be considered, though it must be noted that they require relatively longer training times. In the context of deep learning applications, Zhou and Tartakovsky [201] integrated Deep Convolutional Neural Network (DCNN)-based surrogates with adaptive MCMC methods to reduce computational costs and sampling errors in approximating likelihood functions. This method was applied to reconstruct contaminant release history from sparse and noisy solute concentration measurements. Notably, properly trained autoregressive models and RNN can be strong competitors to CNN. Since they act as fixed time-step predictors, they exhibit better generalization capabilities. However, RNN may incur higher costs due to their higher prediction frequency. Conversely, for scenarios focusing on optimization efficiency, integrated frameworks utilizing shallow models remain prevalent. Furthermore, other specialized deep architectures have been leveraged to alleviate computational burdens. Luo et al. [202] employed Temporal Convolutional Networks (TCN) as surrogate models, combining the Metropolis-Hastings algorithm with the Kalman Filter (KF). This integration improved both the accuracy and efficiency of contaminant source identification. Similarly, Li et al. [36] employed Deep Belief Neural Networks (DBNN) with Unscented Kalman Smoother with Multiple Data Assimilation (UKS-MDA), which successfully boosted inversion efficiency by 12% versus ES-MDA for the joint identification of contaminant sources and hydraulic conductivity fields. Conversely, integrated frameworks utilizing shallow learning models remain prevalent for optimization tasks where training cost is a primary constraint. Conversely, integrated frameworks utilizing shallow learning models remain prevalent for optimization tasks where training cost is a primary constraint.Chang et al. [203] integrated adaptive kriging surrogate model (AKSM), ES, and DEMC methodologies to identify contaminant source attributes and model parameters in groundwater pollution. Similarly, Wang et al. [172] combined kriging surrogates with Ensemble Kalman Filter (EnKF) and Adaptive Step Length Ant Colony Optimization (ASACO) algorithms to enhance efficiency in OPSI. To further enhance robustness and generalization, Bian et al. [204] proposed a Bayesian ensemble machine learning surrogate method (incorporating

Gaussian process, SVR, and Kernel Extreme Learning Machine (KELM)). This surrogate approach serves to boost generalization capability while reducing computational load during inversion iterations. It was effectively integrated with a hyper-heuristic homotopy algorithm to enhance search ergodicity in inversion, alongside a homotopy-based swarm intelligence algorithm for improved stability in inversion.

However, regardless of the specific surrogate architecture employed, data scarcity critically hinders efficient deployment in scientific applications by substantially increasing training costs, thereby limiting acceleration potential. To mitigate this, transfer learning strategies have been adopted to leverage knowledge from related tasks [189]. Jiang and Durlofsky [205] reduced high-fidelity simulation costs via low-fidelity data transfer, while Zhang et al. [206] integrated analytical model knowledge into DL frameworks for complex groundwater flow. Furthermore, Fu et al. [189] developed an image-to-sequence surrogate using decoupled learning and geological posterior sampling with random maximum likelihood (RML). Beyond data constraints, model optimization remains a critical step. While standard hyperparameter optimization efficiently enhances classical ML surrogate models (e.g., ANN, Kriging) for low-to-moderate dimensional problems through direct input-output mapping [207], complex inversion tasks requiring DL surrogates necessitate adaptive closed-loop frameworks to maintain robustness. Chang et al. [208] demonstrated this by integrating ResNet-18 with an ES algorithm, where variable-density grid search optimizes hyperparameters during iterative LNAPLs source identification. Similarly, Xu et al. [56] leveraged ResNet's feature extraction within an IEPF, capturing nonlinear relationships while dynamically refining the surrogate model through inversion feedback.

Furthermore, distinct contamination source types (point, line, area) pose unique computational efficiency challenges for pollution source identification. While point sources dominate current research [171], pipeline leaks represent critical line sources requiring specialized inversion approaches [209]. Such scenarios create non-Gaussian hydraulic conductivity fields around conduits, compromising conventional surrogate models [210]. To address these specific challenges, specialized surrogate frameworks have been developed. Zhang et al. [171] introduced the AEdiffusion-ILUES-ARNW surrogate model. The framework was applied for inversion modeling on synthetic non-Gaussian hydraulic conductivity fields with line-source contamination, nevertheless, the computational process may be subject to error accumulation. Zheng et al. [169] employed a CNN-based optimized OANW surrogate within hybrid frameworks (GAN-ILUES and GAN-OANW-ILUES), enabling simultaneous estimation of line-source parameters and heterogeneous conductivity. Pan et al. [211] integrated DL (Simple CNN, ResNet, UNet) with DA (DREAM$_{(ZS)}$, ESMDA, ILUES) to identify 50 parameters of five pollution sources in rivers. UNet achieved the highest surrogate modeling accuracy, while ILUES demonstrated optimal performance among assimilation methods. The UNet-ILUES framework enhanced computational efficiency by 406 times compared to river water quality model -ILUES, however, it exhibits a strong dependence on data.

In summary, the transition from traditional shallow learning to deep learning architectures represents a fundamental shift in surrogate modeling strategies, each offering distinct trade-offs between computational efficiency, physical fidelity, and data requirements. While deep learning models demonstrate superior capabilities in capturing high-dimensional heterogeneity and non-linear dynamics, they often demand extensive training datasets compared to traditional methods. To provide a clear framework for method selection, Table 3 presents a systematic classification and critical comparison of these surrogate paradigms, evaluating their underlying mapping mechanisms (data type), specific advantages, inherent limitations, and optimal applicability scenarios in organic pollutant source identification.

In summary, this section addressed key innovations in characterizing non-Gaussian parameter fields and accelerating computations via surrogate models. However, the practical implementation of these strategies relies on the continuous evolution of underlying algorithms. Consequently, Section 4 will critically examine the progression of inversion methodologies from traditional optimization to physics-informed learning, demonstrating how these algorithmic advances specifically tackle the complex hydrogeological challenges identified above.

4. ML-based inversion algorithm evolution

4.1. Optimization-based approaches

Optimization-based source identification methods determine

Table 3
Critical comparison of surrogate modeling in organic pollutant source identification.

Category	Model	Data type	Advantages	Limitations	Optimal Applicability
Traditional Shallow Learning	SVM, RF, MLP, Kriging, ANN, AKSM	Point	Extremely fast to build with small datasets; Easy to implement and integrate with standard optimization algorithms; Robust for linear or weakly non-linear problems.	Performance degrades rapidly as parameter count increases; Cannot capture complex spatial heterogeneity.	Low-dimensional inversion; Homogeneous or simple heterogeneous aquifers; Scenarios with very limited training data.
DL	CNN, DCN, RDCNN, DBNN, FNN	Point/ Image	Overcomes the curse of dimensionality; captures intricate spatial connectivity and heterogeneity;	Requires massive amounts of high-fidelity simulation data for training; Incurs significantly longer training times; Deep architectures prone to overfitting.	High-dimensional parameter estimation; Non-Gaussian heterogeneous fields; Complex spatial pattern reconstruction.
	GAN, VAE, UNet, ResNet	Image	Automatically extracts high-level features without manual engineering.		
DL	TCN, LSTM	Sequence/ Image	Exhibits better generalization capabilities for time-varying systems compared to static CNN; Effective for reconstructing contaminant release history from time-series concentration data.	Can incur higher computational costs during inference due to high frequency of time-step predictions; Standard RNN may suffer from vanishing gradients in long sequences.	Identification of time-varying release histories; Dynamic monitoring data assimilation; Real-time forecasting.
Physics-Informed & Theory-Guided	TgFCNN	Image	Ensures mass conservation and plausible geological realism, reducing "black-box" uncertainty.	Integrating PDE residuals complicates the loss landscape, potentially leading to convergence issues; Requires deep understanding of both DL and numerical differentiation.	Scenarios with sparse monitoring data; Problems where physical validity is non-negotiable; Strongly non-linear reactive transport.

unknown contaminant source characteristics through systematic minimization discrepancies between model outputs and field observations [212]. While pioneering linear optimization methods demonstrated efficacy in characterizing pollution sources within simplified 1D steady-state and two-dimensional (2D) transient flow systems through linear programming and response matrix-based least-squares regression [52], their applicability is fundamentally constrained by pervasive nonlinearities in field conditions. Conventional nonlinear optimization techniques frequently converge to local optima when initial parameter estimates are suboptimal, creating significant solution limitations. To overcome these constraints, heuristic algorithms including GA, Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Simulated Annealing (SA) enable global optimization, enhancing computational robustness [41]. Nevertheless, these methods incur substantial computational costs through repeated transport simulations, and their performance remains vulnerable to observational noise, particularly evident in adaptive simulated annealing's sensitivity to initialization quality. This prompted the development of hybrid optimization frameworks that strategically integrate global and local search mechanisms. For instance, sequential hybrids achieve complementary advantages by refining GA-derived solutions through local optimization, balancing convergence efficiency with solution precision [213]. A sequential hybrid framework combining simulated annealing and tabu search effectively resolves source localization and contaminant release history reconstruction, utilizing tabu search for spatial identification and simulated annealing for release history reconstruction [214].

Recent methodological innovations directly confront persistent computational bottlenecks through adaptive metaheuristics typified by ASACO, where dynamic parameter modulation accelerates convergence while circumventing local optima [172]. Differential evolution (DE) algorithms frame the source inversion as an optimization problem, minimizing the sum of squared errors between observed and predicted contaminant concentrations to estimate source parameters including location, mass flux, and release timing [215]. When facing critical limitations, including exponential computational scaling in high-dimensional spaces, local optima convergence, and intensive simulation requirements, these challenges are now overcome through surrogate-optimization frameworks. The Multiverse Optimization (MVO) algorithm excels at balancing parameter space exploration and convergence to optimal solutions, proving particularly valuable for complex nonlinear relationships in datasets [216]. DL surrogates like entity-aware sequential long short-term memory (EAS-LSTM) networks coupled with MVO achieve over 1000 times speed up while accurately identifying release rates and transport parameters [216].

4.2. Stochastic-based approaches

Stochastic-based approaches analyze large sample sets to characterize distribution patterns and quantify measurement uncertainty. Rooted in probability theory, these methods construct prior distributions from existing data and update posteriors via observations, enabling contaminant source inversion. Critically, they explicitly quantify uncertainty in inversion results [217]. Probabilistic frameworks further advanced methodological rigor. The backward probability method introduced backward location and travel-time probabilities to identify potential prior source locations and release timing, respectively [218]. Recent innovations include multi-probability-density-function integration for identifying multiple sources under complex hydrogeological conditions involving 3D transient variably saturated flow [219], and vectorized models resolving anomalous transport in arbitrarily HA [220].

However, the aforementioned probabilistic frameworks often fail to adequately characterize spatial correlations in heterogeneous media, thereby increasing error probability. To address this structural limitation. Geostatistical (GS) approaches were introduced, offering significant advantages by explicitly modeling spatial dependence, offering significant advantages by explicitly modeling spatial dependence. The fundamental GS method assumes the unknown source function is a random variable with a known structural form but unknown correlation parameters; identification is achieved by maximizing the likelihood function while strictly preserving this assumed correlation structure [221]. While early applications demonstrated efficacy in simplified systems, such as river pollution with linear attenuation [222], scaling these methods to complex subsurface heterogeneity presented significant challenges. To resolve source identification in two-dimensional non-uniform flow fields, Butera et al. [223] developed an innovative geostatistical methodology for simultaneous reconstruction of contaminant source locations and release histories in 2D non-uniform flow fields, this approach delineates suspect source zones, inverts release functions for all potential sources using GS, and ultimately identifies source locations as subdomains exhibiting maximum cumulative contaminant mass release. Subsequent research sought to enhance robustness through hybrid frameworks, for instance, Gzyl et al. [224] combined pumping tests, particle backtracking, and quasi-linear GS. However, a critical limitation of this multi-step approach is its dependence on approximate source locations for initialization, limiting its utility in scenarios with no prior information. Most recently, capturing high-dimensional geological realism has become the focal point. Park [225] incorporated manifold embedding a non-Euclidean technique into GS inversion. Crucially, this advanced formulation generates broader ensembles of geologically plausible models than traditional Kalman filtering (KF), significantly enhancing spatial stability by ensuring all models maintain hydraulically equivalent responses.

While this GS approach advances geological plausibility and spatial stability, it retains limitations in systematically quantifying comprehensive uncertainty, Bayesian inference provides a theoretical framework for quantifying uncertainties in heterogeneous media modeling by treating parameters as random variables and deriving posterior distributions through prior-data integration. This approach delivers optimal parameter estimates with uncertainty quantification [226]. Early applications to groundwater contaminant source identification were restricted to 1D steady-state flows with single point sources, failing to ensure non-negative concentrations [227]. To overcome these physical and dimensional constraints, subsequent research significantly expanded the framework's applicability. Subsequent advances include: (i) Hierarchical Bayesian methods for inverse advection-diffusion equations (ADE) in heterogeneous media [228]; (ii) Transient source characterization (location, release time, intensity [229]) extended to field scales [230]; (iii) To develop an efficient Bayesian framework for jointly inferring a scalar field and its hyperparameters [186]; (iv) Bayesian hybrid kernel extreme learning machine (BHK-ELM) to capture DNAPLs source–transport–distribution relationships in high-probability-density regions [191].

A key enabler of the previous Bayesian advancements lies in the evolution of stochastic sampling techniques, which are indispensable for numerical posterior inference especially in subsurface systems characterized by high dimensionality and large inherent uncertainty. To accommodate the large uncertainty of subsurface media, initial reliance on techniques like MRE and Generalized Likelihood Uncertainty Estimation (GLUE) gave way to MCMC [231] and filtering techniques [56] as the dominant paradigm. However, Single-chain MCMC methods, including Metropolis-Hastings (MH) [202] and Delay Rejection Adaptive Metropolis (DRAM) [232] frequently exhibited inadequate randomness propagation through complex posterior distributions, leading to convergence uncertainties. This limitation has driven widespread adoption of multi-chain stochastic frameworks, where algorithms such as Differential Evolution Markov Chain (DEMC) [233] and differential evolution adaptive metropolis ($DREAM_{(ZS)}$) [232,234] utilize parallel interacting Markov chains to enhance random exploration efficiency in high-dimensional parameter spaces. The persistent computational intensity of rigorous stochastic sampling demands innovative frameworks that preserve probabilistic integrity while

accelerating convergence. This limitation catalyzes the emergence of multilevel MCMC as an efficient framework, which optimizes the stochastic exploration process through hierarchical resource allocation, algorithmic improvements such as multilevel Monte Carlo (MC) [235]. Similarly inspired by these algorithmic advances, inspired by these algorithmic advances, the multilevel GLUE method employs multi-resolution spatial discretization for distributed models. It reduces high-resolution model runs, easing computational and time burdens for inversion [236].

Besides the above-discussed stochastic sampling techniques, filtering methods also serve as a core implementation pathway for posterior inference and among these, the EnKF has gained prominent application in subsurface systems. The EnKF effectively estimates multiple parameters (e.g., hydraulic conductivity and dispersivity) in large-scale nonlinear inverse problems [237]. However, standard EnKF and related Ensemble Smoothers (ES-MDA) face two critical bottlenecks: their recursive updating requires repeated simulation restarts, incurring prohibitive computational costs, and their applicability is fundamentally constrained by inherent Gaussian assumptions. To specifically overcome this Gaussian limitation and accurately represent complex subsurface heterogeneity. Hybrid strategies extend ES-MDA with multiple-point statistics [238]. These methods have evolved into the direct sampling-ensemble smoother (ES-DS) and ES-MDA-DS to better preserve geological structures while reducing computational costs [138,239], with demonstrated benefits for DNAPLs source characterization [33]. Further addressing the complexity of variable-density flows where density differences between contaminants and groundwater significantly alter transport pathways recent research has targeted the challenge of identifying multiple sources in 3D heterogeneous aquifers. To mitigate the spurious correlations inherent in such high-dimensional inversions, a novel inversion method based on Ensemble Smoothing with clustering-based covariance localization was proposed. This approach not only improved identification accuracy but also provided the critical insight that, contrary to common simplifications, variable-density dynamics can actually enhance the data information content available for inversion [240]. Inspired by ES-MDA, hybrid approaches utilizing an Improved Butterfly Optimization Algorithm (IBOA) with ES-MDA have addressed computational efficiency, accuracy, posterior distributions, and noise robustness for complex contaminant transport problems [193]. Similarly, the UKS-MDA was developed to enhance the identification of hydraulic conductivity and contamination sources [36]. Representing a further advancement in handling strong non-linearity, Zhang et al. [241] proposed the ILUES. This framework is distinguished by its capability to resolve both unimodal and multimodal problems, successfully achieving the simultaneous identification of contaminant source characteristics and model parameters. In parallel, Filtering techniques (e.g., KF [202], particle filter (PF) [56]) offer a robust alternative for fundamentally non-linear and non-Gaussian systems where Kalman-based methods often struggle. However, despite their theoretical advantages, PF face significant challenges regarding computational efficiency and generalizability. While innovations such as ensemble learning frameworks and swarm evolutionary algorithms attempt to balance robustness with feasibility, these trade-offs remain a persistent bottleneck [242]. To specifically resolve these efficiency constraints, diverse intelligent variants have been developed to enhance the traditional filtering framework. Notably, approaches such as the Intelligent Particle Filter (IPF) [243] and Intelligence-Enhanced Particle Filter (IEPF) [56] have successfully achieved the simultaneous identification of source characteristics and model parameters with significantly reduced computational costs.

In conclusion, stochastic frameworks are indispensable for quantifying uncertainty in heterogeneous aquifers, yet they necessitate a strategic balance between computational efficiency and the rigor of posterior inference. The evolution from traditional geostatistical methods to advanced Bayesian sampling and ensemble filtering reflects a continuous effort to resolve the tension between high-dimensional geological realism and algorithmic cost. To facilitate method selection under varying hydrogeological complexities, Table 4 provides a critical classification and performance assessment of these stochastic paradigms, summarizing their core mechanisms, comparative advantages, limitations, and optimal application scenarios.

4.3. Physics-informed inversion approaches

The governing equations for groundwater contaminant transport, typically the advection-diffusion-reaction equation (ADRE), provide the mathematical foundation for inverse source identification. To relate observed contaminant concentrations to unknown source parameters such as location and release history, these equations are reformulated as integral equations or ill-posed linear inverse operators. However, inferring source characteristics from limited concentration measurements constitutes an inherently ill-posed inverse problem. This condition frequently leads to solution non-uniqueness. To overcome such constraints, PINN have emerged as a mesh-free DL framework [244]. By leveraging automatic differentiation instead of predefined basis functions, PINN directly embed governing equations into neural network loss functions. This approach solves both forward problems and inverse problems without domain discretization [245].

Despite these advances, Standard PINN formulations struggle to resolve the inherent nonlinear dynamics and coupled boundary conditions in such systems [246,247]. To address these limitations, recent work has progressed in three directions focused on enhancing model representation and computational efficiency. First, parameter field representation has been enhanced to handle heterogeneity. Frameworks like physics-informed ML with conditional Karhunen-Loève expansion (PICKLE) and PI-CKL-NN effectively handle heterogeneous parameters and nonlinear coupling [248], while power series-expanded PINN address variable-coefficient fractional ADE in multidimensional domains [249]. Complementing this, the Theory-guided Neural Network constrained with geostatistical information introduces a dual-network architecture that simultaneously approximates random model parameters and solution fields. By honoring prior geostatistical information, this approach achieves robust identification even under conditions of sparse spatial measurements or imprecise prior statistics [250]. Second, inverse problem capabilities have been strengthened via hybrid and efficient strategies. Zhan et al. [251] augment PINN with LASSO regression and sequential optimization to address sparse data scenarios, enabling robust PDE identification, while Hou et al. [71] develop hybrid PINN incorporating locally adaptive residual networks and dynamic sampling strategies, which simultaneously learn diffusion coefficients and dynamically correct transport models. Inverse problem capabilities have been strengthened; specifically, incorporating pre-training strategies and domain decomposition methods has proven effective in significantly enhancing training efficiency and convergence stability, even for nonlinear adsorption models [252]. To further resolve the computational bottleneck of modeling coupled flow and reactive transport, the Multi-Physics Generative Pre-trained PINN (MP-GPT-PINN) introduces a meta-learning framework with a parallel dual-network architecture. By building a compact library of solutions to avoid costly retraining, this approach accelerates online predictions by four orders of magnitude, making large-scale parameter inversion feasible [253]. Third, noise resistance and interpretability have been enhanced. Beyond RBF-activated PINN [254] and Bayesian uncertainty quantification, improving robustness under highly noisy and nonlinear conditions [255], framework interpretability in sparse monitoring scenarios has been advanced by explicitly embedding hydraulic and concentration gradients as physical features; this approach has been validated in field-based industrial cases to maintain robust source localization against irregular sampling [256].

Beyond algorithmic improvements, a critical bottleneck remains in the physical fidelity of the models, as standard PINN applications often simplify the governing physics to the advection-diffusion equation while

Table 4
Critical classification and performance assessment of stochastic-based approaches for organic pollutant source identification.

Category	Model	Core Mechanism	Advantages	Limitations	Optimal Scenario
Geostatistical & Probabilistic	Backward Probability, Quasi-linear GS, Manifold Embedding GS	Spatial Dependence Modeling: Infers sources by maximizing likelihood while strictly preserving spatial correlation structures.	Manifold methods generate hydraulically equivalent and geologically plausible models	Often requires approximate source locations for initialization.	Scenarios with prior knowledge of source zones; Complex geological fields requiring strict structural preservation.
Bayesian MCMC Sampling	MH, DRAM, DREAM, ML-MCMC	Uses Markov chains to rigorously explore the full parameter space and derive posterior distributions via prior-data integration.	The theoretical benchmark for quantifying parameter and predictive uncertainty; Multi-chain methods effectively avoid local optima in high-dimensional spaces.	The computational cost of complex 3D models is prohibitively high; Single-chain methods struggle with complex, multimodal distributions.	Studies dependent on comprehensive uncertainty characterization; Problems with high inherent uncertainty but manageable model runtimes.
Models suitable for Gaussian assumptions	EnKF, ES-MDA, ES-DS	Updates an ensemble of models by minimizing variance based on the linear covariance matrix between parameters and data.	Handles large-scale inversion orders of magnitude faster than MCMC; Naturally suited for high-performance parallel computing.	Performance degrades significantly in non-Gaussian fields; High degrees of freedom lead to false correlations.	Large-scale, high-dimensional industrial applications; Problems where the Gaussian assumption is approximately valid.
Models suitable for Non-Gaussian	ILUES, IEPF	Employs local ensemble approximation or particle resampling/weighting to resolve complex non-Gaussian posterior distributions.	Multimodal Resolution: ILUES effectively resolves multimodal posterior distributions; Handles highly non-linear kinetics better than EnKF.	Standard PF is computationally heavier than EnKF; ILUES requires careful tuning of local domains.	Strongly heterogeneous aquifers; Scenarios with multiple, distinct potential source configuration.

neglecting reactive decay terms [257]. To surmount this barrier and address the complexity of reactive transport, the PH-PINN framework successfully integrates physical constraints with the PHREEQC geochemical module. Validated using field monitoring data from the Datong Basin, this approach explicitly captures multi-pathway reaction networks (e.g., iron-sulfur-carbon-nitrogen cycles) driving arsenic mobilization, significantly outperforming traditional PINNs by capturing non-stationary dynamics and reducing RMSE by over 50% [258]. Taking this a step further to target structural uncertainty where the governing mechanism itself is ambiguous, the Theory-guided U-net (TgU-net) models three potential equilibrium sorption isotherm types (i.e., Linear, Freundlich, and Langmuir) through a single surrogate. This framework demonstrates the feasibility of learning a cluster of equations, though it exhibits sensitivity to data availability and observational noise [259]. For scenarios where the functional form is entirely unknown, the Finite Volume Neural Network merges numerical methods with deep learning to learn arbitrary constitutive relations. Applied to diffusion-sorption problems, this approach flexibly models sorption isotherms without being restricted to predefined parametric models, outperforming calibrated PDE-based models in generalization across varying boundary conditions [260].

Finally, addressing optimization and uncertainty challenges, the recently proposed Time-Space Bayesian PINN (TSBPINN) offers a significant methodological advance. Unlike traditional PINN that often suffer from loss imbalance and convergence to local minima, TSBPINN employs a two-stage Bayesian strategy that facilitates the stable propagation of physical constraints. This design has demonstrated superior accuracy in locating pollutant sources and estimating initial concentrations, even under high sensor noise levels (up to 25%), while providing credible intervals that appropriately reflect increased uncertainty [261]. However, the current framework is limited to 1D synthetic scenarios with idealized boundaries, and its reliance on Gaussian assumptions may fail to capture heavy-tailed errors inherent in field data. Consequently, future research must prioritize extending such methods to 3D heterogeneous aquifers and addressing the irregular, sparse monitoring networks typical of real-world sites.

In conclusion, while deep learning accelerates inversion by bypassing repetitive simulations, its "black-box" nature and data dependency often compromise physical consistency. Physics-informed learning attempts to mitigate this but introduces new optimization challenges. Table 5 critically summarizes different PINN frameworks, clarifying their trade-offs in computational efficiency, physical fidelity, and data requirements to guide model selection in complex groundwater-related scenarios.

4.4. Hybrid methods

Traditional single-method approaches prove inadequate for

Table 5
Comparative of physics-informed inversion frameworks.

Problem Focus	Specific Frameworks	Advantages	Limitations	Optimal Scenario
Heterogeneity Representation	PICKLE, PI-CKL-NN, Geostat-Constrained TgNN, Power-Series PINN	Effectively handles heterogeneous parameters coupled with nonlinear dynamics; Robust identification with sparse measurements by honoring prior statistic	Computational cost explodes with the number of stochastic dimensions	High-dimensional heterogeneous aquifers; Scenarios with reliable geological prior statistics
Computational Efficiency	MP-GPT-PINN, Hybrid PINN	MP-GPT-PINN accelerates online predictions by 4 orders of magnitude; Domain decomposition stabilizes training for stiff nonlinear adsorption.	Meta-learning requires complex offline pre-training phases; Accuracy depends on the pre-built library's coverage.	Large-scale inversion requiring real-time results; Stiff nonlinear problems
Reactive Fidelity	PH-PINN, TgU-net, Finite Volume NN	Captures multi-pathway cycles, reducing RMSE by $>$ 50%; FVNN models arbitrary relations without predefined forms.	Coupling stiff geochemical solvers with NN backpropagation is prone to instability; Learning constitutive laws is highly sensitive to noise.	Contaminants driven by complex multi-component reactions; Unknown sorption mechanisms
Uncertainty & Robustness	TSBPINN, Gradient-PINN, RBF-PINN	TSBPINN remains accurate under 25% sensor noise; Gradient embedding improves physical consistency in sparse networks.	Bayesian frameworks (TSBPINN) currently limited to 1D synthetic scenarios; Reliance on Gaussian assumptions may fail for heavy-tailed field data.	Field sites with high measurement noise or irregular sampling.

inverting organic contaminant sources in heterogeneous aquifers. Optimization algorithms exhibit excessive initial-value dependency and frequent convergence to local optima. Stochastic methods require prohibitively high computational costs to achieve statistical significance. PINN constrained by homogeneity assumptions, fail to resolve highly nonlinear dynamics. These collective limitations induce substantial errors in reconstructing contaminant source parameters (locations, release histories, and intensities), which inhibits the accurate inversion of organic contaminant sources in heterogeneous aquifers.

Currently, the methods for identifying contamination sources mainly fall into three distinct categories: simulation optimization, Bayesian inference, and DA. Each method has its own advantages and disadvantages under specific site conditions. To navigate these challenges and establish a selection baseline, a recent comparative study systematically evaluated three representative algorithms—IBOA for deterministic simulation-optimization, the ES-MDA for DA, and the DiffeRential Evolution Adaptive Metropolis with a Snooker Update and Sampling from a Past Archive ($DREAM_{(ZS)}$) for Bayesian inference- across scenarios involving conservative solute transport and LNAPLs transport with biodegradation, This critical evaluation reveals distinct trade-offs between computational efficiency and posterior reliability. Results indicate that while ES-MDA achieves the shortest elapsed time, making it suitable for rapid preliminary assessments, it exhibits significant deviations in source location estimation under high uncertainty. Conversely, $DREAM_{(ZS)}$ demonstrates superior accuracy and robustness against observational noise, successfully capturing complex posterior distributions. Crucially, the study highlights a fundamental limitation of deterministic optimization: unlike ES-MDA and $DREAM_{(ZS)}$, IBOA fails to quantify uncertainty, rendering it unsuitable for high-risk decision-making in heterogeneous aquifers despite its algorithmic simplicity. Consequently, a strategic framework suggests employing ES-MDA for efficiency-driven tasks and $DREAM_{(ZS)}$ when maximizing posterior coverage and accuracy is paramount [193].

To transcend these individual limitations and leverage synergistic benefits, hybrid methodologies have evolved along three distinct technical pathways. The first pathway involves fusing stochastic and optimization methods, such as the integration of MH algorithms with KF [202], particle swarm optimization-MC hybrids [262],the coupling of EnKF and ant colony optimization (ACO) to reduce parameter uncertainty [263]. More recent innovations include the IPF, which approximates distributions via adaptive particle weighting but remains prone to local optima. Improvements such as IBOA-directed update [243], and IEPF with Bayesian-guided GA enhance convergence and robustness [56]. The second pathway emphasizes the integration of DL with traditional modeling framework. More recent studies have advanced physics-informed architectures, such as PINN with locally adaptive residual learning, which reduce parameter estimation errors to below 1% [71], and transformer encoder–global average pooling, which compresses features via attention mechanisms while noise-injection augmentation with SHAP-based interpretability enhances robustness and transparency [212]. In parallel, deep generative models have been coupled with groundwater simulators; for instance, conditional GAN integrated with MODFLOW–MT3DMS enable multi-source contaminant identification [64]. The third pathway centers on global-local refinement through combinatorial frameworks, such as the fusion of EnKF and IBOA [41], ASACO-EnKF for simultaneous source–parameter inversion [172]. Subsequent advances include covariance matrix adaptation evolution strategy accelerates global search but introduces Bayesian deviations [264]. To resolve these limitations, BHK-ELM embeds sensitivity-related dynamic swarm intelligence within a PSO framework, resolving equivalence issues and premature convergence in multimodal searches while enhancing inversion accuracy and computational efficiency [191].

In conclusion, while single algorithms have specific merits, complex hydrogeological scenarios often necessitate hybrid architectures to synergize computational speed with physical rigor. However, these combinations increase structural complexity and tuning demands. Table 6 provides a comprehensive taxonomy of these core data-driven algorithms, summarizing their inductive mechanisms, applicable data types, and key characteristics. This classification serves as a foundational reference for constructing tailored inversion frameworks.

5. Limitations of ML-based organic pollution identification

5.1. Algorithmic limitations in computational efficiency and interpretability

Identifying OPS in HA faces two core limitations: computational inefficiency and limited model interpretability. Computational inefficiency arises from the high-dimensional, nonlinear nature of pollution datasets: environmental pollution is typically complex, and it can include hundreds of organic contaminants potentially coexisting within a single polluted area [266], and extracting robust features often requires repeated iterative training [45]. Compounding this challenge, modeling heterogeneous aquifer structures via stochastic generation techniques requires fusing geophysical data, borehole data, and hydrological observation data. Surrogate-based inversions can reduce runtime but typically fail to provide reliable parameter estimates under complex structures and boundary conditions [267]. Moreover, many frameworks rely on stationarity assumptions that cannot capture complex spatial patterns and non-stationary behaviors [268], while achieving different inversion objectives typically requires distinct ML models. The process of selecting and adjusting these models is time-consuming, and the poor adaptability of the algorithms and models reduces the efficiency gains from ML and DL techniques [269].

ML models are often regarded as “black boxes” because of their limited interpretability and lack of transparent decision-making processes (Fig. 5). This hinders the identification of OPS, as evaluating the importance of individual features and their interactions becomes difficult. For example, the nonlinear relationship between adsorption capacity and multiple influencing factors adds complexity to the analysis [270]. Furthermore, identifying pollution sources is hindered because pollutant distribution depends on various factors and their interactions within the source-flow-sink process. Several interpretability tools have been introduced but remain limited. Geographical detectors can reveal coupling effects but generally capture only simple additive interactions [271]. Similarly, post-hoc interpretation methods such as SHAP and local interpretable model-agnostic explanations (LIME) are mainly designed for straightforward classification or regression tasks and provide limited insights into complex hydrogeological processes [272]. Recent advancements have successfully repurposed these tools from passive explanation to active sampling guidance. To overcome data sparsity, Wang et al. [273] proposed a SHAP-Guided Two-stage Sampling method, which successfully obtains a more precise and robust hydraulic conductivity field compared to conventional random sampling. Furthermore, addressing the challenge of sparse spatiotemporal data, Zhang et al. [256] demonstrated that this approach enables the accurate reconstruction of contaminant source location, release timing, and intensity. Crucially, the spatiotemporal behavior of OP follows physical laws. Most ML algorithms lack these physical constraints, often leading to identified “interaction effects” that are spurious correlations. Some algorithms, like PINN and TgFCNN, incorporate physical equations or constraints into their training or loss functions. But, currently physics-informed ML models often face a dilemma: they are either overly constrained by physics or too flexible, potentially learning non-physical relationships [260].

5.2. Data sparsity constraints in monitoring networks

Groundwater contamination monitoring networks serve as essential tools for tracking pollutant dispersion, groundwater level dynamics, and toxicological-exposure risk mitigation. Machine learning offers a data-

Table 6
Classification and characteristic analysis of data-driven core algorithms [265].

Inductive Core	Data Task Type	Model	Features
Feature characterization	Dominated by image data, evolving towards multi-modal data fusion	PCA	Dimensionality reduction via eigenvalue decomposition.
		CAE	More suitable for high-dimensional problems than PCA.
		CVAE	More applicable to the characterization of non-Gaussian geological structures than CAE.
		AAE	Suitable for the characterization of non-Gaussian geological structures.
Forward prediction	Image data and sequential data	CNN	Establishes mapping relationships among high-dimensional data.
		ResNet	Achieves higher prediction accuracy than CNN.
		U-Net	Network architecture enhances prediction accuracy.
		LSTM	Applicable to the prediction of sequential data.
		GRU	Shorter training time compared with LSTM.
		Tansformer	High prediction accuracy and computational efficiency.
Inversion simulation	Integration of image data, sequential data and optimization problems	Optimization algorithm	Simple implementation, but computationally inefficient.
		Stochastic algorithm	Effective for inversion in Gaussian fields.
		DA (DL)	Superior for inversion in non-Gaussian or complex heterogeneous fields.
Mechanism exploration	Image data and sequential data	PINN	Improves the physical interpretability and accuracy of meshes.
		TGNN	Overcomes the high-dimensional sampling challenge of PINN.

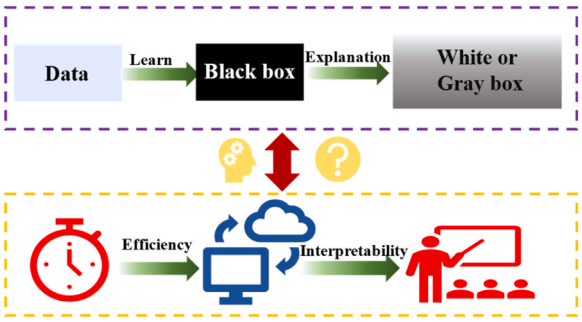

Fig. 5. Dynamic interplay between computational efficiency and model interpretability.

driven approach to optimizing these networks, with promising potential for broad application. However, machine learning has the following limitations: a strong dependence on the quality and scale of monitoring data, limited adaptability to complex contamination scenarios, and insufficient consideration of dynamic hydrogeological processes.

The machine learning algorithms relies heavily on prior knowledge of contamination, high-quality monitoring data, and robust stochastic simulations. Most machine learning models require known or assumed contamination sources and tend to be constrained to areas with minimal observational coverage [126]. In regions where pollution sources are unknown, such as cases involving concealed or illegal discharges, or where historical data are lacking, the resulting optimized monitoring networks are subject to significant uncertainty. Existing monitoring wells are often hydrologically unrepresentative, for instance, many wells are constructed by farmers primarily for agricultural use, leading to spatial clustering in farmland areas [274]. Moreover, critical structural details of these wells are frequently missing, making it difficult to capture vertical heterogeneity in the saturated zone. This limitation may cause models to overlook the vertical distribution of contaminants. In many studies, observed contamination data are simplified into binary classifications [275]. Expanding to multi-class or continuous classifications (e.g., concentration gradients) would require substantially larger datasets. However, sample size limitations often preclude such approaches. Additionally, biases or noise in monitoring data can propagate through stochastic simulations, thereby constraining the predictive accuracy of machine learning-based optimization frameworks.

Moreover, to the best of our knowledge, current studies on hydrological monitoring network optimization are mainly designed for single contaminants (e.g., heavy metals) and specific site conditions (e.g., landfill settings). However, a monitoring variable also influence the evolution of others through coupled physico-chemical processes in real-world hydrological processes [276]. Barcellos and Souza [277] analyzed 6328 water quality observations collected between 1971 and 2021 from 35 stations across 27 watersheds in Brazil, covering 60 water quality parameters. The results indicate that the inclusion of a greater number of monitoring parameters enables the identification of more complex and informative association patterns, offering valuable insights for the optimization of water quality monitoring strategies. When applied to regions affected by multiple contaminants (e.g., heavy metals and organic compounds) and diverse pollution sources (e.g., industrial facilities and agricultural land), the transport behavior of pollutants becomes substantially more complex [1]. This complexity necessitates the support of multiple machine learning models and shifts the focus away from a single optimization objective, thereby increasing the overall complexity of machine learning applications in monitoring network design [278].

At present, most existing studies on monitoring network optimization assume of steady-state contamination scenarios and are typically conducted under fixed boundary conditions in terms of temporal sequences and spatial domains. However, changes in precipitation patterns driven by climate change are expected to intensify both the infiltration depth and spatial extent of surface runoff, thereby altering pollutant transport dynamics [279]. Intra-annual seasonal variability directly affects contaminant migration pathways and the timing of concentration peaks, while interannual hydrological cycles can shift aquifer hydraulic gradients, leading to year-to-year variations in plume dispersion rates [280]. How to effectively incorporate such dynamic environmental and boundary conditions into the design of optimized monitoring networks remains an open and pressing research question.

5.3. Uncertainty analysis in pollution source identification and assessment

Identifying OPS in complex HA is inherently uncertain due to data scarcity and aquifer complexity. Data scarcity and observational errors contribute to data uncertainty, while conceptual simplifications of physical processes lead to structural uncertainty in models. Additionally, temporal variability in pollution sources introduces uncertainty in source term characterization [281]. These uncertainties accumulate and propagate through the system, consequently causing significant errors or failure in contaminant source inversion (Fig. 6).

Sparse monitoring data may miss localized high-pollution areas, while short-term sampling often fails to capture the periodic evolution of OP. Additionally, collected data can contain outliers, noise, and other biases due to natural environmental interference [282]. When applying algorithms such as CNN, LSTM, Transformer, or GNN to fuse such

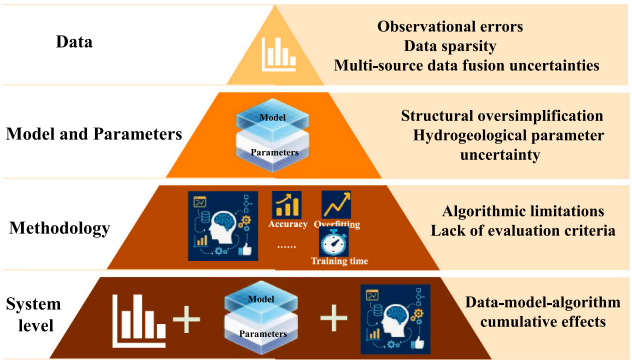

Fig. 6. Multi-dimensional uncertainty hierarchical framework.

multi-source heterogeneous data, critical low-dimensional features may be underrepresented. This can lead to overfitting and amplify result uncertainties. Furthermore, a lack of unified evaluation standards makes it difficult to assess result reliability. To establish a rigorous comparison baseline, a standardized set of validation metrics is urgently needed. This framework should extend beyond simple error statistics to include uncertainty reduction for quantifying information gain, posterior coverage for verifying the statistical validity of confidence intervals, noise resistance for testing model robustness under realistic measurement errors, and convergence characteristics for evaluating computational efficiency.

To rigorously address these uncertainties and reconstruct contaminant plumes from such limited data, stochastic inversion frameworks are essential [283]. However, traditional approaches face distinct challenges. Both MCMC and PF methods encounter significant efficiency bottlenecks in high-dimensional systems. If the proposal distribution is not well-tuned, MCMC often exhibits high autocorrelation, significantly reducing sample generation efficiency [232], while PF tend to suffer from rapid degeneracy and collapse. Consequently, their efficiency is often inferior to Gaussian-based approaches in computationally constrained systems [154]. Among these Gaussian-based approaches, the EnKF, a Monte Carlo implementation of the KF successfully overcomes difficulties associated with covariance estimation in nonlinear state transfer functions. However, its standard formulation requires modification to correctly characterize non-Gaussian parameters [284], and its inherent reliance on Gaussian assumptions limits its robustness [155]. Although recent studies propose using DL to construct nonlinear mappings as a substitute for the Kalman gain matrix to overcome Gaussian limitations [285,286], this introduces a substantial computational burden. The selection of DL structures is often subjective, and the hyperparameter tuning process is extremely time-consuming, rendering it inefficient for practical applications [211]. In contrast, while the batch-form ES-MDA achieves higher computational efficiency [287], it remains limited in non-Gaussian scenarios as a Kalman-based method. While it offers some tolerance to mild violations of Gaussian assumptions, this can conversely introduce sampling errors [288]. To address the vulnerability of EnKF and ES in non-Gaussian settings, ILUES is capable of handling complex multi-modal distributions without relying on auxiliary clustering algorithms. Crucially, the algorithm effectively quantifies parameter uncertainty in complex hydrological models, irrespective of the presence of multi-modal distributions [169,241]. However, focusing solely on parameter estimation within a single model framework is insufficient, such approaches still fail to resolve complex system mechanisms. Neglecting nonlinear interactions induces structural uncertainty that parameter calibration cannot fully compensate [289]. Moreover, reliance on single models is unreliable for pollution source identification. For OP, parametric uncertainty is particularly severe, larger uncertainty in dissolution rate relative to biodegradation rate [290]. Consequently, reliance on single conceptual models proves unreliable for robust pollution source identification.

6. Prospects

ML has demonstrably advanced OPSI in HA, yet critical bottlenecks persist. These are characterized by inadequate integration of multi-source data, the black-box nature of ML models with limited explainability, and computational inefficiency that demands accelerated algorithms. Such intertwined constraints severely constrain field applicability. To overcome these challenges, organic contaminant source identification must be reconceptualized toward individual-level precision, holistic systems integration, and intelligent computational frameworks.

(1) Enhancing Data Support and Optimizing Monitoring Networks: A core challenge in current inversion practices is insufficient data support. Key limitations include low integration of multi-source data (geological, geophysical, hydrological, and chemical), constrained spatiotemporal representativeness of monitoring networks, and a scarcity of dynamic data. To mitigate these issues, recent research has introduced a multivariate network design framework that leverages joint entropy to quantify the uncertainty of multicomponent responses. Validated under high-noise conditions, this deep learning-accelerated approach enables the robust fusion of heterogeneous datasets, including hydrological measurements and geophysical survey lines, thereby enhancing parameter estimation even when measurement accuracy is limited [137]. However, despite such algorithmic advancements, complete in-situ groundwater contamination datasets are also scarce in many regions, which inevitably limits the performance of machine learning models for OPSI. Notably, the idealized assumptions of synthetic datasets differ inherently from real-world field data characteristics (e.g., unknown parameters, sparse monitoring, and data noise). The acquisition of in-situ monitoring data at actual sites faces multiple practical constraints: most contaminated sites lack dense monitoring well networks, and installing additional wells is restricted by both objective site conditions and regulatory approval requirements, significantly raising the difficulty of on-site sampling. Meanwhile, high-resolution in-situ monitoring under minimal aquifer disturbance is technically challenging in practice. Owing to the scarcity and inaccessibility of in-situ data, synthetic datasets have become an essential support for OPSI model development and validation. To narrow the gap between synthetic and real-world data, existing studies have introduced noise into simulated data to mimic on-site monitoring errors, thus minimizing model performance deviations from data scenario discrepancies.

(2) ML Advances and Model Representation: Progress in ML offers promising pathways. Embedding physical constraints from governing equations within hybrid optimization algorithms can restrict solution spaces and enhance realism. Meanwhile, deep generative models (e.g., GAN, VAE, latent diffusion) show strong potential for capturing heterogeneous geological structures and simulating multi-process interactions, yet they must be integrated into physics-informed frameworks to avoid physically irrelevant outputs. Finally, ensemble data assimilation techniques, widely applied in hydrogeology, require rigorous optimization. Careful hyperparameter tuning is essential to sustain their performance under strong nonlinearity and to enhance their feature extraction capacity.

(3) Interpretable High-Efficiency ML: The inherent "black-box" nature of ML models critically impedes interpretability of inversion outcomes, particularly compromising reliability in organic contaminant source identification involving complex biogeochemical processes. To address this, future research must develop interpretable, high-efficiency ML frameworks. Physics-informed DL (PIDL) should embed contaminant-specific mechanisms (e.g., Monod kinetics) to constrain inversion outputs with physicochemical principles, avoiding purely data-driven misinterpretations. Additionally, interpretable ML models must reveal causal relationships, such as electron acceptor gradients and microbial community dynamics, rather than relying on mere statistical correlations [63]. When applying post hoc tools like SHAP and LIME,

feature contributions must be contextualized within spatiotemporal evolution patterns of organic contaminants (e.g., plume zoning) to mitigate errors from overlooked biochemical couplings [290]. Implementing these interpretability frameworks demands substantial computational resources for large-scale or high-dimensional organic contaminant scenarios. Looking ahead, emerging paradigms like quantum computing present a transformative potential for specific bottlenecks. Theoretical studies indicate that quantum computing-integrated optimization algorithms can achieve faster and more accurate contaminant source identification [291], while quantum-enhanced models demonstrate superior stability and cost-efficiency over baseline models [292]. Notably, for highly heterogeneous media such as fractured systems, quantum algorithms offer significant speedups in solving linear systems and provide "free" uncertainty quantification by reducing the computational cost of ensemble simulations to that of a single realization. However, the direct application of quantum computing currently faces significant hurdles due to hardware limitations and the scarcity of specialized algorithms. Consequently, its utility remains largely restricted to theoretical verification, with fault-tolerant implementations for practical hydrogeological problems projected as a decade-long objective rather than an immediate solution [293].

(4) Uncertainty Quantification Analysis : The inversion of organic contaminant sources is challenged by coupled uncertainties stemming from geological heterogeneity, ambiguity in source localization, sparse and noisy measurements, and incomplete aquifer characterization—factors that collectively undermine the reliability of inversion outcomes. To address this, future research should focus on three interconnected innovations in intelligent uncertainty quantification. First, advancing stochastic inversion within Bayesian frameworks, supported by efficient sampling algorithms, will enable robust resolution of parameter-structural uncertainties. Second, leveraging ensemble machine learning techniques to systematically compare uncertainty propagation pathways between independent and coupled inversion strategies through multi-stage data analysis can inform the selection of optimal inversion frameworks. Together, these advances will reduce systematic errors in contaminant source inversion, substantially enhancing the reliability of pollution localization.

7. Conclusions

Deep integration of ML algorithms into OPSI within HA represents the most accurate and efficient approach currently available. This superiority stems from ML's intrinsic capability to approximate complex non-linear mapping functions in high-dimensional spaces, effectively capturing the intricate spatial connectivity and coupled reaction kinetics that are often oversimplified by traditional approaches. However, practical deployment faces significant challenges, including the high variability of organic contaminants, non-Gaussian aquifer heterogeneity, data scarcity, the "black-box" nature of ML models, and high computational complexity. Based on this review, we offer the following practical recommendations to bridge the gap between academic research and industrial application.

(1) The synthesis confirms that monitoring network optimization grounded in information entropy theory is fundamental for maintaining high robustness under high-noise conditions. By maximizing information content, this theoretical framework serves as the critical enabler for effective multi-source data fusion, ensuring reliable identification accuracy even when data quality is compromised.

(2) Regarding surrogate modeling, a critical trade-off is established. While image-to-image deep learning algorithms demonstrate exceptional precision in field reconstruction, our analysis reveals that shallow machine learning algorithms, when hybridized with advanced optimization methods, can achieve comparable performance in source parameter identification with significantly reduced computational costs.

(3) In the realm of stochastic inversion, Bayesian ensemble frameworks have been identified as having the highest applicability for characterizing high-dimensional non-Gaussian heterogeneity. Unlike traditional methods constrained by linearity, these integrated frameworks effectively manage the multimodal posterior distributions inherent in organic pollutant transport, balancing uncertainty quantification with computational feasibility.

(4) While PINN have been widely applied to transport equations, their integration with complex geochemical reaction kinetics remains rare. Although PINN-based surrogates demonstrate superior physical consistency over pure data-driven deep learning algorithms in theory, their practical deployment in high-dimensional, strongly heterogeneous media remains hindered by optimization convergence issues, representing a key frontier for future research.

(5) Regarding emerging paradigms, while quantum computing offers transformative potential for solving large-scale linear systems, its immediate application is constrained by hardware limitations and a scarcity of specialized algorithms. Consequently, practical deployment in the near term remains reliant on optimizing surrogate acceleration within classical high-performance computing architectures, coupled with advancements in parallel computing strategies to handle the computational load of complex surrogate models.

In summary, this study provides a structured scientific basis for model selection by mapping specific algorithmic strengths directly to distinct hydrogeological challenges, ranging from characterizing non-Gaussian heterogeneity to resolving complex reactive transport. By delineating these boundaries, the review effectively bridges the gap between theoretical algorithms and the practical demand for source identification, ultimately facilitating cost-effective remediation decisions and sustainable groundwater management.

Abbreviation

Abbreviation	Meaning
OPS	Organic Pollutant Sources
OP	Organic Pollutants
OPSI	Organic Pollutant Source Identification
PFAS	Polyfluoroalkyl Substances
PCBs	Polychlorinated Biphenyls
PAEs	Phthalates
HA	Heterogeneous Aquifers
PCA	Principal Component Analysis
3D	Three-Dimensional
CNN	Convolutional Neural Networks
TPH	Total Petroleum Hydrocarbon
BTEX	Benzene Series
PAHs	Polycyclic Aromatic Hydrocarbons
NAPLs	Non-Aqueous Phase Liquids
LNAPLs	Light Non-Aqueous Phase Liquids
DNAPLs	Dense Non-Aqueous Phase Liquids
ML	Machine Learning
DL	Deep Learning
ANN	Artificial Neural Networks
SVM	Support Vector Machine
GAN	Generative Adversarial Networks
SVR	Support Vector Regression
PINN	Physics-Informed Neural Networks
PIDL	Physics-Informed Deep Learning
DA	Data Assimilation
EnKF	Ensemble Kalman Filters
ILUES	Iterative Local Updating Ensemble Smoother
DRDCN	Deep Residual Dense Convolutional Networks
DBNN	Deep Belief Neural Networks
AKSM	Adaptive Kriging Surrogate Model
RDCNN	Residual Dense Convolutional Networks
DCNN	Deep Convolutional Neural Network
CVAE	Convolutional Variational Autoencoder
VAE	Variational Autoencoders
AI	Artificial Intelligence
aGNN	Attention-Based Graph Neural Network
GA	Genetic Algorithm
BME	Bayesian Maximum Entropy
KELM	Kernel Extreme Learning Machine

(*continued on next page*)

(*continued*)

Abbreviation	Meaning
DCGAN	Deep Convolutional Generative Adversarial Network
MCMC	Markov Chain Monte Carlo
MIMR	Maximum Information Minimum Redundancy
MC	Monte Carlo
ASACO	Adaptive Step Length Ant Colony Optimization
KELM	Kernel Extreme Learning Machine
DE	Differential Evolution
SO	Simulation Optimization
PSO	Particle Swarm Optimization
GWO	Grey Wolf Optimizer
SA	Simulated Annealing
MRE	Minimum Relative Entropy
MVO	Multiverse Optimization
EAS-LSTM	Entity-Aware Sequential Long Short-Term Memory
Bi-GAN	Bidirectional Generative Adversarial Network
RWPTM	Random Walk Particle Tracking Method
1D	One-Dimensional
2D	Two-Dimensional
GS	Geostatistical
ES	Ensemble Smoother
KF	Kalman Filtering
ADE	Advection-Diffusion Equations
BMARS	Bagging Multivariate Adaptive Regression Splines
BHK-ELM	Bayesian Hybrid Kernel Extreme Learning Machine
GLUE	Generalized Likelihood Uncertainty Estimation
MH	Metropolis-Hastings
DEMC	Differential Evolution Markov Chain
DREAM	Differential Evolution Adaptive Metropolis
ADRE	Advection-Diffusion-Reaction Equation
ES-MDA	Ensemble Smoother with Multiple Data Assimilation
PF	Particle Filter
NS-EnKF	Normal-Score Ensemble Kalman Filter
IES	Iterative Ensemble Smoother
ES	Ensemble Smoother
ES-LM	Ensemble Smoother Based On Levenberg-Marquardt
IBOA	Improved Butterfly Optimization Algorithm
UKS-MDA	Unscented Kalman Smoother With Multiple Data Assimilation
ResNet	Deep Residual Network
IEPF	Intelligence-Enhanced Particle Filter
PDE	Partial Differential Equations
ODE	Ordinary Differential Equations
TgU-net	Theory-guided U-net
MP-GPT-PINN	Multi-Physics Generative Pre-trained PINN
PICKLE	Physics-Informed Machine Learning with Conditional Karhunen-LoÈVe Expansion
RBF	Radial Basis Functions
ACO	Ant Colony Optimization
IPF	Intelligent Particle Filter
TE	Transformer Encoder
GAP	Global Average Pooling
MPS	Multipoint Statistics
EnPAT	Ensemble Pattern
SGAN	Spatial Generative Adversarial Networks
StyleGAN	Style-Based Generative Adversarial Networks
ERT	Electrical Resistivity Tomography
SP	Self-Potential
HT	Hydraulic Tomography
LSTM	Long Short-Term Memory
TSBPINN	Time-Space Bayesian PINN
NS-ES	Normal Score-Ensemble Smoother
ES-DS	Direct Sampling-Ensemble Smoother
GeoSinGAN	Geological Single-Image Generative Adversarial Networks
DOCRN	Deep Octave Convolutional Residual Dense Networks
MLP	Multilayer Perceptron
RNN	Recurrent Neural Network
FNN	Fully Connected Neural Network
TPOT	Tree-Based Pipeline Optimization Tool
XGBoost	Extreme Gradient Boosting
RF	Random Forest
ETR	Extra Trees Regressor
ARNN	Autoregressive Neural Networks
TgFCNN	Theory-guided Fully Convolutional Neural Network
SHAP	SHapley Additive ExPlanations
LIME	Local Interpretable Model-agnostic Explanations
DFML	Direct Forward Machine Learning
OHML	One-Hot Machine Learning
RML	Random Maximum Likelihood
OANW	Optimized AR-Net-WL

(*continued on next column*)

(*continued*)

Nomenclature

Symbol	Parameter	Unit
P_β	Fluid pressure in phase β	$ML^{-1}T^{-2}$
ρ_β	Density of phase β	ML^{-3}
μ_β	Viscosity of phase β	$ML^{-1}T^{-1}$
g	Gravity acceleration vector	LT^{-2}
v_i	Actual average groundwater velocity along the x_i direction	LT^{-1}
D_{ii}	Principal component of the tensor of hydrodynamic dispersion coefficients	/
$R_{reac,n}$	Accounts for concentration changes due to the (bio) chemical reactions	$M{\cdot}L^3$
q_s	Volume flux per unit volume of the aquifer	L^3T^{-1}
θ	Porosity	/
C_n^s	Concentration of the n-th component in the source/sink term	$M{\cdot}L^3$
C_n	Total concentration of the n-th component	$M{\cdot}L^3$
Q_e	Amount of adsorbate adsorbed per unit mass of adsorbent at adsorption equilibrium	mg/g
K_l	Constant related to adsorption capacity in Langmuir model	/
Q_{max}	Maximum adsorption capacity of sorbent	mg/g
C_e	The equilibrium concentration of the solution	mg/L
$\varphi(0 \le \varphi \le 1)$	Coverage fraction	/
k_a	Adsorption rate constants	/
k_d	Desorption rate constants	/
C_0	Initial adsorbate concentration in solution	mg/L
k_2	The pseudo-second-order rate constant	g/(mg·min)
q_t	The adsorbate concentration in the solid phase at time t	mg/g
q_e	The adsorbate concentration in the solid phase at time equilibrium	mg/g
K_f	Adsorption capacity and adsorption strength in Freundlich model	/
n	Constant indicating the nonlinear magnitude of the adsorption isotherm	/
x_l	Summed mass fraction of solid phase exhibiting linear sorption	/
K_{Dr}	Mass-averaged partition coefficient for the summed linear components	/
$(x_{nl})_i$	The mass fraction of the i-th nonlinearly sorbing component	/
R	Gas constant	kJ/(mol•K)
T	Absolute temperature	K
C_s	Solubility of adsorbate	mg/L
Q_T	Total adsorption	mg/kg
K_{om}	Partition coefficient	L/kg
$Q_{max,D}$	Saturated adsorption capacity	mg/kg
S_t	Sorbent concentration at time t	mg/kg
S_0	Sorbent concentration at time 0	mg/kg
F_r, F_s and F_{vs}	The fractions of rapid, slow desorption, and very slow domains, respectively	/
k_r, k_s, and k_{vs}	The kinetic constants of the three desorption domains.	h^{-1}
a	Initial adsorption rate constant	mg/(g·min)
b	Adsorption resistance constant related to surface coverage	g/mg
X	The biomass concentrations	M/L^3
E	Electron acceptor concentrations	M/L^3
S	Substrate concentrations	M/L^3
K_s, K_E	The half-saturations of the substrate and electron acceptor respectively	M/L^3
μ_{max}	Maximum specific growth rate	T^{-1}
Y	The biomass yield coefficient	/

(*continued on next page*)

(*continued*)

Symbol	Parameter	Unit
b_1	The first-order decay coefficient	T^{-1}
μ	Specific growth rate	h^{-1}
S_i	The substrate concentration	mg/L
Ks	The half saturation constant	mg/L
S	The substrate concentration	mg/L
K_i	The inhibition constant	mg/L
S_m	The maximum substrate concentration above which growth ceases	mg/L
S_0	Initial substrate concentration	mg/L
k_0	The zero-order reaction rate constant	/
C	Organic pollutants content at different times	%
C_0	Initial organic pollutants content	%
$t_{1/2}$	Time required for microorganisms to degrade the organic pollutants content by half	d

CRediT authorship contribution statement

Yue Zhang: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Mingxu Cao:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis. **Zhenxue Dai:** Writing – review & editing, Visualization, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Hao Wang:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Sida Jia:** Writing – review & editing, Formal analysis, Conceptualization. **Lulu Xu:** Writing – review & editing, Formal analysis, Conceptualization. **Xiaoying Zhang:** Writing – review & editing, Formal analysis, Conceptualization. **Mohamad Reza Soltanian:** Writing – review & editing, Formal analysis, Conceptualization. **Javier Samper Calvete:** Writing – review & editing, Formal analysis, Conceptualization. **Huichao Yin:** Writing – review & editing, Formal analysis, Conceptualization. **Kenneth C. Carroll:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

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.

Acknowledgments

This work was funded by the National Key Research and Development Program of China (No.2024YFC3713800), the National Natural Science Foundation of China (No.U2267217, No.42141011), the Shandong Key Water Conservancy Science, and Technology Project (No.2024370203001957).

Data availability

Data will be made available on request.

References

[1] O.D. Ekpe, H. Moon, J. Pyo, J.-E. Oh, Prioritization of monitoring compounds from SNTS identified organic micropollutants in contaminated groundwater using a machine learning optimized ToxPi model, Water Res. 270 (2025) 122824, https://doi.org/10.1016/j.watres.2024.122824.

[2] H. Lü, J.-L. Wei, G.-X. Tang, Y.-S. Chen, Y.-H. Huang, R. Hu, C.-H. Mo, H.-M. Zhao, L. Xiang, Y.-W. Li, Q.-Y. Cai, Q.X. Li, Microbial consortium degrading of organic pollutants: source, degradation efficiency, pathway, mechanism and application, J. Clean. Prod. 451 (2024) 141913, https://doi.org/10.1016/j.jclepro.2024.141913.

[3] K. Chen, F. Wu, L. Li, K. Zhang, J. Huang, F. Cheng, Z. Yu, A.L. Hicks, J. You, Prioritizing organic pollutants for shale gas exploitation: life cycle environmental risk assessments in China and the US, Environ. Sci. & Technol. 58 (19) (2024) 8149–8160, https://doi.org/10.1021/acs.est.3c10288.

[4] Y. Wang, C. Wang, S. Guo, C. Wang, J. Wang, D. Yang, Z. Chen, T. Wu, B. Wang, Interception behaviors of mixed chlorinated hydrocarbons in groundwater by DDAC-amended LPBs: Control mechanism and CFD simulation, J. Hazard. Mater. 480 (2024) 135923, https://doi.org/10.1016/j.jhazmat.2024.135923.

[5] L. Zhang, Y. Zhong, Q. Fan, S. Li, J. Zhu, X. Ma, Y. Zhu, R. Wu, Z. Zhang, F. Zhou, Y. Wu, M. Cai, Y. Ma, Coupled physical-biogeochemical dynamics of polycyclic aromatic compounds in the East China Sea, Environ. Sci. & Technol. 59 (9) (2025) 4684–4698, https://doi.org/10.1021/acs.est.4c11906.

[6] R.D. Arcega, R.-J. Chen, P.-S. Chih, Y.-H. Huang, W.-H. Chang, T.-K. Kong, C.-C. Lee, T. Mahmudiono, C.-C. Tsui, W.-C. Hou, H.-T. Hsueh, H.-L. Chen, Toxicity prediction: an application of alternative testing and computational toxicology in contaminated groundwater sites in Taiwan, J. Environ. Manag. 328 (2023) 116982, https://doi.org/10.1016/j.jenvman.2022.116982.

[7] E.Y. Lee, D. Fathy, X. Xiang, D. Spahić, M.S. Ahmed, E. Fathi, M. Sami, Middle Miocene syn-rift sequence on the central Gulf of Suez, Egypt: depositional environment, diagenesis, and their roles in reservoir quality, Mar. Pet. Geol. 174 (2025) 107305, https://doi.org/10.1016/j.marpetgeo.2025.107305.

[8] C.J. Newell, D.T. Adamson, P.R. Kulkarni, B.N. Nzeribe, J.A. Connor, J. Popovic, H.F. Stroo, Monitored natural attenuation to manage PFAS impacts to groundwater: scientific basis, Groundw. Monit. Remediat. 41 (4) (2021) 76–89, https://doi.org/10.1111/gwmr.12486.

[9] Z. Liang, X. Tu, H. Liu, K. Zhang, Q. Pan, X. He, Y. Jia, Y. Sang, Occurrence of volatile and semi-volatile organic compounds in solid waste landfills and their pollution risk to groundwater, J. Hazard. Mater. 488 (2025) 137456, https://doi.org/10.1016/j.jhazmat.2025.137456.

[10] X. Yang, P.-J. De Buyck, R. Zhang, D. Manhaeghe, H. Wang, L. Chen, Y. Zhao, K. Demeestere, S.W.H. Van Hulle, Enhanced removal of refractory humic- and fulvic-like organics from biotreated landfill leachate by ozonation in packed bubble columns, Sci. Total Environ. 807 (2022) 150762, https://doi.org/10.1016/j.scitotenv.2021.150762.

[11] M. Vijayanand, A. Ramakrishnan, R. Subramanian, P.K. Issac, M. Nasr, K.S. Khoo, R. Rajagopal, B. Greff, N.I. Wan Azelee, B.-H. Jeon, S.W. Chang, B. Ravindran, Polyaromatic hydrocarbons (PAHs) in the water environment: a review on toxicity, microbial biodegradation, systematic biological advancements, and environmental fate, Environ. Res. 227 (2023) 115716, https://doi.org/10.1016/j.envres.2023.115716.

[12] F. Qiao, J. Wang, Z. Chen, S. Zheng, A.K. Kwaw, Y. Zhao, J. Huang, Experimental research on the transport-transformation of organic contaminants under the influence of multi-field coupling at a site scale, J. Hazard. Mater. 470 (2024) 134222, https://doi.org/10.1016/j.jhazmat.2024.134222.

[13] H. Qiu, J. Xu, Y. Yuan, E.J. Alesi, X. Liang, B. Cao, Low-disturbance land remediation using vertical groundwater circulation well technology: the first commercial deployment in an operational chemical plant, Sci. Total Environ. 944 (2024) 173804, https://doi.org/10.1016/j.scitotenv.2024.173804.

[14] M. Herraiz-Carboné, A. Santos, A. Checa-Fernández, C.M. Domínguez, S. Cotillas, Removal of organochlorine pollutants from DNAPL-saturated groundwater using electrolysis with MMO anodes, Chem. Eng. J. 486 (2024) 150238, https://doi.org/10.1016/j.cej.2024.150238.

[15] R.J. Lenhard, J.L. Rayner, J. García-Rincón, Testing an analytical model for predicting subsurface LNAPL distributions from current and historic fluid levels in monitoring wells: a preliminary test considering hysteresis, Water 11 (11) (2019) 2404, https://doi.org/10.3390/w11112404.

[16] S.K. Ngueleu, F. Rezanezhad, R.I. Al-Raoush, P. Van Cappellen, Sorption of benzene and naphthalene on (semi)-arid coastal soil as a function of salinity and temperature, J. Contam. Hydrol. 219 (2018) 61–71, https://doi.org/10.1016/j.jconhyd.2018.11.001.

[17] X. Xiao, Q. Zheng, R. Shen, K. Huang, H. Xu, B. Tu, Y. Zhao, Patterns of groundwater bacterial communities along the petroleum hydrocarbon gradient, J. Environ. Chem. Eng. 10 (6) (2022) 108773, https://doi.org/10.1016/j.jece.2022.108773.

[18] D. Kaown, K.C. Carroll, J. Mahlknecht, Y.J. Kim, J.-Y. Shin, S.-S. Lee, K.-K. Lee, Influence of saline water and heavy rain on the fate of chlorinated ethenes in groundwater characterized by compound-specific isotope and microbial data, J. Hazard. Mater. 487 (2025) 137238, https://doi.org/10.1016/j.jhazmat.2025.137238.

[19] J.L. Shelton, J.C. McIntosh, P.D. Warwick, A.L.Z. Yi, Fate of injected CO2 in the Wilcox Group, Louisiana, Gulf Coast Basin: Chemical and isotopic tracers of microbial–brine–rock–CO2 interactions, Appl. Geochem. 51 (2014) 155–169, https://doi.org/10.1016/j.apgeochem.2014.09.015.

[20] K.S. Lari, G.B. Davis, J.L. Rayner, T.P. Bastow, G.J. Puzon, Natural source zone depletion of LNAPL: a critical review supporting modelling approaches, Water Res. 157 (2019) 630–646, https://doi.org/10.1016/j.watres.2019.04.001.

[21] W. Hu, J. Zhang, D. Li, Y. Yuan, Y. Tang, K. Hui, Y. Jiang, W. Tan, Study on factors influencing the transport and transformation of polycyclic aromatic hydrocarbons in soil–groundwater systems, Emerg. Contam. 11 (2) (2025) 100472, https://doi.org/10.1016/j.emcon.2025.100472.

[22] Y. Yang, J. Li, N. Lv, H. Wang, H. Zhang, Multiphase migration and transformation of BTEX on groundwater table fluctuation in riparian petrochemical sites, Environ. Sci. Pollut. Res. 30 (19) (2023) 55756–55767, https://doi.org/10.1007/s11356-023-26393-8.

[23] A. Cavelan, F. Golfier, S. Colombano, H. Davarzani, J. Deparis, P. Faure, A critical review of the influence of groundwater level fluctuations and temperature on LNAPL contaminations in the context of climate change, Sci. Total Environ. 806 (2022) 150412, https://doi.org/10.1016/j.scitotenv.2021.150412.

[24] L.P. Wilson, E.J. Bouwer, Biodegradation of aromatic compounds under mixed oxygen/denitrifying conditions: a review, J. Ind. Microbiol. Biotechnol. 18 (2-3) (1997) 116–130, https://doi.org/10.1038/sj.jim.2900288.

[25] U.S. EPA, Superfund Remedy Report, (2025).

[26] J.G. Langeveld, J. Post, K.F. Makris, B. Palsma, M. Kuiper, E. Liefting, Monitoring organic micropollutants in stormwater runoff with the method of fingerprinting, Water Res. 235 (2023) 119883, https://doi.org/10.1016/j.watres.2023.119883.
[27] X. Duan, J. Li, Y. Li, Y. Xu, S. Chao, Y. Shi, Accumulation of typical persistent organic pollutants and heavy metals in bioretention facilities: Distribution, risk assessment, and microbial community impact, Environ. Res. 252 (2024) 119107, https://doi.org/10.1016/j.envres.2024.119107.
[28] Y. Wu, M. Xu, S. Liu, Generative Artificial Intelligence: A New Engine for Advancing Environmental Science and Engineering, Environ. Sci. & Technol. 58 (40) (2024) 17524–17528, https://doi.org/10.1021/acs.est.4c07216.
[29] F. Ma, J. Chen, Z. Dai, F. Cai, D. Wang, Y. Ma, Impact of groundwater extraction intensity on the monitoring design for seawater intrusion in heterogeneous coastal aquifers, J. Hydrol. 661 (2025) 133638, https://doi.org/10.1016/j.jhydrol.2025.133638.
[30] A. Wolfsberg, Z. Dai, L. Zhu, P. Reimus, T. Xiao, D. Ware, Colloid-Facilitated Plutonium Transport in Fractured Tuffaceous Rock, Environ. Sci. & Technol. 51 (10) (2017) 5582–5590, https://doi.org/10.1021/acs.est.7b00968.
[31] S.J. Berg, W.A. Illman, Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments, Water Resour. Res. 47 (9) (2011), https://doi.org/10.1029/2011WR010429.
[32] A. Anshuman, T.I. Eldho, A parallel workflow framework using encoder-decoder LSTMs for uncertainty quantification in contaminant source identification in groundwater, J. Hydrol. 619 (2023) 129296, https://doi.org/10.1016/j.jhydrol.2023.129296.
[33] Q. Guo, X. Shi, X. Kang, S. Hao, L. Liu, J. Wu, Evaluation of the benefits of improved permeability estimation on high-resolution characterization of DNAPL distribution in aquifers with low-permeability lenses, J. Hydrol. 603 (2021) 126955, https://doi.org/10.1016/j.jhydrol.2021.126955.
[34] L. He, H. Cheng, Z. Nan, Y. Gong, H. Guo, J. Mao, J. Zhang, Improving joint identification of groundwater contaminant source and non-Gaussian distributed conductivity field using a deep learning-based ensemble smoother, J. Hydrol. 658 (2025) 133202, https://doi.org/10.1016/j.jhydrol.2025.133202.
[35] Z. Pan, Z. Guo, K. Chen, W. Lu, C. Zheng, A deep adaptive bidirectional generative adversarial neural network (Bi-GAN) for groundwater contamination source estimation, J. Hydrol. 653 (2025) 132753, https://doi.org/10.1016/j.jhydrol.2025.132753.
[36] J. Li, Z. Wu, S. Zhang, W. Lu, Joint identification of hydraulic conductivity and groundwater pollution sources using unscented Kalman smoother with multiple data assimilation and deep learning, Ecotoxicol. Environ. Saf. 295 (2025) 118134, https://doi.org/10.1016/j.ecoenv.2025.118134.
[37] F.H. Maina, F. Delay, P. Ackerer, Estimating initial conditions for groundwater flow modeling using an adaptive inverse method, J. Hydrol. 552 (2017) 52–61, https://doi.org/10.1016/j.jhydrol.2017.06.041.
[38] M. Pang, C.A. Shoemaker, Comparison of parallel optimization algorithms on computationally expensive groundwater remediation designs, Sci. Total Environ. 857 (2023) 159544, https://doi.org/10.1016/j.scitotenv.2022.159544.
[39] W. Zhang, T. Xu, Z. Chen, J.J. Gómez-Hernández, C. Lu, J. Yang, Y. Ye, M. Jing, Simultaneous identification of a non-point contaminant source with Gaussian spatially distributed release and heterogeneous hydraulic conductivity in an aquifer using the LES-MDA method, J. Hydrol. 630 (2024) 130745, https://doi.org/10.1016/j.jhydrol.2024.130745.
[40] Y. Ji, Y. Zha, X. Gong, Exploring different representations of hydraulic tomographic data for deep learning: Sequence or image, J. Hydrol. 648 (2025) 132368, https://doi.org/10.1016/j.jhydrol.2024.132368.
[41] Z. Wang, W. Lu, Z. Chang, J. Luo, A combined search method based on a deep learning combined surrogate model for groundwater DNAPL contamination source identification, J. Hydrol. 616 (2023) 128854, https://doi.org/10.1016/j.jhydrol.2022.128854.
[42] Q. Guo, Y. He, M. Liu, Y. Zhao, Y. Liu, J. Luo, Reduced Geostatistical Approach With a Fourier Neural Operator Surrogate Model for Inverse Modeling of Hydraulic Tomography, Water Resour. Res. 60 (6) (2024) e2023WR034939, https://doi.org/10.1029/2023WR034939.
[43] D. Ruan, J. Bian, Y. Wang, J. Wu, Z. Gu, Identification of groundwater pollution sources and health risk assessment in the Songnen Plain based on PCA-APCS-MLR and trapezoidal fuzzy number-Monte Carlo stochastic simulation model, J. Hydrol. 632 (2024) 130897, https://doi.org/10.1016/j.jhydrol.2024.130897.
[44] Y. Zhang, Y. Liu, A. Zhou, L. zhang, Identification of groundwater pollution from livestock farming using fluorescence spectroscopy coupled with multivariate statistical methods, Water Res. 206 (2021) 117754, https://doi.org/10.1016/j.watres.2021.117754.
[45] O.D. Ekpe, G. Choo, J.-K. Kang, S.-T. Yun, J.-E. Oh, Identification of organic chemical indicators for tracking pollution sources in groundwater by machine learning from GC-HRMS-based suspect and non-target screening data, Water Res. 252 (2024) 121130, https://doi.org/10.1016/j.watres.2024.121130.
[46] A.A. Abdelhady, B. Seuss, S. Jain, D. Fathy, M. Sami, A. Ali, A. Elsheikh, M. S. Ahmed, A.M.T. Elewa, A.M. Hussain, Molecular technology in paleontology and paleobiology: Applications and limitations, Quat. Int. 685 (2024) 24–38, https://doi.org/10.1016/j.quaint.2024.01.006.
[47] Z. Chen, L. Zong, J.J. Gómez-Hernández, T. Xu, Y. Jiang, Q. Zhou, H. Yang, Z. Jia, S. Mei, Contaminant source and aquifer characterization: An application of ES-MDA demonstrating the assimilation of geophysical data, Adv. Water Resour. 181 (2023) 104555, https://doi.org/10.1016/j.advwatres.2023.104555.
[48] L. Lévy, T.S. Bording, G. Fiandaca, A.V. Christiansen, L.M. Madsen, L. F. Bennedsen, T.H. Jørgensen, L. MacKinnon, J.F. Christensen, Managing the remediation strategy of contaminated megasites using field-scale calibration of geo-electrical imaging with chemical monitoring, Sci. Total Environ. 920 (2024) 171013, https://doi.org/10.1016/j.scitotenv.2024.171013.
[49] Y. Wu, H. Xie, J. Wang, Y. Shi, M. Zhang, Deep learning-enhanced inverse framework for high-fidelity characterization of heterogeneous aquifers and DNAPL contamination zones using sparse geophysical data, J. Hydrol. 661 (2025) 133693, https://doi.org/10.1016/j.jhydrol.2025.133693.
[50] J.T. McGarr, D.C. McAvoy, J. Hobbs, L. Lupton, E. Poston, T. Marsh, D. M. Sturmer, C. Dietsch, M.R. Soltanian, An Integrated Approach to Mapping Per- and Polyfluoroalkyl Substances Sorption in Sediments Using Electromagnetic Induction, ACS Earth Space Chem. (2025), https://doi.org/10.1021/acsearthspacechem.5c00081.
[51] L. Chen, G. Ding, J. Lu, Y. Liu, S. Wei, X. Guo, C. Tang, H. Sun, H. Zuo, Gas tube effect: A transport mode of deeply buried volatile DNAPLs to shallow strata, J. Hydrol. 630 (2024) 130696, https://doi.org/10.1016/j.jhydrol.2024.130696.
[52] S.M. Gorelick, B. Evans, I. Remson, Identifying sources of groundwater pollution: An optimization approach, Water Resour. Res. 19 (3) (1983) 779–790, https://doi.org/10.1029/WR019i003p00779.
[53] A.C. Bagtzoglou, D.E. Dougherty, A.F.B. Tompson, Application of particle methods to reliable identification of groundwater pollution sources, Water Resour. Manag. 6 (1) (1992) 15–23, https://doi.org/10.1007/BF00872184.
[54] B.J. Wagner, Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling, J. Hydrol. 135 (1) (1992) 275–303, https://doi.org/10.1016/0022-1694(92)90092-A.
[55] A.M. Michalak, P.K. Kitanidis, Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, Water Resour. Res. 40 (8) (2004), https://doi.org/10.1029/2004WR003214.
[56] Y. Xu, W. Lu, Z. Pan, Z. Wang, C. Luo, Y. Bai, Intelligent enhanced particle filter with deep residual network surrogate for accurate groundwater pollution source characterization, J. Hydrol. 642 (2024) 131904, https://doi.org/10.1016/j.jhydrol.2024.131904.
[57] K.P. Tripathy, A.K. Mishra, Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions, J. Hydrol. 628 (2024) 130458, https://doi.org/10.1016/j.jhydrol.2023.130458.
[58] T. Rajaee, H. Ebrahimi, V. Nourani, A review of the artificial intelligence methods in groundwater level modeling, J. Hydrol. 572 (2019) 336–351, https://doi.org/10.1016/j.jhydrol.2018.12.037.
[59] K.B.W. Boo, A. El-Shafie, F. Othman, M.M.H. Khan, A.H. Birima, A.N. Ahmed, Groundwater level forecasting with machine learning models: A review, Water Res. 252 (2024) 121249, https://doi.org/10.1016/j.watres.2024.121249.
[60] R. Haggerty, J. Sun, H. Yu, Y. Li, Application of machine learning in groundwater quality modeling - A comprehensive review, Water Res. 233 (2023) 119745, https://doi.org/10.1016/j.watres.2023.119745.
[61] H. Pandya, K. Jaiswal, M. Shah, A Comprehensive Review of Machine Learning Algorithms and Its Application in Groundwater Quality Prediction, Arch. Comput. Methods Eng. 31 (8) (2024) 4633–4654, https://doi.org/10.1007/s11831-024-10126-2.
[62] C. Zhan, Z. Dai, Z. Yang, X. Zhang, Z. Ma, H.V. Thanh, M.R. Soltanian, Subsurface sedimentary structure identification using deep learning: A review, EarthSci. Rev. 239 (2023) 104370, https://doi.org/10.1016/j.earscirev.2023.104370.
[63] Z. Dai, C. Zhan, H. Yin, J. Chen, L. Xu, Y. Xia, S. Yang, W. Chen, M. Cao, Z. Du, X. Zhang, B. Yan, Y. Ma, H. Wang, F. Moeini, M.R. Soltanian, H.V. Thanh, K. C. Carroll, Incorporating Deep Learning Into Hydrogeological Modeling: Advancements, Challenges, and Future Directions, J. Geophys. Res. Mach. Learn. Comput. 2 (2) (2025) e2025JH000703, https://doi.org/10.1029/2025JH000703.
[64] H. Yan, Q. Zheng, L. Zeng, Conditional generative adversarial networks for groundwater contamination characterization and source identification, J. Hydrol. 632 (2024) 130900, https://doi.org/10.1016/j.jhydrol.2024.130900.
[65] C. Luo, X. Wang, Y.J. Xu, Q. Lv, X. Ji, B. Mao, S. Jia, Z. Liu, Y. Rong, Y. Dai, Multi-machine learning methods for rapid and synergistic inversion of groundwater contamination source, hydrogeologic parameter and boundary condition, J. Contam. Hydrol. 273 (2025) 104599, https://doi.org/10.1016/j.jconhyd.2025.104599.
[66] Z. Zhang, Q. Li, Q. Hu, J. Xue, T. Liu, Z. Tang, F. Wang, C. Zhang, C. Lu, Z. Wang, M. Gao, C. Liu, Deep learning approach for reconstructing three-dimensional distribution of NO2 on an urban scale, Remote Sens. Environ. 321 (2025) 114678, https://doi.org/10.1016/j.rse.2025.114678.
[67] J. Bi, Y. Li, X. Zhang, H. Yuan, Z. Wang, J. Zhang, M. Zhou, Multi-Indicator Water Quality Prediction Using Multimodal Bottleneck Fusion and ITransformer with Attention, IEEE Int. Conf. Syst. Man Cybern. (SMC) 2024 (2024) 2367–2372.
[68] Z. Pan, W. Lu, Y. Bai, Groundwater contamination source estimation based on a refined particle filter associated with a deep residual neural network surrogate, Hydrogeol. J. 30 (3) (2022) 881–897, https://doi.org/10.1007/s10040-022-02454-z.
[69] Q. Guo, Y. Zhao, C. Lu, J. Luo, High-dimensional inverse modeling of hydraulic tomography by physics informed neural network (HT-PINN), J. Hydrol. 616 (2023) 128828, https://doi.org/10.1016/j.jhydrol.2022.128828.
[70] Y. Ji, Y. Zha, T.-C.J. Yeh, L. Shi, Y. Wang, Groundwater inverse modeling: Physics-informed neural network with disentangled constraints and errors, J. Hydrol. 640 (2024) 131703, https://doi.org/10.1016/j.jhydrol.2024.131703.
[71] Q. Hou, X. Xu, Z. Sun, J. Wang, V.P. Singh, Physics informed neural network for forward and inverse multispecies contaminant transport with variable parameters, J. Hydrol. 655 (2025) 132977, https://doi.org/10.1016/j.jhydrol.2025.132977.

[72] H.I. Essaid, B.A. Bekins, I.M. Cozzarelli, Organic contaminant transport and fate in the subsurface: Evolution of knowledge and understanding, Water Resour. Res. 51 (7) (2015) 4861–4902, https://doi.org/10.1002/2015WR017121.
[73] B. Anneser, F. Einsiedl, R.U. Meckenstock, L. Richters, F. Wisotzky, C. Griebler, High-resolution monitoring of biogeochemical gradients in a tar oil-contaminated aquifer, Appl. Geochem. 23 (6) (2008) 1715–1730, https://doi.org/10.1016/j.apgeochem.2008.02.003.
[74] E. Warren, B.A. Bekins, Relating subsurface temperature changes to microbial activity at a crude oil-contaminated site, J. Contam. Hydrol. 182 (2015) 183–193, https://doi.org/10.1016/j.jconhyd.2015.09.007.
[75] K. Pruess, A. Battistelli, TMVOC, A Numerical Simulator for Three-Phase Non-isothermal Flows of Multicomponent Hydrocarbon Mixtures in Variably Saturated Heterogeneous Media, Office of Scientific & Technical Information Technical Reports (2005).
[76] R. Ai, R. Zha, X. Kang, J.H. Lee, M. Zhang, X. Shi, Integrating CSIA and reactive transport modeling to characterize DNAPL source zone architecture during natural attenuation in biogeochemically heterogeneous aquifers, J. Hydrol. 661 (2025) 133565, https://doi.org/10.1016/j.jhydrol.2025.133565.
[77] C. Feng, F. Liu, F. Huang, L. Chen, E. Bi, Dense nonaqueous phase liquids back diffusion controlled by biodegradation and heterogeneous sorption-desorption, J. Clean. Prod. 382 (2023) 135370, https://doi.org/10.1016/j.jclepro.2022.135370.
[78] W. Liu, F. Cheng, W. Li, B. Xing, S. Tao, Desorption behaviors of BDE-28 and BDE-47 from natural soils with different organic carbon contents, Environ. Pollut. 163 (2012) 235–242, https://doi.org/10.1016/j.envpol.2011.12.043.
[79] M.L. Brusseau, N. Khan, Y. Wang, N. Yan, S. Van Glubt, K.C. Carroll, Nonideal Transport and Extended Elution Tailing of PFOS in Soil, Environ. Sci. Technol. 53 (18) (2019) 10654–10664, https://doi.org/10.1021/acs.est.9b02343.
[80] V. Rafiei, A.P. Nejadhashemi, Watershed scale PFAS fate and transport model for source identification and management implications, Water Res. 240 (2023) 120073, https://doi.org/10.1016/j.watres.2023.120073.
[81] A.C. Umeh, R. Naidu, E. Olisa, Y. Liu, F. Qi, D. Bekele, A systematic investigation of single solute, binary and ternary PFAS transport in water-saturated soil using batch and 1-dimensional column studies: Focus on mixture effects, J. Hazard. Mater. 461 (2024) 132688, https://doi.org/10.1016/j.jhazmat.2023.132688.
[82] X.-Y. Yu, G.-G. Ying, R.S. Kookana, Sorption and Desorption Behaviors of Diuron in Soils Amended with Charcoal, J. Agric. Food Chem. 54 (22) (2006) 8545–8550, https://doi.org/10.1021/jf061354y.
[83] J. Wang, X. Guo, Adsorption kinetic models: physical meanings, applications, and solving methods, J. Hazard. Mater. 390 (2020) 122156, https://doi.org/10.1016/j.jhazmat.2020.122156.
[84] Y. Yang, G. Sheng, Enhanced pesticide sorption by soils containing particulate matter from crop residue burns, Environ. Sci. Technol. 37 (16) (2003) 3635–3639, https://doi.org/10.1021/es034006a.
[85] A. Yadav, N. Bagotia, A.K. Sharma, S. Kumar, Simultaneous adsorptive removal of conventional and emerging contaminants in multi-component systems for wastewater remediation: A critical review, Sci. Total Environ. 799 (2021) 149500, https://doi.org/10.1016/j.scitotenv.2021.149500.
[86] J. Walter, Jr Weber, M. Paul, Lynn McGinley, E. Katz, A distributed reactivity model for sorption by soils and sediments. 1. Conceptual basis and equilibrium assessments, Environ. Sci. & Technol. 26 (10) (1992) 1955–1962.
[87] B. Xing, J.J. Pignatello, Dual-mode sorption of low-polarity compounds in glassy poly(Vinyl Chloride) and soil organic matter, Environ. Sci. Technol. 31 (3) (1997) 792–799, https://doi.org/10.1021/es960481f.
[88] L. Liping, C. Guanghuan, D. Jingyou, S. Mingyang, C. Huanyu, Y. Qiang, X. Xinhua, Mechanism of and relation between the sorption and desorption of nonylphenol on black carbon-inclusive sediment, Environ. Pollut. 190 (2014) 101–108, https://doi.org/10.1016/j.envpol.2014.03.027.
[89] K. Sven, Vetenskapsakad, About the theory of so-called adsorption of soluble substances, (1962).
[90] J.S.T. Adadevoh, C.A. Ramsburg, R.M. Ford, Chemotaxis increases the retention of bacteria in porous media with residual NAPL entrapment, Environ. Sci. Technol. 52 (13) (2018) 7289–7295, https://doi.org/10.1021/acs.est.8b01172.
[91] P. Li, C.D. Wallace, J.T. McGarr, F. Moeini, Z. Dai, M.R. Soltanian, Investigating key drivers of N2O emissions in heterogeneous riparian sediments: Reactive transport modeling and statistical analysis, Sci. Total Environ. 905 (2023) 166930, https://doi.org/10.1016/j.scitotenv.2023.166930.
[92] Y. Meng, P. Li, V. Elumalai, Factors affecting distribution and ecological risk assessment of volatile organic compounds (VOCs) in groundwater of the Huazhou district in northwestern China, Environ. Pollut. 363 (2024) 125243, https://doi.org/10.1016/j.envpol.2024.125243.
[93] M.J. Lønborg, P. Engesgaard, P.L. Bjerg, D. Rosbjerg, A steady state redox zone approach for modeling the transport and degradation of xenobiotic organic compounds from a landfill site, J. Contam. Hydrol. 87 (3) (2006) 191–210, https://doi.org/10.1016/j.jconhyd.2006.05.004.
[94] Y. Meng, P. Li, Effects of environmental factors on groundwater BTEX pollution: a quantitative analysis, J. Environ. Chem. Eng. 13 (6) (2025) 119915, https://doi.org/10.1016/j.jece.2025.119915.
[95] R. Zuo, K. Han, D. Xu, Q. Li, J. Liu, Z. Xue, X. Zhao, J. Wang, Response of environmental factors to attenuation of toluene in vadose zone, J. Environ. Manag. 302 (2022) 113968, https://doi.org/10.1016/j.jenvman.2021.113968.
[96] Y. Cheng, K. Zhang, K. Huang, H. Zhang, Meta-analysis and machine learning models for anaerobic biodegradation rates of organic contaminants in sediments and sludge, Environ. Sci. Technol. 58 (29) (2024) 12976–12988, https://doi.org/10.1021/acs.est.4c01033.
[97] S. Bajaj, D.K. Singh, Biodegradation of persistent organic pollutants in soil, water and pristine sites by cold-adapted microorganisms: mini review, Int. Biodeterior. Biodegrad. 100 (2015) 98–105, https://doi.org/10.1016/j.ibiod.2015.02.023.
[98] N. Koproch, A. Dahmke, R. Köber, The aqueous solubility of common organic groundwater contaminants as a function of temperature between 5 and 70 °C, Chemosphere 217 (2019) 166–175, https://doi.org/10.1016/j.chemosphere.2018.10.153.
[99] Q. Wang, S. Guo, M. Ali, X. Song, Z. Tang, Z. Zhang, M. Zhang, Y. Luo, Thermally enhanced bioremediation: a review of the fundamentals and applications in soil and groundwater remediation, J. Hazard. Mater. 433 (2022) 128749, https://doi.org/10.1016/j.jhazmat.2022.128749.
[100] P. Hosseininoosheri, H.R. Lashgari, K. Sepehrnoori, A novel method to model and characterize in-situ bio-surfactant production in microbial enhanced oil recovery, Fuel 183 (2016) 501–511, https://doi.org/10.1016/j.fuel.2016.06.035.
[101] J.S. Meyer, M.D. Marcus, H.L. Bergman, Inhibitory interactions of aromatic organics during microbial degradation, Environ. Toxicol. Chem. 3 (4) (1984) 583–587, https://doi.org/10.1002/etc.5620030408.
[102] E. Wantz, A. Kane, M. Lhuissier, A. Amrane, J.-L. Audic, A. Couvert, A mathematical model for VOCs removal in a treatment process coupling absorption and biodegradation, Chem. Eng. J. 423 (2021) 130106, https://doi.org/10.1016/j.cej.2021.130106.
[103] J. Monod, The growth of bacterial cultures, Sel. Pap. Mol. Biol. Jacques Monod 139 (2012) 606.
[104] K.S. Lari, G.B. Davis, J.L. Rayner, T.P. Bastow, Advective and diffusive gas phase transport in vadose zones: importance for defining vapour risks and natural source zone depletion of petroleum hydrocarbons, Water Res. 255 (2024) 121455, https://doi.org/10.1016/j.watres.2024.121455.
[105] M. Luo, X. Zhang, X. Zhu, T. Long, S. Cao, R. Yu, Bioremediation of chlorinated ethenes contaminated groundwater and the reactive transport modeling – A review, Environ. Res. 240 (2024) 117389, https://doi.org/10.1016/j.envres.2023.117389.
[106] N. Singh, C. Balomajumder, Batch growth kinetic studies for elimination of phenol and cyanide using mixed microbial culture, J. Water Process Eng. 11 (2016) 130–137, https://doi.org/10.1016/j.jwpe.2016.04.006.
[107] M. Luo, X. Zhang, S. Cao, Q. Chen, X. Zhu, C. Xu, D. Yu, M. Zhan, R. Yu, T. Long, Modeling the elongation of commingled BTEX and chlorinated ethene plumes undergoing biodegradation with a multi-level substrate interaction module, J. Hazard. Mater. 491 (2025) 137929, https://doi.org/10.1016/j.jhazmat.2025.137929.
[108] M.H. El-Naas, J.A. Acio, A.E. El Telib, Aerobic biodegradation of BTEX: Progresses and Prospects, J. Environ. Chem. Eng. 2 (2) (2014) 1104–1122, https://doi.org/10.1016/j.jece.2014.04.009.
[109] G. Swain, R.K. Sonwani, R.S. Singh, R.P. Jaiswal, B.N. Rai, A comparative study of 4-chlorophenol biodegradation in a packed bed and moving bed bioreactor: performance evaluation and toxicity analysis, Environ. Technol. Innov. 24 (2021) 101820, https://doi.org/10.1016/j.eti.2021.101820.
[110] A. Mathur, C. Majumder, Kinetics modelling of the biodegradation of benzene, toluene and phenol as single substrate and mixed substrate by using Pseudomonas putida, Chem. Biochem. Eng. Q. 24 (1) (2010) 101–109, https://doi.org/10.1038/cgt.2009.66.
[111] A.R. Bielefeldt, H.D. Stensel, Modeling competitive inhibition effects during biodegradation of BTEX mixtures, Water Res. 33 (3) (1999) 707–714, https://doi.org/10.1016/S0043-1354(98)00256-5.
[112] C.-W. Lin, Y.-W. Cheng, S.-L. Tsai, Multi-substrate biodegradation kinetics of MTBE and BTEX mixtures by Pseudomonas aeruginosa, Process Biochem. 42 (8) (2007) 1211–1217, https://doi.org/10.1016/j.procbio.2007.05.020.
[113] D.E.G. Trigueros, A.N. Módenes, A.D. Kroumov, F.R. Espinoza-Quiñones, Modeling of biodegradation process of BTEX compounds: kinetic parameters estimation by using Particle Swarm Global Optimizer, Process Biochem. 45 (8) (2010) 1355–1361, https://doi.org/10.1016/j.procbio.2010.05.007.
[114] J.V. Littlejohns, A.J. Daugulis, Kinetics and interactions of BTEX compounds during degradation by a bacterial consortium, Process Biochem. 43 (10) (2008) 1068–1076, https://doi.org/10.1016/j.procbio.2008.05.010.
[115] H. Yoon, G. Klinzing, H.W. Blanch, Competition for mixed substrates by microbial populations, Biotechnol. Bioeng. 19 (8) (1977) 1193–1210, https://doi.org/10.1002/bit.260190809.
[116] W. Chen, F. Wang, L. Zeng, Q. Li, Bioremediation of petroleum-contaminated soil by semi-aerobic aged refuse biofilter: Optimization and mechanism, J. Clean. Prod. 294 (2021) 125354, https://doi.org/10.1016/j.jclepro.2020.125354.
[117] S. Koley, Future perspectives and mitigation strategies towards groundwater arsenic contamination in West Bengal, India, Environ. Qual. Manag. 31 (4) (2022) 75–97, https://doi.org/10.1002/tqem.21784.
[118] F. Becher Quinodoz, A. Cabrera, M. Blarasin, E. Matteoda, M. Pascuini, S. Prámparo, L. Boumaiza, I. Matiatos, G. Schroeter, V. Lutri, D. Giacobone, Chemical and isotopic tracers combined with mixing models for tracking nitrate contamination in the Pampa de Pocho aquifer, Argentina, Environ. Res. 259 (2024) 119571, https://doi.org/10.1016/j.envres.2024.119571.
[119] X. Wang, C. Xiao, W. Yang, X. Liang, L. Zhang, J. Zhang, Analysis of the quality, source identification and apportionment of the groundwater in a typical arid and semi-arid region, J. Hydrol. 625 (2023) 130169, https://doi.org/10.1016/j.jhydrol.2023.130169.
[120] J. Jiang, S. Tang, D. Han, G. Fu, D. Solomatine, Y. Zheng, A comprehensive review on the design and optimization of surface water quality monitoring networks, Environ. Model. & Softw. 132 (2020) 104792, https://doi.org/10.1016/j.envsoft.2020.104792.

[121] M. Meggiorin, N. Naranjo-Fernández, G. Passadore, A. Sottani, G. Botter, A. Rinaldo, Data-driven statistical optimization of a groundwater monitoring network, J. Hydrol. 631 (2024) 130667, https://doi.org/10.1016/j.jhydrol.2024.130667.
[122] Y. Xiong, J. Luo, X. Liu, Y. Liu, X. Xin, S. Wang, Machine learning-based optimal design of groundwater pollution monitoring network, Environ. Res. 211 (2022) 113022, https://doi.org/10.1016/j.envres.2022.113022.
[123] R. Jia, J. Wu, Y. Zhang, Z. Luo, Site prioritization and performance assessment of groundwater monitoring network by using information-based methodology, Environ. Res. 212 (2022) 113181, https://doi.org/10.1016/j.envres.2022.113181.
[124] M.T. Ayvaz, A. Elçi, Identification of the optimum groundwater quality monitoring network using a genetic algorithm based optimization approach, J. Hydrol. 563 (2018) 1078–1091, https://doi.org/10.1016/j.jhydrol.2018.06.006.
[125] R. Salman, M.R. Nikoo, S.A. Shojaeezadeh, P.H.B. Beiglou, M. Sadegh, J. F. Adamowski, N. Alamdari, A novel Bayesian maximum entropy-based approach for optimal design of water quality monitoring networks in rivers, J. Hydrol. 603 (2021) 126822, https://doi.org/10.1016/j.jhydrol.2021.126822.
[126] S. Teimoori, M.H. Olya, C.J. Miller, Groundwater level monitoring network design with machine learning methods, J. Hydrol. 625 (2023) 130145, https://doi.org/10.1016/j.jhydrol.2023.130145.
[127] Y. Xia, C. Zhan, Z. Dai, J. Wu, X. Zhang, H. Yin, J. Yan, J. Chen, Z. Wang, M. R. Soltanian, K.C. Carroll, Impact of observation and surrogate-model noises on deep learning-based subsurface heterogeneous structure identification through monitoring network optimization, Adv. Water Resour. 207 (2026) 105204, https://doi.org/10.1016/j.advwatres.2025.105204.
[128] Q.K. Ha, V.T. Dang, L.P. Vo, D.H. Dang, Integrated approaches to track saline intrusion for fresh groundwater resource protection in the Mekong Delta, Groundw. Sustain. Dev. 23 (2023) 101046, https://doi.org/10.1016/j.gsd.2023.101046.
[129] N. Igwebuike, M. Ajayi, C. Okolie, T. Kanyerere, T. Halihan, Application of machine learning and deep learning for predicting groundwater levels in the West Coast Aquifer System, South Africa, Earth Sci. Inform. 18 (1) (2024) 6, https://doi.org/10.1007/s12145-024-01623-w.
[130] T.G. Sanders, D.D. Adrian, Sampling frequency for river quality monitoring, Water Resour. Res. 14 (4) (1978) 569–576, https://doi.org/10.1029/WR014i004p00569.
[131] M. Hosseini, R. Kerachian, A data fusion-based methodology for optimal redesign of groundwater monitoring networks, J. Hydrol. 552 (2017) 267–282, https://doi.org/10.1016/j.jhydrol.2017.06.046.
[132] T.H. Nguyen, B. Helm, H. Hettiarachchi, S. Caucci, P. Krebs, The selection of design methods for river water quality monitoring networks: a review, Environ. Earth Sci. 78 (3) (2019) 96, https://doi.org/10.1007/s12665-019-8110-x.
[133] R. Yang, J. Jiang, T. Pang, Z. Yang, F. Han, H. Li, H. Wang, Y. Zheng, Crucial time of emergency monitoring for reliable numerical pollution source identification, Water Res. 265 (2024) 122303, https://doi.org/10.1016/j.watres.2024.122303.
[134] M. Pang, E. Du, C. Zheng, Contaminant Transport Modeling and Source Attribution With Attention-Based Graph Neural Network, Water Resour. Res. 60 (6) (2024) e2023WR035278, https://doi.org/10.1029/2023WR035278.
[135] J. Chen, Z. Dai, S. Dong, X. Zhang, G. Sun, J. Wu, R. Ershadnia, S. Yin, M. R. Soltanian, Integration of Deep Learning and Information Theory for Designing Monitoring Networks in Heterogeneous Aquifer Systems, Water Resour. Res. 58 (10) (2022) e2022WR032429, https://doi.org/10.1029/2022WR032429.
[136] M. Cao, Z. Dai, J. Chen, H. Yin, X. Zhang, J. Wu, H.V. Thanh, M.R. Soltanian, An integrated framework of deep learning and entropy theory for enhanced high-dimensional permeability field identification in heterogeneous aquifers, Water Res. 268 (2025) 122706, https://doi.org/10.1016/j.watres.2024.122706.
[137] M. Cao, Z. Dai, J. Chen, Entropy-Guided Multivariate Groundwater Network Design for Multi-Source Data Assimilation Under Observational Uncertainty, Geophys. Res. Lett. 52 (19) (2025) e2025GL117466, https://doi.org/10.1029/2025GL117466.
[138] X. Kang, X. Shi, A. Revil, Z. Cao, L. Li, T. Lan, J. Wu, Coupled hydrogeophysical inversion to identify non-Gaussian hydraulic conductivity field by jointly assimilating geochemical and time-lapse geophysical data, J. Hydrol. 578 (2019) 124092, https://doi.org/10.1016/j.jhydrol.2019.124092.
[139] J. Irving, K. Singha, Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, Water Resour. Res. 46 (11) (2010), https://doi.org/10.1029/2009WR008340.
[140] J.P. Boyd, J.E. Chambers, P.B. Wilkinson, P.I. Meldrum, E. Bruce, A. Binley, Coupled Hydrogeophysical Modeling to Constrain Unsaturated Soil Parameters for a Slow-Moving Landslide, Water Resour. Res. 60 (10) (2024) e2023WR036319, https://doi.org/10.1029/2023WR036319.
[141] T. De Clercq, A. Jardani, P. Fischer, L. Thanberger, T.M. Vu, D. Pitaval, J.-M. Côme, P. Begassat, The use of electrical resistivity tomograms as a parameterization for the hydraulic characterization of a contaminated aquifer, J. Hydrol. 587 (2020) 124986, https://doi.org/10.1016/j.jhydrol.2020.124986.
[142] S.J. Ikard, K.C. Carroll, S.C. Brooks, D.F. Rucker, G. Smith-Vega, A. Elwes, Self-Potential Tomography Preconditioned by Particle Swarm Optimization—Application to Monitoring Hyporheic Exchange in a Bedrock River, Water Resour. Res. 60 (10) (2024) e2024WR037549, https://doi.org/10.1029/2024WR037549.
[143] Z. Han, X. Kang, J. Wu, X. Shi, Characterization of the non-Gaussian hydraulic conductivity field via deep learning-based inversion of hydraulic-head and self-potential data, J. Hydrol. 610 (2022) 127830, https://doi.org/10.1016/j.jhydrol.2022.127830.
[144] Q. Guo, M. Liu, J. Luo, Predictive Deep Learning for High-Dimensional Inverse Modeling of Hydraulic Tomography in Gaussian and Non-Gaussian Fields, Water Resour. Res. 59 (10) (2023) e2023WR035408, https://doi.org/10.1029/2023WR035408.
[145] Z. Dai, A. Wolfsberg, Z. Lu, H. Deng, Scale dependence of sorption coefficients for contaminant transport in saturated fractured rock, Geophys. Res. Lett. 36 (1) (2009), https://doi.org/10.1029/2008GL036516.
[146] X. Zhang, F. Ma, S. Yin, C.D. Wallace, M.R. Soltanian, Z. Dai, R.W. Ritzi, Z. Ma, C. Zhan, X. Lü, Application of upscaling methods for fluid flow and mass transport in multi-scale heterogeneous media: A critical review, Appl. Energy 303 (2021) 117603, https://doi.org/10.1016/j.apenergy.2021.117603.
[147] F. Jiang, Y. Guo, T. Tsuji, Y. Kato, M. Shimokawara, L. Esteban, M. Seyyedi, M. Pervukhina, M. Lebedev, R. Kitamura, Upscaling Permeability Using Multiscale X-Ray-CT Images With Digital Rock Modeling and Deep Learning Techniques, Water Resour. Res. 59 (3) (2023) e2022WR033267, https://doi.org/10.1029/2022WR033267.
[148] N. You, Y.E. Li, A. Cheng, 3D Carbonate Digital Rock Reconstruction Using Progressive Growing GAN, J. Geophys. Res. Solid Earth 126 (5) (2021) e2021JB021687, https://doi.org/10.1029/2021JB021687.
[149] J.E. Santos, D. Xu, H. Jo, C.J. Landry, M. Prodanović, M.J. Pyrcz, PoreFlow-Net: A 3D convolutional neural network to predict fluid flow through porous media, Adv. Water Resour. 138 (2020) 103539, https://doi.org/10.1016/j.advwatres.2020.103539.
[150] A. Bárdossy, S. Hörning, Gaussian and non-Gaussian inverse modeling of groundwater flow using copulas and random mixing, Water Resour. Res. 52 (6) (2016) 4504–4526, https://doi.org/10.1002/2014WR016820.
[151] L. Li, L. Stetler, Z. Cao, A. Davis, An iterative normal-score ensemble smoother for dealing with non-Gaussianity in data assimilation, J. Hydrol. 567 (2018) 759–766, https://doi.org/10.1016/j.jhydrol.2018.01.038.
[152] V. Todaro, M. D'Oria, A. Zanini, J.J. Gómez-Hernández, M.G. Tanda, Experimental sandbox tracer tests to characterize a two-facies aquifer via an ensemble smoother, Hydrogeol. J. 31 (6) (2023) 1665–1678, https://doi.org/10.1007/s10040-023-02662-1.
[153] H. Jung, H. Jo, S. Kim, K. Lee, J. Choe, Recursive update of channel information for reliable history matching of channel reservoirs using EnKF with DCT, J. Pet. Sci. Eng. 154 (2017) 19–37, https://doi.org/10.1016/j.petrol.2017.04.016.
[154] M. Ramgraber, R. Weatherl, F. Blumensaat, M. Schirmer, Non-Gaussian Parameter Inference for Hydrogeological Models Using Stein Variational Gradient Descent, Water Resour. Res. 57 (4) (2021) e2020WR029339, https://doi.org/10.1029/2020WR029339.
[155] C. Cao, J. Zhang, W. Gan, T. Nan, C. Lu, A Deep Learning-Based Data Assimilation Approach to Characterizing Coastal Aquifers Amid Non-Linearity and Non-Gaussianity Challenges, Water Resour. Res. 60 (7) (2024) e2023WR036899, https://doi.org/10.1029/2023WR036899.
[156] E. Laloy, R. Hérault, J. Lee, D. Jacques, N. Linde, 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, https://doi.org/10.1016/j.advwatres.2017.09.029.
[157] 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, https://doi.org/10.1016/j.cageo.2019.04.006.
[158] Z. Chen, T. Xu, J.J. Gómez-Hernández, A. Zanini, Q. Zhou, Reconstructing the release history of a contaminant source with different precision via the ensemble smoother with multiple data assimilation, J. Contam. Hydrol. 252 (2023) 104115, https://doi.org/10.1016/j.jconhyd.2022.104115.
[159] E. Laloy, R. Hérault, D. Jacques, N. Linde, Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network, Water Resour. Res. 54 (1) (2018) 381–406, https://doi.org/10.1002/2017WR022148.
[160] F. Moeini, R. Ershadnia, R.L. Rubinstein, R. Versteeg, P. Li, J.T. McGarr, A. Meyal, C.D. Wallace, Z. Dai, K.C. Carroll, M.R. Soltanian, Employing generative adversarial neural networks as surrogate model for reactive transport modeling in the hyporheic zone, J. Hydrol. 639 (2024) 131485, https://doi.org/10.1016/j.jhydrol.2024.131485.
[161] J. Bao, L. Li, A. Davis, Variational Autoencoder or Generative Adversarial Networks? A Comparison of Two Deep Learning Methods for Flow and Transport Data Assimilation, Math. Geosci. 54 (6) (2022) 1017–1042, https://doi.org/10.1007/s11004-022-10003-3.
[162] N. Janssens, M. Huysmans, R. Swennen, Computed Tomography 3D Super-Resolution with Generative Adversarial Neural Networks: Implications on Unsaturated and Two-Phase Fluid Flow, Materials 13 (6) (2020) 1397, https://doi.org/10.3390/ma13061397.
[163] J. Lopez-Alvis, E. Laloy, F. Nguyen, T. Hermans, Deep generative models in inversion: The impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder, Comput. & Geosci. 152 (2021) 104762, https://doi.org/10.1016/j.cageo.2021.104762.
[164] A.Y. Sun, Discovering State-Parameter Mappings in Subsurface Models Using Generative Adversarial Networks, Geophys. Res. Lett. 45 (20) (2018) 11,137–11,146, https://doi.org/10.1029/2018GL080404.
[165] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, Adv. Neural Inf. Process. Syst. 34 (2021) 852–863.
[166] A. Jabbar, X. Li, B. Omar, A Survey on Generative Adversarial Networks: Variants, Applications, and Training, Article 157. ACM Comput. Surv. 54 (8) (2021) https://doi.org/10.1145/3463475.

[167] C. Zhan, Z. Dai, M.R. Soltanian, X. Zhang, Stage-Wise Stochastic Deep Learning Inversion Framework for Subsurface Sedimentary Structure Identification, Geophys. Res. Lett. 49 (1) (2022) e2021GL095823, https://doi.org/10.1029/2021GL095823.
[168] C. Zhan, Z. Dai, J. Samper, S. Yin, R. Ershadnia, X. Zhang, Y. Wang, Z. Yang, X. Luan, M.R. Soltanian, An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network, J. Hydrol. 610 (2022) 127844, https://doi.org/10.1016/j.jhydrol.2022.127844.
[169] N. Zheng, Z. Li, X. Xia, S. Gu, X. Li, S. Jiang, Estimating line contaminant sources in non-Gaussian groundwater conductivity fields using deep learning-based framework, J. Hydrol. 630 (2024) 130727, https://doi.org/10.1016/j.jhydrol.2024.130727.
[170] X. Zhang, S. Jiang, J. Wei, C. Wu, X. Xia, X. Wang, N. Zheng, J. Xing, Non-gaussian hydraulic conductivity and potential contaminant source identification: A comparison of two advanced DLPM-based inversion framework, J. Hydrol. 638 (2024) 131540, https://doi.org/10.1016/j.jhydrol.2024.131540.
[171] X. Zhang, S. Jiang, N. Zheng, X. Xia, Z. Li, R. Zhang, J. Zhang, X. Wang, Integration of DDPM and ILUES for Simultaneous Identification of Contaminant Source Parameters and Non-Gaussian Channelized Hydraulic Conductivity Field, Water Resour. Res. 60 (9) (2024) e2023WR036893, https://doi.org/10.1029/2023WR036893.
[172] Z. Wang, W. Lu, Z. Chang, H. Wang, Simultaneous identification of groundwater contaminant source and simulation model parameters based on an ensemble Kalman filter – Adaptive step length ant colony optimization algorithm, J. Hydrol. 605 (2022) 127352, https://doi.org/10.1016/j.jhydrol.2021.127352.
[173] F.-K. Huang, Y.-F. Lin, H.-D. Yeh, G.S. Wang, Analytical framework for fast identification of hydrogeological boundaries and aquifer parameters in confined aquifers, J. Hydrol. 661 (2025) 133454, https://doi.org/10.1016/j.jhydrol.2025.133454.
[174] J. Qian, W. Wang, L. Ma, B. Dang, X. Sun, Identification of preferential flow paths by hydraulic tomography compared with tracer test and the groundwater contour map in coal mine water hazard area, J. Hydrol. 631 (2024) 130816, https://doi.org/10.1016/j.jhydrol.2024.130816.
[175] N. Zheng, S. Jiang, X. Xia, W. Kong, Z. Li, S. Gu, Z. Wu, Efficient estimation of groundwater contaminant source and hydraulic conductivity by an ILUES framework combining GAN and CNN, J. Hydrol. 621 (2023) 129677, https://doi.org/10.1016/j.jhydrol.2023.129677.
[176] Z. Wang, W. Lu, Z. Chang, T. Zhang, Joint identification of groundwater pollution source information, model parameters, and boundary conditions based on a novel ES-MDA with a wheel battle strategy, J. Hydrol. 636 (2024) 131320, https://doi.org/10.1016/j.jhydrol.2024.131320.
[177] Y. Xu, W. Lu, Z. Pan, C. Luo, Y. Bai, S. Qiu, Groundwater contaminant source identification considering unknown boundary condition based on an automated machine learning surrogate, Geosci. Front. 15 (1) (2024) 101732, https://doi.org/10.1016/j.gsf.2023.101732.
[178] J. Luo, Y. Ji, W. Lu, Comparison of Surrogate Models Based on Different Sampling Methods for Groundwater Remediation, J. Water Resour. Plan. Manag. 145 (5) (2019) 04019015, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001062.
[179] J. Luo, Y. Liu, X. Li, X. Xin, W. Lu, Inversion of groundwater contamination source based on a two-stage adaptive surrogate model-assisted trust region genetic algorithm framework, Appl. Math. Model. 112 (2022) 262–281, https://doi.org/10.1016/j.apm.2022.07.035.
[180] J. Luo, X. Ma, Y. Ji, X. Li, Z. Song, W. Lu, Review of machine learning-based surrogate models of groundwater contaminant modeling, Environ. Res. 238 (2023) 117268, https://doi.org/10.1016/j.envres.2023.117268.
[181] M. Wu, L. Wang, J. Xu, Z. Wang, P. Hu, H. Tang, Multiobjective ensemble surrogate-based optimization algorithm for groundwater optimization designs, J. Hydrol. 612 (2022) 128159, https://doi.org/10.1016/j.jhydrol.2022.128159.
[182] Z. Xing, R. Qu, Y. Zhao, Q. Fu, Y. Ji, W. Lu, Identifying the release history of a groundwater contaminant source based on an ensemble surrogate model, J. Hydrol. 572 (2019) 501–516, https://doi.org/10.1016/j.jhydrol.2019.03.020.
[183] Y. Liu, X. Chen, Z. Wang, J. Dong, Operator Inference for Physical and Generalized Surrogate Groundwater Modeling, Water Resour. Res. 62 (1) (2026) e2025WR039961, https://doi.org/10.1029/2025WR039961.
[184] L. Rice, E. Wong, Z. Kolter, Overfitting in adversarially robust deep learning. International conference on machine learning, PMLR, 2020, pp. 8093–8104.
[185] S. Razavi, B.A. Tolson, D.H. Burn, Review of surrogate modeling in water resources, Water Resour. Res. 48 (7) (2012), https://doi.org/10.1029/2011WR011527.
[186] N. Polette, O. Le Maître, P. Sochala, A. Gesret, Change of measure for Bayesian field inversion with hierarchical hyperparameters sampling, J. Comput. Phys. 529 (2025) 113888, https://doi.org/10.1016/j.jcp.2025.113888.
[187] S. Chaturantabut, D.C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput. 32 (5) (2010) 2737–2764, https://doi.org/10.1137/090766498.
[188] M. Gosses, W. Nowak, T. Wöhling, Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models, Adv. Water Resour. 115 (2018) 160–171, https://doi.org/10.1016/j.advwatres.2018.03.011.
[189] W. Fu, P. Liu, K. Zhang, J. Zhang, X. Chen, L. Zhang, X. Yan, Use deep transfer learning for efficient time-series updating of subsurface flow surrogate mode, Eng. Appl. Artif. Intell. 153 (2025) 110873, https://doi.org/10.1016/j.engappai.2025.110873.
[190] J. Li, W. Lu, J. Luo, Groundwater contamination sources identification based on the Long-Short Term Memory network, J. Hydrol. 601 (2021) 126670, https://doi.org/10.1016/j.jhydrol.2021.126670.
[191] Z. Hou, K. Zhao, S. Wang, Y. Wang, W. Lu, Bayesian hybrid-kernel machine-learning-assisted sensitivity analysis and sensitivity-relevant inverse modeling for groundwater DNAPL contamination, J. Hydrol. 633 (2024) 131009, https://doi.org/10.1016/j.jhydrol.2024.131009.
[192] S. Funk, A. Airoud Basmaji, U. Nackenhorst, Globally supported surrogate model based on support vector regression for nonlinear structural engineering applications, Arch. Appl. Mech. 93 (2) (2023) 825–839, https://doi.org/10.1007/s00419-022-02301-3.
[193] Z. Chang, Z. Guo, K. Chen, Z. Wang, Y. Zhan, W. Lu, C. Zheng, A Comparison of Inversion Methods for Surrogate-Based Groundwater Contamination Source Identification With Varying Degrees of Model Complexity, Water Resour. Res. 60 (4) (2024) e2023WR036051, https://doi.org/10.1029/2023WR036051.
[194] Y. Zhu, N. Zabaras, Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification, J. Comput. Phys. 366 (2018) 415–447, https://doi.org/10.1016/j.jcp.2018.04.018.
[195] S. Mo, Y. Zhu, N. Zabaras, X. Shi, J. Wu, Deep Convolutional Encoder-Decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Heterogeneous Media, Water Resour. Res. 55 (1) (2019) 703–728, https://doi.org/10.1029/2018WR023528.
[196] S. Mo, N. Zabaras, X. Shi, J. Wu, Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification, Water Resour. Res. 55 (5) (2019) 3856–3881, https://doi.org/10.1029/2018WR024638.
[197] S. Mo, N. Zabaras, X. Shi, J. Wu, Integration of Adversarial Autoencoders With Residual Dense Convolutional Networks for Estimation of Non-Gaussian Hydraulic Conductivities, Water Resour. Res. 56 (2) (2020) e2019WR026082, https://doi.org/10.1029/2019WR026082.
[198] X. Kang, A. Kokkinaki, P.K. Kitanidis, X. Shi, J. Lee, S. Mo, J. Wu, Hydrogeophysical Characterization of Nonstationary DNAPL Source Zones by Integrating a Convolutional Variational Autoencoder and Ensemble Smoother, Water Resour. Res. 57 (2) (2021) e2020WR028538, https://doi.org/10.1029/2020WR028538.
[199] T. He, N. Wang, D. Zhang, Theory-guided full convolutional neural network: An efficient surrogate model for inverse problems in subsurface contaminant transport, Adv. Water Resour. 157 (2021) 104051, https://doi.org/10.1016/j.advwatres.2021.104051.
[200] X. Xia, S. Jiang, N. Zhou, J. Cui, X. Li, Groundwater contamination source identification and high-dimensional parameter inversion using residual dense convolutional neural network, J. Hydrol. 617 (2023) 129013, https://doi.org/10.1016/j.jhydrol.2022.129013.
[201] Z. Zhou, D.M. Tartakovsky, Markov chain Monte Carlo with neural network surrogates: application to contaminant source identification, Stoch. Environ. Res. Risk Assess. 35 (3) (2021) 639–651, https://doi.org/10.1007/s00477-020-01888-9.
[202] J. Luo, X. Li, Y. Xiong, Y. Liu, Groundwater pollution source identification using Metropolis-Hasting algorithm combined with Kalman filter algorithm, J. Hydrol. 626 (2023) 130258, https://doi.org/10.1016/j.jhydrol.2023.130258.
[203] Z. Chang, W. Lu, Z. Wang, A differential evolutionary Markov chain algorithm with ensemble smoother initial point selection for the identification of groundwater contaminant sources, J. Hydrol. 603 (2021) 126918, https://doi.org/10.1016/j.jhydrol.2021.126918.
[204] J. Bian, D. Ruan, Y. Wang, X. Sun, Z. Gu, Bayesian ensemble machine learning-assisted deterministic and stochastic groundwater DNAPL source inversion with a homotopy-based progressive search mechanism, J. Hydrol. 624 (2023) 129925, https://doi.org/10.1016/j.jhydrol.2023.129925.
[205] S. Jiang, L.J. Durlofsky, Use of multifidelity training data and transfer learning for efficient construction of subsurface flow surrogate models, J. Comput. Phys. 474 (2023) 111800, https://doi.org/10.1016/j.jcp.2022.111800.
[206] J. Zhang, X. Liang, L. Zeng, X. Chen, E. Ma, Y. Zhou, Y.-K. Zhang, Deep transfer learning for groundwater flow in heterogeneous aquifers using a simple analytical model, J. Hydrol. 626 (2023) 130293, https://doi.org/10.1016/j.jhydrol.2023.130293.
[207] C. Wang, Z. Dou, Y. Zhu, Z. Yang, Z. Zou, Breaking the mold of simulation-optimization: Direct forward machine learning methods for groundwater contaminant source identification, J. Hydrol. 642 (2024) 131759, https://doi.org/10.1016/j.jhydrol.2024.131759.
[208] Z. Chang, W. Lu, Z. Wang, Study on source identification and source-sink relationship of LNAPLs pollution in groundwater by the adaptive cyclic improved iterative process and Monte Carlo stochastic simulation, J. Hydrol. 612 (2022) 128109, https://doi.org/10.1016/j.jhydrol.2022.128109.
[209] S. Bonvicini, G. Antonioni, V. Cozzani, Assessment of the risk related to environmental damage following major accidents in onshore pipelines, J. Loss Prev. Process Ind. 56 (2018) 505–516, https://doi.org/10.1016/j.jlp.2018.11.005.
[210] Y. Zhao, W. Lu, C. Xiao, A Kriging surrogate model coupled in simulation–optimization approach for identifying release history of groundwater sources, J. Contam. Hydrol. 185-186 (2016) 51–60, https://doi.org/10.1016/j.jconhyd.2016.01.004.
[211] H. Pan, Y. Li, J. Zhang, C. Cao, Y. Cheng, Y. Zhou, Y. Wang, S. Bai, J. Liu, Q. Jin, C. Gualtieri, Identifying urban river pollution sources from wet-weather discharges using an integrated deep learning and data assimilation approach, J. Hydrol. 661 (2025) 133797, https://doi.org/10.1016/j.jhydrol.2025.133797.

[212] L. Zhu, W. Lu, C. Luo, A high-precision and interpretability-enhanced direct inversion framework for groundwater contaminant source identification using multiple machine learning techniques, J. Hydrol. 659 (2025) 133237, https://doi.org/10.1016/j.jhydrol.2025.133237.
[213] G. Mahinthakumar, M. Sayeed, Hybrid Genetic Algorithm—Local Search Methods for Solving Groundwater Source Identification Inverse Problems, J. Water Resour. Plan. Manag. 131 (1) (2005) 45–57, https://doi.org/10.1061/(ASCE)0733-9496(2005)131:1(45).
[214] H.D. Yeh, T.H. Chang, Y.C. Lin, Groundwater contaminant source identification by a hybrid heuristic approach, Water Resour. Res. 43 (9) (2007) 2005WR004731, https://doi.org/10.1029/2005WR004731.
[215] D. Jia, L. Zhao, J. Song, D. Guo, X. Liu, Traceability of surface water pollution based on the SSO+DE algorithm, Alex. Eng. J. 114 (2025) 112–122, https://doi.org/10.1016/j.aej.2024.11.007.
[216] A. Anshuman, T.I. Eldho, Entity aware sequence to sequence learning using LSTMs for estimation of groundwater contamination release history and transport parameters, J. Hydrol. 608 (2022) 127662, https://doi.org/10.1016/j.jhydrol.2022.127662.
[217] J.J. Gómez-Hernández, T. Xu, Contaminant Source Identification in Aquifers: A Critical View, Math. Geosci. 54 (2) (2022) 437–458, https://doi.org/10.1007/s11004-021-09976-4.
[218] P.S. Mahar, B. Datta, Identification of Pollution Sources in Transient Groundwater Systems, Water Resour. Manag. 14 (3) (2000) 209–227, https://doi.org/10.1023/A:1026527901213.
[219] H.-T. Hwang, S.-W. Jeen, D. Kaown, S.-S. Lee, E.A. Sudicky, D.T. Steinmoeller, K.-K. Lee, Backward Probability Model for Identifying Multiple Contaminant Source Zones Under Transient Variably Saturated Flow Conditions, Water Resour. Res. 56 (4) (2020) e2019WR025400, https://doi.org/10.1029/2019WR025400.
[220] Y. Zhang, Backward Particle Tracking of Anomalous Transport in Multi-Dimensional Aquifers, Water Resour. Res. 58 (10) (2022) e2022WR032396, https://doi.org/10.1029/2022WR032396.
[221] L.D. Lemke, L.M. Abriola, P. Goovaerts, Dense nonaqueous phase liquid (DNAPL) source zone characterization: Influence of hydraulic property correlation on predictions of DNAPL infiltration and entrapment, Water Resour. Res. 40 (1) (2004), https://doi.org/10.1029/2003WR001980.
[222] F. Boano, R. Revelli, L. Ridolfi, Source identification in river pollution problems: A geostatistical approach, Water Resour. Res. 41 (7) (2005), https://doi.org/10.1029/2004WR003754.
[223] I. Butera, M.G. Tanda, A. Zanini, Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach, Stoch. Environ. Res. Risk Assess. 27 (5) (2013) 1269–1280, https://doi.org/10.1007/s00477-012-0662-1.
[224] G. Gzyl, A. Zanini, R. Frączek, K. Kura, Contaminant source and release history identification in groundwater: A multi-step approach, J. Contam. Hydrol. 157 (2014) 59–72, https://doi.org/10.1016/j.jconhyd.2013.11.006.
[225] E. Park, Manifold embedding in geostatistical inversion: Redefining optimality in subsurface characterization, J. Hydrol. 661 (2025) 133576, https://doi.org/10.1016/j.jhydrol.2025.133576.
[226] H. Yang, D. Shao, B. Liu, J. Huang, X. Ye, Multi-point source identification of sudden water pollution accidents in surface waters based on differential evolution and Metropolis–Hastings–Markov Chain Monte Carlo, Stoch. Environ. Res. Risk Assess. 30 (2) (2016) 507–522, https://doi.org/10.1007/s00477-015-1191-5.
[227] M.F. Snodgrass, P.K. Kitanidis, A geostatistical approach to contaminant source identification, Water Resour. Res. 33 (4) (1997) 537–546, https://doi.org/10.1029/96WR03753.
[228] J. Wang, N. Zabaras, A Markov random field model of contamination source identification in porous media flow, Int. J. Heat. Mass Transf. 49 (5) (2006) 939–950, https://doi.org/10.1016/j.ijheatmasstransfer.2005.09.016.
[229] A. Hazart, J.F. Giovannelli, S. Dubost, L. Chatellier, Contaminant source estimation in a two-layers porous environment using a Bayesian approach, IEEE Int. Geosci. Remote Sens. Symp. 2007 (2007) 4757–4760.
[230] A. Hazart, J.-F. Giovannelli, S. Dubost, L. Chatellier, Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC, Signal Process. 96 (2014) 346–361, https://doi.org/10.1016/j.sigpro.2013.08.013.
[231] G. Brunetti, J. Šimunek, T. Wöhling, C. Stumpp, An in-depth analysis of Markov-Chain Monte Carlo ensemble samplers for inverse vadose zone modeling, J. Hydrol. 624 (2023) 129822, https://doi.org/10.1016/j.jhydrol.2023.129822.
[232] T.-D. Nguyen, D.H. Nguyen, H.-H. Kwon, D.-H. Bae, A novel framework for uncertainty quantification of rainfall–runoff models based on a Bayesian approach focused on transboundary river basins, J. Hydrol. Reg. Stud. 57 (2025) 102095, https://doi.org/10.1016/j.ejrh.2024.102095.
[233] Y. Bai, W. Lu, J. Li, Z. Chang, H. Wang, Groundwater contamination source identification using improved differential evolution Markov chain algorithm, Environ. Sci. Pollut. Res. 29 (13) (2022) 19679–19692, https://doi.org/10.1007/s11356-021-17120-2.
[234] Y. Zhu, H. Cao, Z. Gao, Z. Chen, A DiffeRential Evolution Adaptive Metropolis (DREAM)-based inverse model for continuous release source identification in river pollution incidents: quantitative evaluation and sensitivity analysis, Environ. Pollut. 347 (2024) 123448, https://doi.org/10.1016/j.envpol.2024.123448.
[235] T. Cui, G. Detommaso, R. Scheichl, Multilevel dimension-independent likelihood-informed MCMC for large-scale inverse problems, Inverse Probl. 40 (3) (2024) 035005, https://doi.org/10.1088/1361-6420/ad1e2c.
[236] M.G. Rudolph, T. Wöhling, T. Wagener, A. Hartmann, Extending GLUE with multilevel methods to accelerate statistical inversion of hydrological models, Water Resour. Res. 60 (10) (2024) e2024WR037735, https://doi.org/10.1029/2024WR037735.
[237] L. Li, H. Zhou, J.J. Gómez-Hernández, H.-J. Hendricks Franssen, Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter, J. Hydrol. 428-429 (2012) 152–169, https://doi.org/10.1016/j.jhydrol.2012.01.037.
[238] D.H. Le, R. Younis, A.C. Reynolds, A history matching procedure for non-gaussian facies based on ES-MDA, SPE Reserv. Simul. Symp. (2015).
[239] Z. Cao, L. Li, K. Chen, Bridging iterative Ensemble Smoother and multiple-point geostatistics for better flow and transport modeling, J. Hydrol. 565 (2018) 411–421, https://doi.org/10.1016/j.jhydrol.2018.08.023.
[240] S. Yoon, S. Lee, J. Zhang, L. Zeng, P.K. Kang, Inverse estimation of multiple contaminant sources in three-dimensional heterogeneous aquifers with variable-density flows, J. Hydrol. 617 (2023) 129041, https://doi.org/10.1016/j.jhydrol.2022.129041.
[241] J. Zhang, G. Lin, W. Li, L. Wu, L. Zeng, AN Iterative Local Updating Ensemble Smoother for Estimation and Uncertainty Assessment of Hydrologic Model Parameters with Multimodal Distributions, Water Resour. Res. 54 (3) (2018) 1716–1733, https://doi.org/10.1002/2017WR020906.
[242] Z. Pan, W. Lu, H. Wang, Y. Bai, Groundwater contaminant source identification based on an ensemble learning search framework associated with an auto xgboost surrogate, Environ. Model. Softw. 159 (2023) 105588, https://doi.org/10.1016/j.envsoft.2022.105588.
[243] Z. Wang, W. Lu, Z. Chang, Joint inverse estimation of groundwater pollution source characteristics and model parameters based on an intelligent particle filter, J. Hydrol. 625 (2023) 129965, https://doi.org/10.1016/j.jhydrol.2023.129965.
[244] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.
[245] Q. He, A.M. Tartakovsky, Physics-Informed Neural Network Method for Forward and Backward Advection-Dispersion Equations, Water Resour. Res. 57 (7) (2021) e2020WR029479, https://doi.org/10.1029/2020WR029479.
[246] Z. Miao, Y. Chen, VC-PINN: variable coefficient physics-informed neural network for forward and inverse problems of PDEs with variable coefficient, Physica D Nonlinear Phenomena 456 (2023) 133945, https://doi.org/10.1016/j.physd.2023.133945.
[247] S. Lin, Y. Chen, Gradient-enhanced physics-informed neural networks based on transfer learning for inverse problems of the variable coefficient differential equations, Physica D Nonlinear Phenomena 459 (2024) 134023, https://doi.org/10.1016/j.physd.2023.134023.
[248] J. Li, A.M. Tartakovsky, Physics-informed Karhunen-Loéve and neural network approximations for solving inverse differential equation problems, J. Comput. Phys. 462 (2022) 111230, https://doi.org/10.1016/j.jcp.2022.111230.
[249] S. Wang, H. Zhang, X. Jiang, Physics-informed neural network algorithm for solving forward and inverse problems of variable-order space-fractional advection–diffusion equations, Neurocomputing 535 (2023) 64–82, https://doi.org/10.1016/j.neucom.2023.03.032.
[250] N. Wang, H. Chang, D. Zhang, Deep-learning-based inverse modeling approaches: a subsurface flow example, J. Geophys. Res. Solid Earth 126 (2) (2021) e2020JB020549, https://doi.org/10.1029/2020JB020549.
[251] Y. Zhan, Z. Guo, B. Yan, K. Chen, Z. Chang, V. Babovic, C. Zheng, Physics-informed identification of PDEs with LASSO regression, examples of groundwater-related equations, J. Hydrol. 638 (2024) 131504, https://doi.org/10.1016/j.jhydrol.2024.131504.
[252] Z.-W. Ke, S.-J. Wei, S.-Y. Yao, S. Chen, Y.-M. Chen, Y.-C. Li, Pre-trained Physics-Informed Neural Networks for Analysis of Contaminant Transport in Soils, Comput. Geotech. 180 (2025) 107055, https://doi.org/10.1016/j.compgeo.2025.107055.
[253] Z. Jiao, X. Zhu, G. Xiong, S. Mo, Y. Meng, J. Wu, J. Wu, An Efficient Multi-Physics GPT-PINN Framework for Predicting Reactive Solute Transport in Parameterized Groundwater Systems, Geophys. Res. Lett. 53 (3) (2026) e2025GL120217, https://doi.org/10.1029/2025GL120217.
[254] F.V. Difonzo, L. Lopez, S.F. Pellegrino, Physics informed neural networks for an inverse problem in peridynamic models, Eng. Comput. (2024), https://doi.org/10.1007/s00366-024-01957-5.
[255] X. Liu, W. Yao, W. Peng, W. Zhou, Bayesian physics-informed extreme learning machine for forward and inverse PDE problems with noisy data, Neurocomputing 549 (2023) 126425, https://doi.org/10.1016/j.neucom.2023.126425.
[256] L. Zhang, J. Liu, D. Mei, X. Qiao, Z. Xie, N. Sun, Z. Ding, Z. Zhang, X. Peng, Physics-informed deep learning for groundwater contamination sources identification under sparse monitoring, J. Hydrol. 665 (2026) 134691, https://doi.org/10.1016/j.jhydrol.2025.134691.
[257] I. Chuprov, D. Derkach, D. Efremenko, A. Kychkin, Application of Physics-Informed Neural Networks for Solving the Inverse Advection-Diffusion Problem to Localize Pollution Sources, Comput. Sci. (2025), https://doi.org/10.48550/arXiv.2503.18849.
[258] F. Deng, J. Wu, Y. Yang, J. Li, X. Xie, J. Jiang, X. Zhu, Improved physics-informed neural network for reactive transport modeling of groundwater arsenic enrichment, Sci. China Earth Sci. 68 (9) (2025) 2781–2796, https://doi.org/10.1007/s11430-024-1590-x.
[259] T. He, H. Chang, D. Zhang, Identification of physical processes and unknown parameters of 3D groundwater contaminant problems via theory-guided U-net, Stoch. Environ. Res. Risk Assess. 38 (3) (2024) 869–900, https://doi.org/10.1007/s00477-023-02604-z.

[260] T. Praditia, M. Karlbauer, S. Otte, S. Oladyshkin, M.V. Butz, W. Nowak, Learning Groundwater Contaminant Diffusion-Sorption Processes With a Finite Volume Neural Network, Water Resour. Res. 58 (12) (2022) e2022WR033149, https://doi.org/10.1029/2022WR033149.
[261] Z. Zhou, X. Yan, C. Hu, Two-Stage Bayesian Physics-Informed Neural Networks for Groundwater Contaminant Source Identification, J. Water Resour. Plan. Manag. 152 (2) (2026) 04025075, https://doi.org/10.1061/JWRMD5.WRENG-7084.
[262] Z. Pan, W. Lu, Z. Chang, H. wang, Simultaneous identification of groundwater pollution source spatial–temporal characteristics and hydraulic parameters based on deep regularization neural network-hybrid heuristic algorithm, J. Hydrol. 600 (2021) 126586, https://doi.org/10.1016/j.jhydrol.2021.126586.
[263] C. Wang, S. Majdalani, V. Guinot, H. Jourde, Solute transport in dual conduit structure: Effects of aperture and flow rate, J. Hydrol. 613 (2022) 128315, https://doi.org/10.1016/j.jhydrol.2022.128315.
[264] S. Yang, F.T.-C. Tsai, P. Bacopoulos, C.E. Kees, Comparative analyses of covariance matrix adaptation and iterative ensemble smoother on high-dimensional inverse problems in high-resolution groundwater modeling, J. Hydrol. 625 (2023) 130075, https://doi.org/10.1016/j.jhydrol.2023.130075.
[265] J. Kang, S. Mo, Xueyuan Kang, J. Dang, C. Cheng, P. Xu, X. Shi, Frontier advances and challenges of machine learning in groundwater science, EarthScienceFrontiers 33 (01) (2026) 483–499, https://doi.org/10.13745/j.esf.sf.2025.10.19.
[266] K.C. Carroll, R. Taylor, E. Gray, M.L. Brusseau, The impact of composition on the physical properties and evaporative mass transfer of a PCE–diesel immiscible liquid, J. Hazard. Mater. 164 (2) (2009) 1074–1081, https://doi.org/10.1016/j.jhazmat.2008.09.003.
[267] L. Yan, T. Zhou, Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems, J. Comput. Phys. 381 (2019) 110–128, https://doi.org/10.1016/j.jcp.2018.12.025.
[268] Z. Dai, C. Zhan, M.R. Soltanian, R.W. Ritzi, X. Zhang, Identifying spatial correlation structure of multimodal permeability in hierarchical media with Markov chain approach, J. Hydrol. 568 (2019) 703–715, https://doi.org/10.1016/j.jhydrol.2018.11.032.
[269] C. Zhan, Z. Dai, J.J. Jiao, M.R. Soltanian, H. Yin, K.C. Carroll, Toward Artificial General Intelligence in Hydrogeological Modeling With an Integrated Latent Diffusion Framework, Geophys. Res. Lett. 52 (3) (2025) e2024GL114298, https://doi.org/10.1029/2024GL114298.
[270] Q. Wang, J. Bian, E. Ma, J. Zhang, Predicting sorption of organic pollutants on soils with interpretable machine learning, Environ. Pollut. 382 (2025) 126665, https://doi.org/10.1016/j.envpol.2025.126665.
[271] D. Duan, P. Wang, X. Rao, J. Zhong, M. Xiao, F. Huang, R. Xiao, Identifying interactive effects of spatial drivers in soil heavy metal pollutants using interpretable machine learning models, Sci. Total Environ. 934 (2024) 173284, https://doi.org/10.1016/j.scitotenv.2024.173284.
[272] Q. Liu, D. Gui, L. Zhang, J. Niu, H. Dai, G. Wei, B.X. Hu, Simulation of regional groundwater levels in arid regions using interpretable machine learning models, Sci. Total Environ. 831 (2022) 154902, https://doi.org/10.1016/j.scitotenv.2022.154902.
[273] Z. Wang, K. Chen, J. Wang, Enhancing surrogate assisted optimization with SHAP guided two-stage sampling, Environ. Model. Softw. 195 (2026) 106755, https://doi.org/10.1016/j.envsoft.2025.106755.
[274] T.S. Narany, M.F. Ramli, K. Fakharian, A.Z. Aris, W.N.A. Sulaiman, Multi-objective based approach for groundwater quality monitoring network optimization, Water Resour. Manag. 29 (14) (2015) 5141–5156, https://doi.org/10.1007/s11269-015-1109-5.
[275] V. Gómez-Escalonilla, E. Montero-González, S. Díaz-Alcaide, M. Martín-Loeches, M.R. del Rosario, P. Martínez-Santos, A machine learning approach to site groundwater contamination monitoring wells, Appl. Water Sci. 14 (12) (2024) 250, https://doi.org/10.1007/s13201-024-02320-1.
[276] H. Gamaleldien, L.-G. Wu, H.K.H. Olierook, C.L. Kirkland, U. Kirscher, Z.-X. Li, T. E. Johnson, S. Makin, Q.-L. Li, Q. Jiang, S.A. Wilde, X.-H. Li, Onset of the Earth's hydrological cycle four billion years ago or earlier, Nat. Geosci. 17 (6) (2024) 560–565, https://doi.org/10.1038/s41561-024-01450-0.
[277] D.D.S. Barcellos, F.T.D. Souza, Optimization of water quality monitoring programs by data mining, Water Res. 221 (2022) 118805, https://doi.org/10.1016/j.watres.2022.118805.
[278] Z. Fang, H. Ke, Y. Ma, S. Zhao, R. Zhou, Z. Ma, Z. Liu, Design optimization of groundwater circulation well based on numerical simulation and machine learning, Sci. Rep. 14 (1) (2024) 11506, https://doi.org/10.1038/s41598-024-62545-7.
[279] C. Shen, Y. Song, M. Clark, J. Liu, J. Halgren, K. Lawson, 2024, Prominent impacts of hydrologic scaling laws on climate risks, (2024). doi: 10.21203/rs.3.rs-4584048/v1.
[280] A.O. Meray, S. Sturla, M.R. Siddiquee, R. Serata, S. Uhlemann, H. Gonzalez-Raymat, M. Denham, H. Upadhyay, L.E. Lagos, C. Eddy-Dilek, PyLEnM: a machine learning framework for long-term groundwater contamination monitoring strategies, Environ. Sci. Technol. 56 (9) (2022) 5973–5983, https://doi.org/10.1021/acs.est.1c07440.
[281] M. Troldborg, W. Nowak, N. Tuxen, P.L. Bjerg, R. Helmig, P.J. Binning, Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework, Water Resour. Res. 46 (12) (2010), https://doi.org/10.1029/2010WR009227.
[282] C. Zhang, Z. Zhu, Y. Li, E. Du, Y. Sun, Z. Liu, Pollution source detection with low-cost low-accuracy sensors through coupling forward data assimilation and inverse optimization, Water Resour. Res. 60 (11) (2024) e2023WR036834, https://doi.org/10.1029/2023WR036834.
[283] X. Kang, A. Kokkinaki, P.K. Kitanidis, X. Shi, A. Revil, J. Lee, A. Soueid Ahmed, J. Wu, Improved characterization of DNAPL source zones via sequential hydrogeophysical inversion of hydraulic-head, self-potential and partitioning tracer data, Water Resour. Res. 56 (8) (2020) e2020WR027627, https://doi.org/10.1029/2020WR027627.
[284] T. Xu, J.J. Gómez-Hernández, Characterization of non-Gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures, Water Resour. Res. 52 (8) (2016) 6111–6136, https://doi.org/10.1002/2016WR019011.
[285] J. Zhang, C. Cao, T. Nan, L. Ju, H. Zhou, L. Zeng, A novel deep learning approach for data assimilation of complex hydrological systems, Water Resour. Res. 60 (2) (2024) e2023WR035389, https://doi.org/10.1029/2023WR035389.
[286] J. Zhang, Q. Zheng, L. Wu, L. Zeng, Using deep learning to improve ensemble smoother: applications to subsurface characterization, Water Resour. Res. 56 (12) (2020) e2020WR027399, https://doi.org/10.1029/2020WR027399.
[287] W. Ling, B. Jafarpour, Improving the parameterization of complex subsurface flow properties with style-based generative adversarial network (StyleGAN), Water Resour. Res. 60 (11) (2024) e2024WR037630, https://doi.org/10.1029/2024WR037630.
[288] M. Sun, Q. Luo, Y. Yang, T. Nan, J. Zhang, L. Ma, Y. Li, H. Ma, M. Lei, Y. Deng, J. Qian, Enhancing Aquifer Characterization With Position-Encoded Hyperparameters: A Novel ES-SIFG Approach, Water Resour. Res. 61 (6) (2025) e2024WR038468, https://doi.org/10.1029/2024WR038468.
[289] T. Xu, A.J. Valocchi, M. Ye, F. Liang, Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model, Water Resour. Res. 53 (5) (2017) 4084–4105, https://doi.org/10.1002/2016WR019831.
[290] A.H. Hosseini, C.V. Deutsch, C.A. Mendoza, K.W. Biggar, Inverse modeling for characterization of uncertainty in transport parameters under uncertainty of source geometry in heterogeneous aquifers, J. Hydrol. 405 (3) (2011) 402–416, https://doi.org/10.1016/j.jhydrol.2011.05.039.
[291] Y. Zhao, R. Qu, Z. Xing, W. Lu, Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms, Adv. Water Resour. 138 (2020) 103540, https://doi.org/10.1016/j.advwatres.2020.103540.
[292] L. Zhu, W. Lu, A quantum-inspired attention integrated scalar long short-term memory model for accurate and stable groundwater contaminant source inversion, Environ. Monit. Assess. 198 (2) (2026) 124, https://doi.org/10.1007/s10661-025-14972-w.
[293] J. Kath, J.K. Golden, A.G. Percus, D. O'Malley, Predicting flow in fracture networks with quantum algorithms, Water Resour. Res. 61 (12) (2025) e2024WR039784, https://doi.org/10.1029/2024WR039784.