Advances in Water Resources 207 (2026) 105204

Contents lists available at ScienceDirect

Advances in Water Resources

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

Impact of observation and surrogate-model noises on deep learning-based subsurface heterogeneous structure identification through monitoring network optimization

Yuzhou Xia[a], Chuanjun Zhan[b,c,*], Zhenxue Dai[a,b,*], Jichun Wu[d], Xiaoying Zhang[a], Huichao Yin[e], Jiahe Yan[f], Junjun Chen[g,h], Zihao Wang[a], Mohamad Reza Soltanian[i], Kenneth C. Carroll[e]

[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, 266520, China*
[c] *Department of Earth Sciences, The University of Hong Kong, Hong Kong*
[d] *Key Laboratory of Surficial Geochemistry of Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China*
[e] *Department of Plant & Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA*
[f] *College of Geo-Exploration Science and Technology, Jilin University, Changchun, 130026, China*
[g] *National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou, China*
[h] *Physical Science and Engineering (PSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia*
[i] *Departments of Geosciences and Environmental Engineering, University of Cincinnati, Cincinnati, OH, USA*

ARTICLE INFO

Keywords:
Subsurface structure identification
Deep learning
Monitoring network optimization
Multisource data
Data worth
Data noise

ABSTRACT

Accurate identification of subsurface structures is essential for enhancing geoscientific modeling and monitoring, thereby deepening the understanding of earth system. Although recent advances in deep learning have greatly enhanced the computational efficiency of structure identification, the identification accuracy can still be degraded by the spatial stochasticity of monitoring locations, observation noise and surrogate model noise. To address these challenges, this study develops a transition probability and entropy-based monitoring network optimization method, which can determine monitoring locations with high data worth. Based on this method, a synthetic contaminant transport experiment was conducted to obtain dynamic observations of hydraulic head, solute concentration, and electrical resistivity tomography (ERT). The impacts of observation and model noise on inversion results for different data types were then quantitatively analyzed. The results indicate that ERT data exhibit strong noise resistance, with average identification accuracy reductions of only 1.59% and 2.24% under 15% observation noise and model noise, respectively. Moreover, incorporating ERT data into data fusion scenarios enhances the robustness of other observation types, particularly solute concentration data, for which the accuracy decrease is limited to 1.24% even under the highest observation noise level. Model noise not only reduces the accuracy of structure identification but also increases uncertainty, and this negative effect is further amplified in data fusion scenarios.

1. Introduction

Accurate identification of subsurface structures is crucial for the sustainable management of groundwater and subsurface resources (Aeschbach-Hertig and Gleeson, 2012; Gleeson et al., 2020; Kuang et al., 2024), enabling pollution prevention (Ferrari et al., 2015), efficient remediation (Barati Moghaddam et al., 2021), and safe geological storage of carbon, hydrogen, and radioactive waste (Xu et al., 2025). Coupled hydrogeophysical inversion within a data assimilation (DA) framework is an important method for identifying heterogeneous subsurface structures (Camporese et al., 2015; Hinnell et al., 2010). The advantages of this method can be summarized as follows: (1) compared with modeling methods that rely on costly lithological observations (Emery, 2008), coupled inversion effectively utilizes dynamic observations to update model parameters (Berg and Illman, 2011; Mo et al., 2020; Xu and Gómez-Hernández, 2016; Yeh and Liu, 2000; Zhu and Yeh,

---

* Corresponding authors.
E-mail addresses: zhanchuanjun@qut.edu.cn (C. Zhan), dzx@jlu.edu.cn (Z. Dai).

https://doi.org/10.1016/j.advwatres.2025.105204
Received 16 November 2025; Received in revised form 29 December 2025; Accepted 29 December 2025

2005). (2) by jointly solving hydrological and geophysical governing equations, this method enables the synergistic inversion of multisource dynamic observations. Unlike conventional methods that separately calibrate hydrological and geophysical models, this mechanism not only reduces reliance on smoothness regularization constraints (Fernández-Muñiz et al., 2019; Jamil et al., 2024; Kang et al., 2018; Tsai et al., 2022; Tso et al., 2024; Yang et al., 2014), but also maximizes the data worth of different observation types, leading to more consistent inversion results (Irving and Singha, 2010; McClymont et al., 2011; Pollock and Cirpka, 2012; Tran et al., 2014). However, not all types of observations contribute to reliable subsurface structure characterization and dynamic response estimation (Dausman et al., 2010; Schilling et al., 2019; Xu and Gómez-Hernández, 2016). Therefore, when designing data collection strategies, priority should be given to observations that are closely related to the target dynamic response to ensure accurate inversion results (de Barros et al., 2012; Zhan et al., 2022b). Among various geophysical techniques, electrical resistivity tomography (ERT) is widely applied in hydrogeophysical studies due to its ability to provide high-resolution subsurface resistivity distributions (Binley et al., 2015; Tso et al., 2020). ERT technology is particularly effective in time-lapse monitoring, making it well-suited for tracking the spatiotemporal evolution of contaminant plumes (Aghasi et al., 2013; Binley et al., 2002; Cheng et al., 2019; Johnson et al., 2012, 2010; McLachlan et al., 2020; Pollock and Cirpka, 2010; Power et al., 2015; Tso et al., 2021). Fig. 1 presents the multisource observations collected through an ERT-monitored tracer test. Specifically, lithological observations are collected from well log and borehole core, a sparsely distributed monitoring network measures solute concentration and hydraulic head, and densely arranged surface electrodes capture dynamic changes in subsurface resistivity.

However, in practical applications, hydrogeophysical coupled inversion method often incurs high computational costs due to the repeated invocation of forward models (Panzeri et al., 2013; Zhang et al., 2020). This challenge becomes even more pronounced in complex scenarios where both hydrogeological and geophysical forward models must be solved concurrently. For instance, the discretization of high-resolution grids with tens of thousands of dimensions can lead to individual iterations requiring several hours and substantial storage capacity (Zhan et al., 2022a; Zhu and Zabaras, 2018). Over the past decade, the widespread adoption of deep learning (DL) has led to exciting advancements in addressing these challenges. DL methods leverages its ability to handle highly nonlinear mappings and extract features from high-dimensional data to develop data-driven surrogate models, thereby approximating the input-output relationship at a significantly lower computational cost (Chan and Elsheikh, 2019; Mo et al., 2020; Siade et al., 2020; Xiao et al., 2021; Xu et al., 2025). DL-based parameterization methods can represent high-dimensional parameters using low-dimensional latent vectors, allowing for customizable computational dimensions while preserving the ability to refine local heterogeneities (Canchumuni et al., 2017; Kang et al., 2021b; Laloy et al., 2018, 2017; Zhang et al., 2019).

In addition to developing efficient inversion algorithms and advanced deep learning models, data worth plays a crucial role in determining the accuracy of inversion results (JafarGandomi and Binley, 2013). For hydraulic head and solute concentration data collected through invasive monitoring, multiple measurements at different times and locations are necessary to ensure that their data worth is sufficient to capture hydrological and geochemical dynamics. However, blindly expanding monitoring sites in practical applications poses dual risks: (1) the deployment and maintenance costs of intrusive sampling equipment grow nonlinearly with the number of monitoring locations (Jackson and Jin, 2005; Yeh and Zhu, 2007); (2) intrusive monitoring methods such as borehole drilling may alter the pore water flow field, leading to distorted plume shapes in solute transport simulations (Dai et al., 2022; Danquigny et al., 2004). To address these challenges, constructing an economical and informative monitoring network has become a key approach to enhancing data worth. (Azadi et al., 2020; Gharasoo et al., 2019; Hosseini and Kerachian, 2017). Among the existing methods, entropy-based optimization frameworks are particularly prominent due to their broad applicability to non-Gaussian distributions and nonlinear systems (Alfonso et al., 2010a). The workflow typically involves establishing candidate monitoring locations (CMLs), estimating entropy via Monte Carlo simulations, and applying the maximum information and minimum redundancy (MIMR) criterion (Chen et al., 2022; Li et al., 2012). While the MIMR criterion has been successfully matured in surface hydrology domains, including river basin, precipitation, and soil moisture monitoring (Kornelsen and Coulibaly, 2015; Mishra and Coulibaly, 2010; Salman et al., 2021; Samuel et al., 2013; Wang et al., 2023), extending this framework to subsurface heterogeneous structures remains fundamentally challenging. A critical challenge is generating a

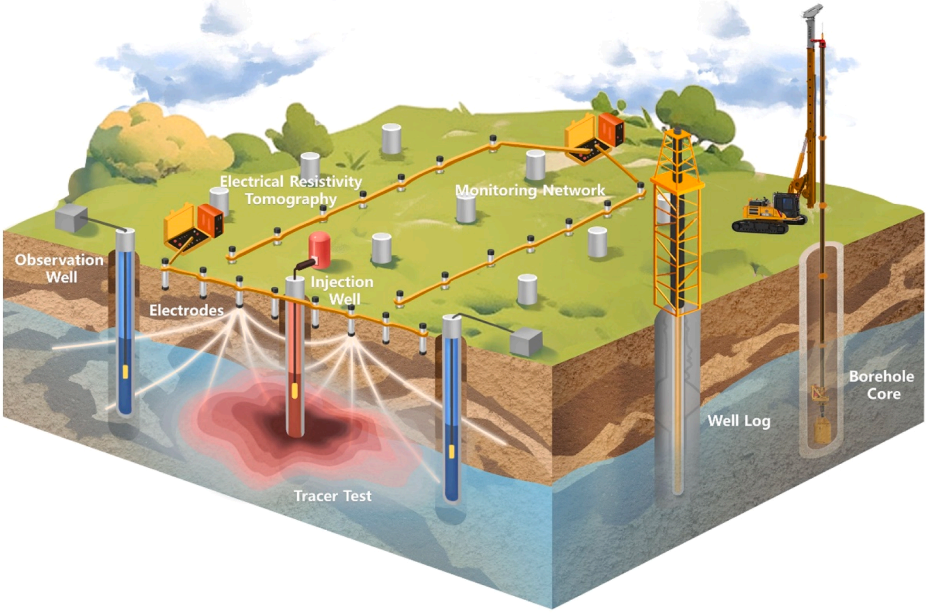


Fig. 1. A schematic representation of the ERT-monitored tracer test.

dataset that accurately quantifies dynamic response uncertainty based on available lithological observations. Furthermore, the primary obstacle lies in the curse of dimensionality: structure heterogeneity involves high-dimensional parameter spaces, requiring massive Monte Carlo simulations to accurately estimate the probability density functions for entropy calculation. Using traditional numerical models for such extensive sampling is often computationally expensive, creating a significant barrier to applying entropy-based optimization in complex groundwater systems.

Data noise (or error) is crucial in inversion studies. Improper noise settings may lead to data overfitting or underfitting, affecting the reliability of inversion results (Tso et al., 2017). In fact, data noise consists of both observation noise and model noise, which correspond to aleatoric uncertainty and epistemic uncertainty in statistical literature, respectively. Observation noise is affected by various uncontrollable factors, including the accuracy of measuring instruments, the skill level of technicians, and interference sources in the field environment. These uncertainties accumulate during data acquisition and propagate through various stages of hydrogeophysical studies, ultimately affecting the stability of inversion results (Huisman et al., 2010; Linde et al., 2015). Model noise arises from errors introduced during forward modeling, such as simplifications in subsurface structure, parameters, boundary conditions, numerical discretization, and physicochemical processes (Binley, 2015). While these simplifications improve computational feasibility, they may also cause model outputs to deviate from actual physical processes. In current studies that integrate hydrogeophysical multisource data to enhance subsurface imaging, observation noise is typically manually set as Gaussian noise at a predefined level, while the impact of model noise is often overlooked (Deng et al., 2024; Han et al., 2022; Kang et al., 2021a, 2021b; Laloy et al., 2018; Lopez-Alvis et al., 2022; Pollock and Cirpka, 2010, 2012). However, with the introduction of data-driven DL surrogate models, the impact of model noise may become more pronounced. DL surrogate models are designed to reduce the computational cost of repeatedly invoking high-fidelity numerical models. Since their training datasets are derived from numerical simulations, these models enhance computational efficiency but inevitably introduce additional sources of error. In this study, model noise specifically refers to the approximation error introduced by replacing the physical forward model with a trained DL surrogate model. Currently, research on DL surrogate models primarily focuses on optimizing network architectures and tuning hyperparameters to maximize prediction accuracy (Chan and Elsheikh, 2019; Mo et al., 2020; Siade et al., 2020; Xiao et al., 2021; Zhan et al., 2022a). However, systematically quantifying and compensating for the model noise introduced by DL models remains an underexplored research area.

This study is structured into three integrated phases as shown in Fig. 2. The first phase focuses on monitoring network optimization, where we propose a strategy to obtain high-worth dynamic observations. This strategy integrates transition probability (TP) stochastic simulation with a DL surrogate model to explicitly resolve the difficulty of generating representative datasets from sparse lithological data and to overcome the high computational costs of massive Monte Carlo simulations. The second phase involves coupled hydrogeophysical inversion. Utilizing the ILUES data assimilation framework, 21 different data fusion scenarios are designed to examine the impact of combining various hydrological datasets with ERT data. Finally, six optimal data fusion strategies are selected, considering varying levels of observation noise and model noise, resulting in a total of 96 scenarios to systematically investigate noise impact mechanisms.

2. Methods

2.1. Transition probability and entropy-based monitoring network optimization

2.1.1. Entropy theory

Marginal entropy represents the expected information content of an unknown random variable and serves as a measure of its uncertainty (Li et al., 2012; Shannon, 1948). Suppose $\boldsymbol{C} = \left\{X^q_{C_1}, X^q_{C_2}, \ldots X^q_{C_M}\right\}$ is a set of dynamic response data of type q collected from M CMLs. The marginal entropy of a single random variable X^q is defined as:

$$H(X^q) = -\sum_{i=1}^{n} p\left(x_i^q\right)\log_2 p\left(x_i^q\right) \tag{1}$$

where $p\left(x_i^q\right)$ denotes the probability of the event $X^q = x_i^q$, with $0 \le p\left(x_i^q\right) \le 1$. n represents the number of possible values of X. In this study, x_i^q represents a dynamic response variable (e.g., hydraulic head or solute concentration) corresponding to a specific subsurface structure at

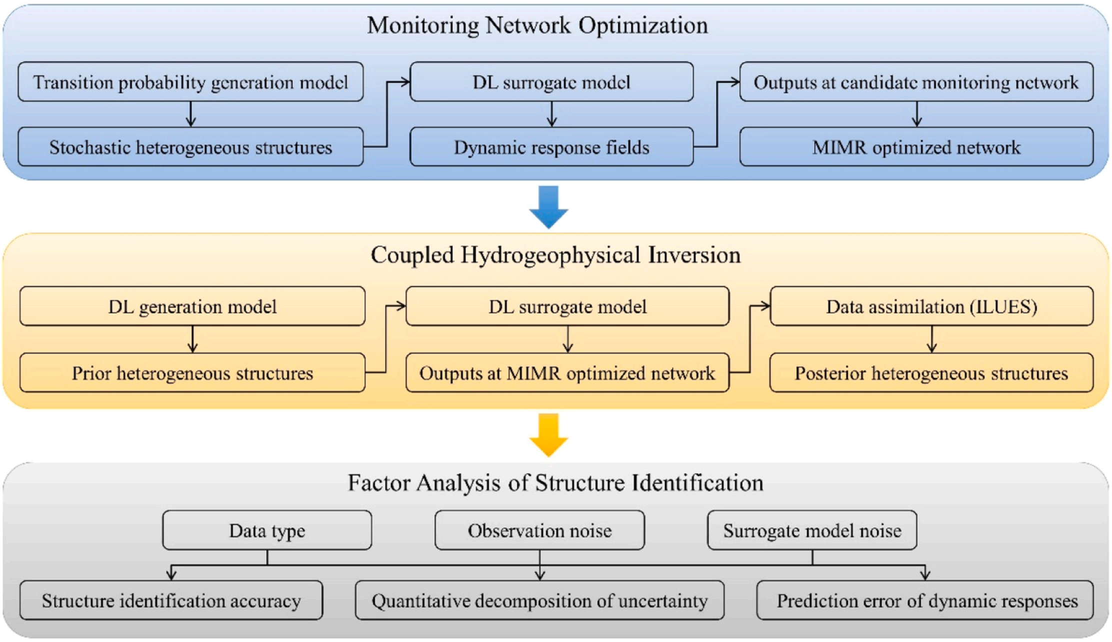


Fig. 2. Workflow of the integrated framework for monitoring network optimization, coupled hydrogeophysical inversion, and factor analysis.

a given CMLs. The value of n depends on the number of iterations. The unit of entropy is the bit, the fundamental unit of binary information (Keum et al., 2017). A higher marginal entropy indicates greater information content and higher uncertainty in a single random variable. When all events occur with equal probability, the marginal entropy reaches its maximum value, given by $H_{max}(X^q) = \log_2 n$.

The total information contained in multiple random variables $\{X_1^q, X_2^q, \cdots, X_N^q\}$ within a system is described by joint entropy:

$$H(X_1^q, \cdots, X_N^q) = -\sum_{i_1=1}^{n_1} \cdots \sum_{i_N=1}^{n_N} p\left(x_{1,i_1}^q, \cdots, x_{N,i_N}^q\right) \log_2 p\left(x_{1,i_1}^q, \cdots, x_{N,i_N}^q\right) \tag{2}$$

where $p\left(x_{1,i_1}^q, \cdots, x_{N,i_N}^q\right)$ is the probability of the event $X_1^q, \cdots, X_N^q = x_{1,i_1}^q, \cdots, x_{N,i_N}^q$, $n_1 \cdots n_N$ represents the total number of discrete intervals of $X_1^q, \cdots, X_N^q$. A detailed explanation of data discretization is provided in Section 2.1.2.

Transinformation is used to measure the shared information between response data (X_1^q, X_2^q) from two monitoring locations. It can be further understood as the amount of information in X_2^q that can be inferred based on the known value of X_1^q. The transformation between (X_1^q, X_2^q) is defined as:

$$T(X_1^q, X_2^q) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} p\left(x_{1,i_1}^q, x_{2,i_2}^q\right) \log_2 \frac{p\left(x_{1,i_1}^q, x_{2,i_2}^q\right)}{p\left(x_{1,i_1}^q\right) p\left(x_{2,i_2}^q\right)} \tag{3}$$

For a monitoring network system consisting of m CMLs ($m \leq M$), denoted as $\boldsymbol{C}' = \left\{X_{C'_1}^q, X_{C'_2}^q, \ldots X_{C'_m}^q\right\}$, the redundant information between different CMLs is measured using total correlation:

$$Tc(\boldsymbol{C}') = \sum_{i=1}^{m} H\left(X_{C'_i}^q\right) - H\left(X_{C'_1}^q, X_{C'_2}^q, \ldots X_{C'_m}^q\right) \tag{4}$$

Fig. 3 illustrates the topological implications of marginal entropy, joint entropy, transformation, and total correlation. Taking a specific type of response data as an example, suppose a monitoring network consists of three CMLs: X_1, X_2, and X_3. In Fig. 3a, the yellow, green, and blue regions represent the marginal entropy of the three variables, respectively. In Fig. 3b, the brown region represents their joint entropy. It is evident that the sum of the marginal entropy of each CML should be greater than the joint entropy of all CMLs, i.e., $\sum_{i=1}^{3} H(X_i) \geq H(X_1, X_2, X_3)$. The total correlation in Eq. (4) is also derived based on this concept (see Fig. 3d, where the gray region represents the total correlation of data across the three CMLs). In Fig. 3c, the yellow-green gradient region and the yellow-green-blue gradient region represent the transformation between (X_1, X_2) and (X_1, X_2, X_3), respectively.

To obtain a monitoring network with maximum joint entropy, maximum transinformation, and minimum total correlation, the MIMR objective function is used to balance different entropy terms during network design. Suppose the selected monitoring locations are a subset $\boldsymbol{S} = \left\{X_{S_1}^q, X_{S_2}^q, \ldots X_{S_k}^q\right\}$ of $\boldsymbol{C}$ and satisfy $H(\boldsymbol{S}) \leq H(\boldsymbol{C})$. Within $\boldsymbol{S}$, the variable $X_{S_1}^q$ with the highest marginal entropy is selected first. Subsequent selections of $X_{S_k}^q (k \geq 2)$ aim to maximize $H\left(X_{S_1}^q, X_{S_2}^q, \ldots X_{S_k}^q\right)$. The unselected monitoring locations are denoted as $\boldsymbol{F} = \left\{X_{F_1}^q, X_{F_2}^q, \ldots X_{F_l}^q\right\}$ with $l + k = M$, and they should share as much transformation as possible with $\boldsymbol{S}$. Finally, the process of selecting CMLs from $\boldsymbol{F}$ to supplement $\boldsymbol{S}$ is achieved by maximizing the MIMR objective function:

$$MIMR_{max} = \lambda_1 \left(H\left(X_{S_1}^q, X_{S_2}^q, \ldots X_{S_k}^q\right) + \sum_{i_{F_i}=1}^{n_{F_i}} T\left(X_{S_1}^q, X_{S_2}^q, \ldots X_{S_k}^q ; X_{F_i}^q\right) \right) - \lambda_2 Tc\left(X_{S_1}^q, X_{S_2}^q, \ldots X_{S_k}^q\right) \tag{5}$$

where λ_1 and λ_2 represent the information-redundancy trade-off weights satisfying $\lambda_1 + \lambda_2 = 1$. In this study, λ_1 and λ_2 are set to 0.8 and 0.2 (Wang et al., 2018). A greedy monitoring network optimization is used to solve the Eq. (5) (Alfonso et al., 2010b; Fahle et al., 2015). While global optimization algorithms (e.g., Genetic Algorithms) can theoretically explore the entire solution space, the combinatorial explosion of candidate subsets makes them computationally prohibitive for high-dimensional groundwater models. Although the greedy algorithm, being a heuristic method, may lead to suboptimal solutions (local optima) as it does not backtrack to revise previous choices, the impact of this suboptimality is limited in the context of groundwater monitoring. Due to the strong spatial correlation of hydraulic and hydrochemical fields, multiple monitoring configurations can yield practically equivalent information content. Therefore, the marginal gain of finding the absolute global optimum is negligible compared to the significant computational savings offered by the greedy approach. Consequently, the process of selecting CMLs from $\boldsymbol{F}$ to $\boldsymbol{S}$ gradually reduces the gap between $H(\boldsymbol{S})$ and $H(\boldsymbol{C})$. In general, a predefined ratio (Pr) is set as a stopping criterion for the optimization process to balance survey costs and the amount of effective information:

$$H(\boldsymbol{S}) \geq Pr \times H(\boldsymbol{C}) \tag{6}$$

A higher value of Pr indicates that more CMLs are needed to meet the termination condition of the optimization process. Although the optimized monitoring network contains more effective information, the

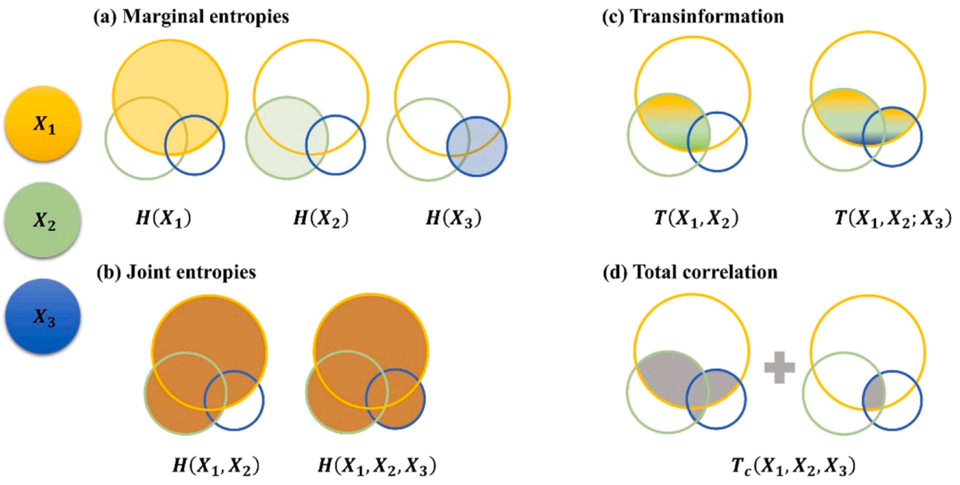


Fig. 3. Topological implications for different entropy-related concepts (Dai et al., 2022).

actual survey cost will also increase significantly. Therefore, the choice of *Pr* does not follow a fixed standard and should be determined based on the specific optimization context.

2.1.2. Data discretization

According to Eq. (1), entropy calculation focuses on the probability of a data occurrence rather than its actual value. Given that hydraulic head and solute concentration responses are continuous, data must be discretized for entropy computation.

This study employs the mathematical floor function for data discretization (Alfonso et al., 2010b; Ruddell and Kumar, 2009):

$$x' = \Delta x \left\lfloor \frac{2x + \Delta x}{2\Delta x} \right\rfloor \tag{7}$$

where $\lfloor \cdots \rfloor$ denotes the floor brackets, x represents the continuous dynamic response, Δx is the bin width (i.e., the discretization constant), and x' is the discretized data.

The bin width Δx is a key factor in entropy computation, and its selection in groundwater monitoring network design is often influenced by measurement noise (Keum and Coulibaly, 2017; Keum et al., 2018). This study compares three empirical formulas for determining Δx (Δx_{st} Sturges, 1926, Δx_{sc} Scott, 1979, Δx_{pb} Piersol and Bendat, 1966) along with three fixed bin width methods. The optimal Δx calculation process using empirical formulas is as follows:

$$\begin{cases} \Delta x_{st} = \dfrac{R_x}{1 + \log_2 N} \\ \Delta x_{sc} = 3.49\sigma(x) N^{-\frac{1}{3}} \\ \Delta x_{pb} = \dfrac{R_x}{1.87(N-1)^{0.4}} \end{cases} \tag{8}$$

where $R_x = x_{max} - x_{in}$ represents the range of continuous data values. $\sigma(x)$ represents the standard deviation of x, and N is the sampling size.

2.2.3. Monitoring network optimization

The uncertainty in dynamic response datasets results in non-uniqueness in monitoring network optimization results, which may miss critical response data and reduce the data worth of the optimized network. To overcome this challenge, we develop a robust optimization framework based on TP stochastic simulation. Its advantage lies in enhancing the diversity of the dynamic response dataset through TP model -driven stochastic geological modeling (Fig. 4a). The main process is as follows: first, structure parameter ranges are defined based on borehole data, and Latin hypercube sampling is used to generate stochastic heterogeneous structures. Dynamic response predictions are performed on these stochastic structures to obtain data at the CMLs. After discretizing the data, optimization is performed based on the MIMR criterion to obtain the current optimal monitoring network. A frequency distribution matrix for monitoring point selection is constructed through N_c Monte Carlo iterations (N_c is set to 100 in this study), where frequency values reflect the information entropy weights of the monitoring points (Fig. 4b). In this paper, the stochastic simulation tool GEOST is used to complete the solution of the TP model (Dai et al., 2014). The dynamic responses are obtained by running the TOUGHREACT simulator (Xu et al., 2006).

Notably, a frequency-based selection strategy alone is impractical in real applications due to the significant cost differences between horizontal placement and vertical drilling. To address this, we propose a monitoring well prioritization strategy: in the CMLs system, monitoring wells with higher information entropy weights are selected first, followed by vertical monitoring point selection within the chosen monitoring well. The feasibility of this strategy has been demonstrated through the Drive-in Multi-Depth Sampling System developed by the University of Hong Kong (Luo et al., 2017), which successfully achieved multi-depth synchronous monitoring at the land-sea interface, confirming the practical applicability of the monitoring well prioritization strategy.

During the iterative process of stochastic optimization, repeated calls to high-fidelity numerical models significantly increase computational costs. To address this issue, two deep learning surrogate models with different output modes are developed to rapidly obtain response data on

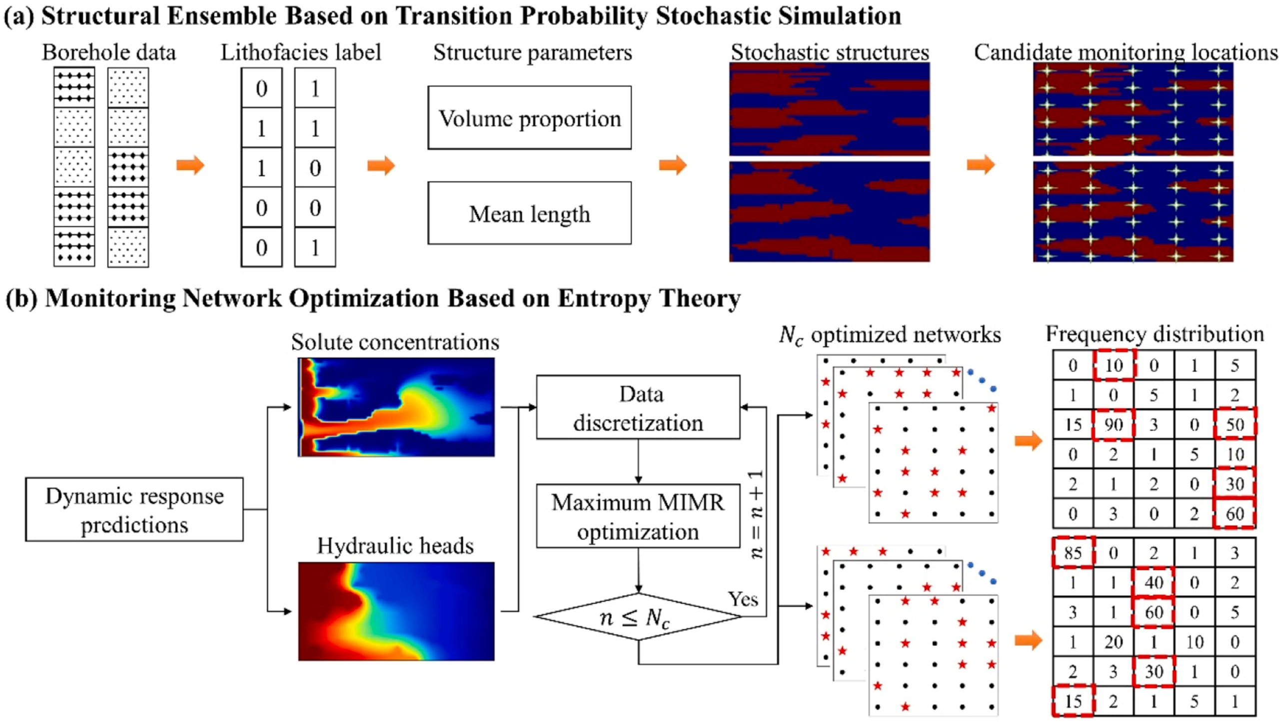


Fig. 4. Transition probability and entropy-based monitoring networks optimization.

CMLs: the Deep Octave Convolution Residual Network for Point-Information (DOCRP) and the Deep Octave Convolution Residual Network for Image-Information (DOCRI). The detailed architectures are shown in Fig. 5. The input structure is treated as image information, where each grid cell is regarded as a pixel, and the data within the grid represents the pixel value. Feature extraction is achieved by repeatedly passing the input through residual in residual dense residual block (RRDRB, Rakotonirina and Rasoanaivo, 2020; Zhan et al., 2022a). The primary difference between DOCRP and DOCRI lies in how the feature matrix output by the final RRDRB is processed: DOCRP flattens and vectorizes the feature matrix before passing it through a fully connected layer, ultimately outputting time-series responses on CMLs (1D output). DOCRI upsamples the feature matrix and then processes it through convolutional layers to reconstruct a dynamic response field with dimensions consistent with the input structure (2D output).

2.2. *General framework for DL-enhanced coupled hydrogeophysical inversion*

A general framework for coupled hydrogeophysical inversion typically consists of: (1) a structure generation model to reduce computational dimensionality; (2) a forward model to generate dynamic responses; (3) a petrophysical relationship that links the hydrological and geophysical models; and (4) an inversion algorithm for parameter updates. This framework iteratively updates the low-dimensional parameters to ensure that the simulated dynamic responses closely match the observations. The updated parameters are then input into the structure generation model to obtain the posterior structure ensemble. Incorporating a DL generation model and a DL surrogate model can enhance this framework (See Fig. 6). In terms of the structure generation model, the primary distinction between non-DL and DL methods lies in the dimensionality of the low-dimensional parameters. For example, the TP generation model used in this study parameterizes high-dimensional heterogeneous structures using structural parameters ($S_i = \{p, L_x, L_z\}$, where p represents volume proportions and L denotes mean lengths in different directions). The DL generation model employs OCAAE (Zhan et al., 2022a) to parameterize high-dimensional heterogeneous structures using a latent vector of customizable dimensions ($m = 200$ in this study). For generating dynamic responses, non-DL methods rely on high-fidelity numerical models and consider only observation noise during the update process. The training set for the DL surrogate model is derived from numerical simulations. While providing efficient dynamic response predictions, it introduces model noise due to deviations from numerical model outputs. Therefore, the update process considers both observation noise and model noise. The subsurface structures discussed in this study are clay-free, so surface resistivity can be neglected. The petrophysical relationship describes the nonlinear relationship between solute concentration and bulk resistivity. Specifically, Archie's law is applied to describe the relationship between pore fluid resistivity and bulk resistivity (Archie, 1942). Pore fluid resistivity is determined by both temperature and solute concentration (Sen and Goode, 1992). Time-lapse ERT data are obtained by using SimPEG (Cockett et al., 2015; Heagy et al., 2017), and the surrogate model employs HROCN (Xia et al., 2026). The architecture and code of the HROCN model are available in the Open Research section. This study employs the Iterative Local Updating Ensemble Smoother (ILUES) to assimilate different types of dynamic observations (Zhang et al., 2018). ILUES effectively identifies multimodal distributed parameters in complex hydrological models, and its local updating strategy significantly reduces computational complexity in high-dimensional parameter spaces, improving efficiency. It is important to note that the primary focus of this study is not to evaluate the superiority of the solver itself. For a systematic comparative study of data assimilation algorithms, readers are referred to the comprehensive evaluation work by Chang et al. (2024).

2.3. *Application example*

The 2D synthetic heterogeneous subsurface structure, as shown in Fig. 7a, spans 80 m in length and 40 m in depth, discretized into 3200 uniform square grids with a 1 m side length. This model is generated using GEOST based on predefined structure parameters, with the estimation process for the prior interval as follows: first, five boreholes

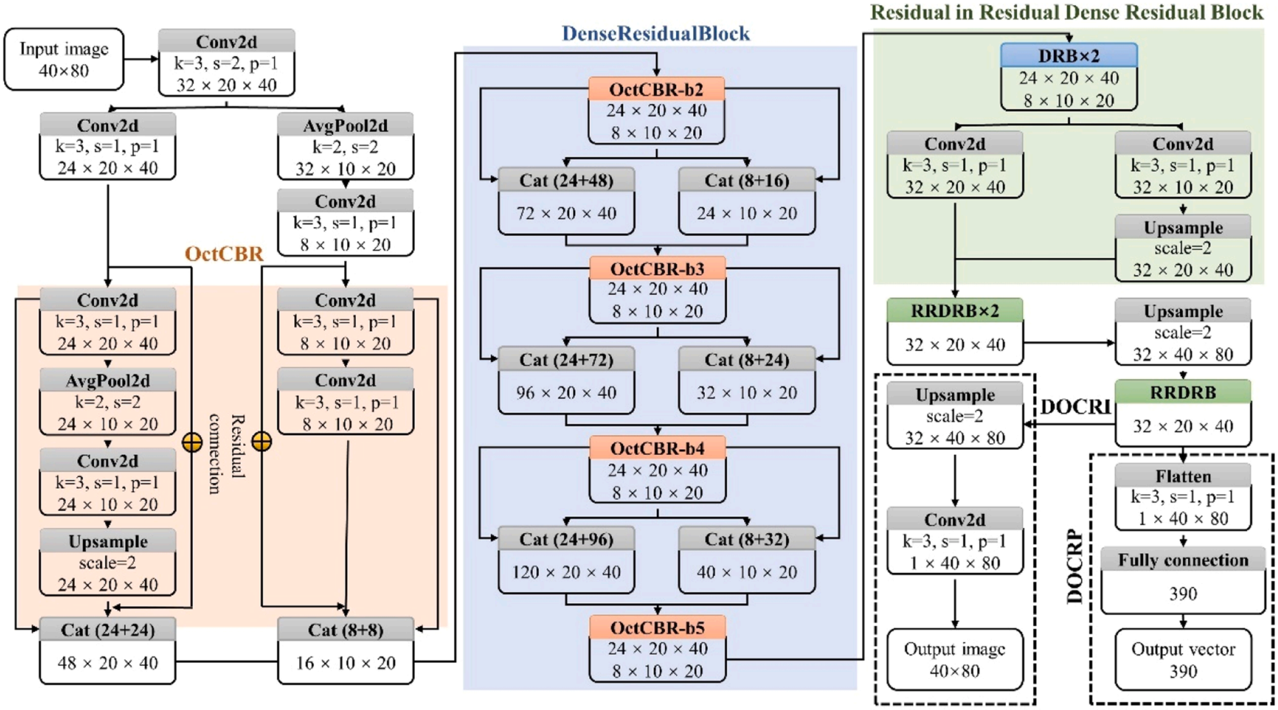


Fig. 5. Model architecture of DOCRP and DOCRI. k represents kernel size, s represents the stride, and p represents the zero-paddings.

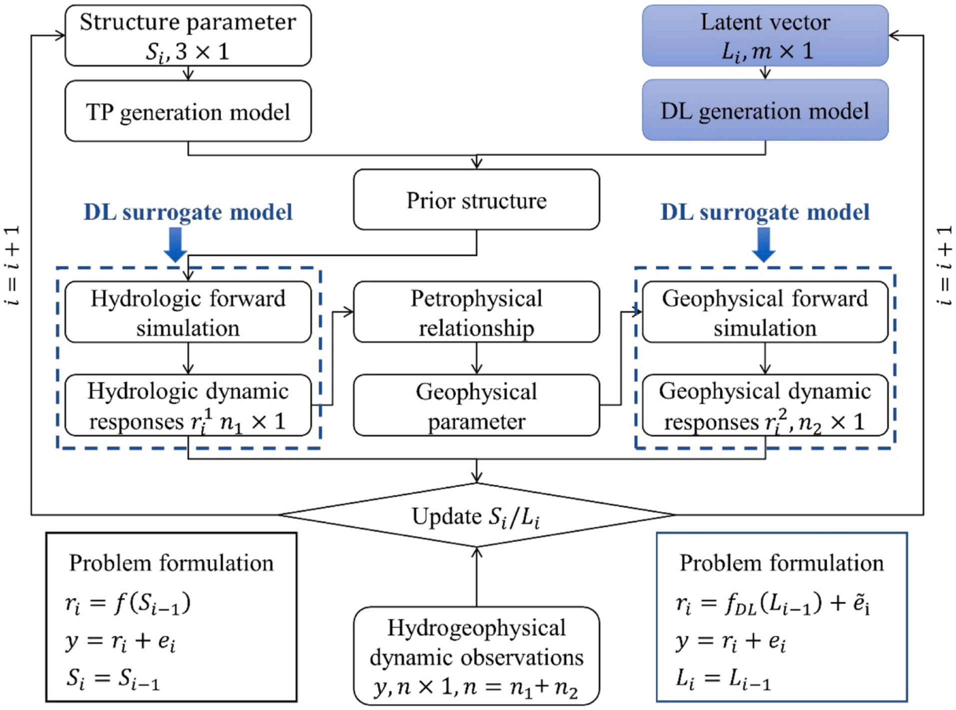


Fig. 6. General framework for DL-enhanced coupled hydrogeophysical inversion.

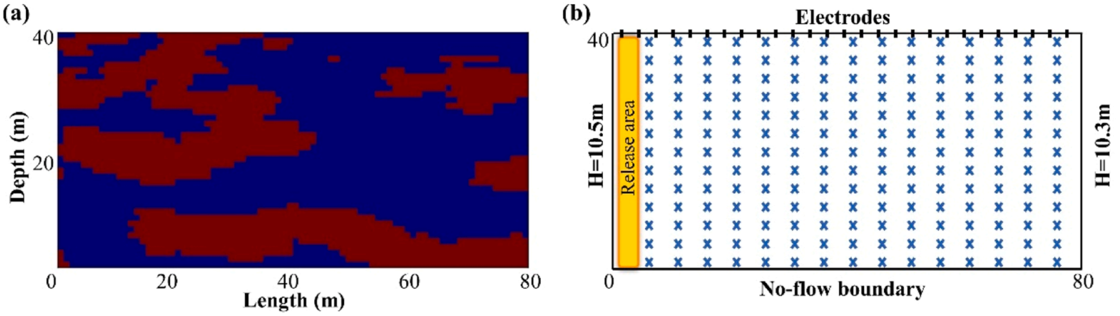


Fig. 7. (a) Distribution of low-permeability (red) and high-permeability (blue) facies. (b) Flow domain of the solute transport model. The yellow area is the release area, the blue ' × ' are the CMLs, and the black rectangles are the electrodes.

penetrating the entire structure are placed within the study area at locations $x = 0$, 20, 40, 60, and 80 m. Based on lithofacies data observed from the boreholes, the volume proportion is estimated as p=0.38, and the mean length in the z-direction is L_z=5.06. The corresponding prior distribution is then set based on this estimate (see Table 1). However, due to the limitations of discrete field investigations in providing reliable horizontal continuity information, the prior distribution of the mean length in the x-direction (L_x) is broader. The structure consists of two facies types: high-permeability (k = 16.799 × 10^{-10} m^2) and low-permeability (k = 0.094 × 10^{-10} m^2). Therefore, the structure parameters of only one facies type are needed to characterize the entire structure. An ERT-monitored tracer test is conducted within the computational domain. As shown in Fig. 7b, $x = 0$ m and $x = 80$ m are constant hydraulic head boundaries, with head values of 10.5 m and 10.3 m, respectively. $z = 0$ m and $z = 40$ m are no-flow boundaries. In the release area near $x = 0$ m, a conservative tracer with a concentration of 0.05 mol/L is continuously injected at a rate of 0.02 kg/s for one month. Observations are collected at $T = 2$ d, 5 d, 8 d, 14 d, 22 d, and 30 d The subsurface temperature is maintained at a constant 25 °C.

In this case, 195 CMLs are uniformly distributed (15 × 13 grid) with x-direction spacing of 5 m and z-direction spacing of 3 m. The optimal monitoring network consists of the most valuable CMLs in terms of data quality. Subsequently, dynamic observations of hydraulic head and solute concentration will be obtained from the optimal monitoring network. Surface electrodes are arranged in a dipole-dipole array with 2 m spacing to collect time-lapse ERT data. Based on different data types and monitoring network configurations, eight identification scenarios are designed (see Table 2). The single monitoring network is optimized using only one type of response data. OS1 and OS4 represent inversion scenarios for hydraulic head and solute concentration based on their respective single monitoring networks. The fusion monitoring network is optimized by considering two types of response data. OS3, OS6, and OS8 represent inversion scenarios using hydraulic head, solute concentration, and fusion data within the fusion monitoring network. The

Table 1
The structure parameters and prior distributions for low-k Facies type.

	P	L_x	L_z
Predefined	0.4	17	5
Prior	[0.3, 0.5]	[10, 50]	[3, 6.5]

Table 2

Definition of identification scenarios combining different dynamic observations and optimized monitoring networks.

	Identification Scenarios	OS1	OS2	OS3	OS4	OS5	OS6	OS7	OS8
Dynamic Observations	Hydraulic head	√	√	√				√	√
	Solute concentration				√	√	√	√	√
Optimized Monitoring network	Single	√			√				
	Stochastic		√			√		√	
	Fusion			√			√		√

stochastic monitoring network is randomly selected from the remaining CMLs, with the same number of monitoring locations as the fusion monitoring network. OS2, OS5, and OS7 are designed to analyze the improvement in inversion performance when using an optimized monitoring network with the same amount of data.

High-worth hydraulic head (H) and solute concentration (C) observations obtained from the optimal monitoring network are integrated with ERT data for fusion analysis. This study evaluates 21 scenarios, categorized into single-data inversion scenarios and fusion-data inversion scenarios. Physically, resistivity variations are driven by changes in C, meaning that both data types originate from the same underlying process but differ in resolution and spatial coverage. Therefore, it is worth investigating whether fusing ERT and C data at the same observation times might lead to redundancy, potentially affecting inversion performance. Based on this, we further expand the C and ERT data into C_1, C_2, and ERT_1, ERT_2, corresponding to observation times of T_1=2 d, 8 d, 22 d and T_2=5 d, 14 d, 30 d, respectively. Table 3 provides details on the seven observation data combination scenarios. To investigate the impact of integrating different types of observations, this stage considers only 1 % Gaussian random observation noise. Finally, to evaluate the impact of observation noise and model noise on data worth, H, C, E, HE, CE, and HCE are analyzed under observation noise levels of 1 %, 5 %, 10 %, and 15 %, as well as model noise levels of 0 %, 5 %, 10 %, and 15 %. This specific range was selected to define the performance boundaries under different scenarios. The 0 % model noise and 1 % observation noise serve as distinct baselines: the former corresponds to an ideal case where the DL surrogate model has no prediction errors, while the latter represents the practical limit of high-precision instruments. In contrast, the 15 % levels simulate difficult conditions with significant model uncertainty and complex measurement errors. Crucially, including intermediate levels (5 % and 10 %) allows us to verify the method's robustness. This ensures that the solution quality declines steadily and predictably as noise increases, rather than failing severely. As a result, a total of 96 scenarios are designed to account for different data fusion types, observation noise, and model noise.

3. Results

In this section, we first evaluate the results of monitoring network optimization and then conduct hydrogeophysical coupled inversion based on the optimal monitoring network. In all scenarios, we assess the ability to identify subsurface structures and the uncertainty associated with such characterization. The TP generation model is used as the structure parameterization method in monitoring network optimization and hydrogeophysical coupled inversion, with evaluation metrics including structure parameters and structure identification accuracy. The structure parameterization method based on the DL model is applied to hydrogeophysical coupled inversion under the same scenarios to compare the effectiveness of different parameterization methods. Based on this, we further explore the impact of observation noise and model noise on structure identification accuracy and errors in dynamic responses prediction.

3.1. Monitoring network optimization results

3.1.1. Spatiotemporal variability of dynamic responses

In the optimization of monitoring networks for dynamic observations, the selection of observation times is often neglected, and optimization is usually conducted using data from a stochastically chosen time point. Additionally, for a standard S-shaped concentration breakthrough curve, concentration represents a single point on the curve, whereas the sum of solute concentrations represents the curve's area. Previous studies suggest that the sum of solute concentrations over different time points better captures the continuous variation characteristics of the breakthrough curve (Dai et al., 2022). The information content of observations is related to uncertainty. Therefore, before optimizing the monitoring network, we conduct a spatiotemporal variability analysis of dynamic responses to guide the selection of observation times.

In practical studies, the true subsurface structure is typically unknown, while borehole data obtained from site investigations is known. Therefore, we use the facies observations from boreholes (as described in Section 2.3) as conditional data. TP generation model is employed to generate 10,000 possible stochastic structure samples, which are subsequently input into high-fidelity numerical model to obtain datasets of dynamic responses. Temporal dynamics are evaluated using the standard deviation of data at each grid point over time:

$$\sigma_i^r = \sqrt{\frac{\sum_{T=1}^{S}\left(\sigma_{i,T}^r - \overline{\sigma_i^r}\right)^2}{S}} \tag{9}$$

where σ_i^r is the standard deviation of the rth type of dynamic responses at grid point i over time, T is the current observation time, S is the total number of observation time points, and $\overline{\sigma_i^r}$ is the mean of $\sigma_{i,T}^r$. The spatial dynamics of dynamic responses are expressed as follows:

Table 3

Definition of fusion-data inversion scenarios for observation types.

Scenarios	HE_1	HE_2	HE	C_1E_1	C_1E_2	C_1E	C_2E_1	C_2E_2	C_2E	CE_1	CE_2	CE	HC	HCE
H	√	√	√										√	√
C_1				√	√	√								
C_2							√	√	√					
C										√	√	√	√	√
ERT_1	√			√			√			√				
ERT_2		√			√			√			√			
ERT			√			√			√			√		√

Abbreviations: H and C represent **hydraulic head** and **solute concentration**, respectively. C_1 and ERT_1 correspond to solute concentration and ERT data collected at **T1=2 d, 8 d, and 22 d**, while C_2 and ERT_2 refer to solute concentration and ERT data collected at **T2=5 d, 14 d, and 30 d**.

$$\sigma_{i,T}^{r}=\sqrt{\frac{\sum_{j=1}^{N}\left(R_{i,\,j}^{r}-\overline{R_{i}^{r}}\right)^{2}}{N}} \tag{10}$$

where $\sigma_{i,T}^{r}$ represents the standard deviation of the rth type of dynamic response at grid point i at time T, $R_{i,\,j}^{r}$ is the predicted value of the rth dynamic responses for the jth state field sample at grid point i, $\overline{R_{i}^{r}}$ is the mean of $R_{i,\,j}^{r}$, and N is the number of samples.

The results of temporal dynamics are shown in Fig. 8. The distribution of solute concentration data exhibit significant temporal variability, whereas the distribution of hydraulic head remains nearly unchanged over time, indicating that hydraulic head data provide nearly identical information at any given time. It is important to note that this conclusion is related to the boundary conditions considered in the forward model. The steady-state flow conditions set in this study result in weak temporal variability in hydraulic head data. Fig. 9 compares the spatial dynamics of solute concentration at different observation times, the sum of observed solute concentration, and hydraulic head at $T = 30$ d Evidently, solute concentration at $T = 30$ d demonstrates the most pronounced spatial variability, suggesting that structure heterogeneity exerts the strongest influence on the solute concentration distribution and that its variability spans a wider spatial extent.

3.1.2. Assessment of DL surrogate models

To assess the performance of DOCRP and DOCRI, datasets are generated based on 10,000 randomly sampled structures and dynamic responses obtained from the high-fidelity numerical model. Detailed information on the predicted data for the different surrogate models is presented in Table 4. Specifically, DOCRP-ST and DOCRI-ST are used to predict the sum of solute concentrations over multiple time points and the hydraulic head at a single time point, while DOCRP-TT and DOCRI-AA are used to predict the solute concentration and hydraulic head at a single time point. The actual output of DOCRI-AA includes dynamic responses across all time points, but only the data corresponding to the target time point is extracted for comparative analysis. The surrogate models are trained with a batch size of 64, a learning rate of 0.001, and a weight decay of 0.00005. After training, 1000 samples are randomly selected from the test set for performance assessment using the coefficient of determination (R^2) and the Root-Mean-Square Error (*RMSE*):

$$R^{2}=\frac{\sum_{i=1}^{n}(\widetilde{y}_{i}-\overline{y})}{\sum_{i=1}^{n}(y_{i}-\overline{y})} \tag{11}$$

and

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widetilde{y}_{i}-y_{i})^{2}} \tag{12}$$

where $\widetilde{y}_i$ represents the output of surrogate model, y_i is the output of the high-precision forward model, $\overline{y}$ is the mean value of y_i.

The predicted dynamic response data at CMLs is shown in Fig. 10, where the horizontal axis represents the output data from the surrogate model, and the vertical axis represents the output data from the high-fidelity numerical model. For a specific type of dynamic response data, DOCRI demonstrates higher prediction accuracy. For example, comparing Fig. 10a and b, when predicting the summed solute concentration data (c_{sum}), DOCRI-ST achieves a higher R^2 value and a lower *RMSE* value compared to DOCRP-ST. The same trend is observed in the prediction of hydraulic head at a single observation time (h_T). Comparing Fig. 10c and d, although the output of DOCRI-AA ($c_{all} + h_{all}$) includes more observation times and involves a more complex nonlinear relationship than the output of DOCRP-TT ($c_T + h_T$), DOCRI still demonstrates superior prediction accuracy. These results indicate that for DL surrogate models with image-information input mode for structures, maintaining an image-information output mode for the dynamic response field is more advantageous for capturing local heterogeneity. Fig. 11 compares the marginal entropy of dynamic responses from the numerical model and DOCRI, where dynamic response data are discretized using the optimal bin width from Eq. (8) and fixed bin widths of Δx=0.05, 0.1, and 0.15. The marginal entropy values (vertical axis) from 195 CMLs based on DOCRI-ST and DOCRI-AA closely match those obtained from the numerical model (horizontal axis). This indicates that DOCRI can provide reliable and efficient entropy estimates for monitoring network optimization.

3.1.3. Assessment of optimized monitoring networks

In this study, the data size is determined using the mean (M_{mean}) and standard deviation (M_{std}) of marginal entropies. A sufficiently large data size should result in a stable trend for marginal entropies. Fig. 12 presents the changes in M_{mean} and M_{std} with increasing data size under different Δx conditions. Except for Δx_{sc}, when the data size reaches 2000, oscillation in M_{mean} and M_{std} significantly decrease. Notably, under fixed Δx conditions, both M_{mean} and M_{std} converge to a constant value. Among the three optimal bin width methods, only under the Δx_{st} condition do M_{mean} and M_{std} exhibit initial oscillations followed by a relatively stable trend as data size increases. In contrast, under Δx_{sc} and Δx_{pb} conditions, M_{mean} and M_{std} continue to increase. This phenomenon occurs because the bin width in optimal bin width methods dynamically changes with data size, but the degree of this influence varies. Therefore, in this study, the fixed bin width method proves to be more suitable for the intended standards.

Under the three fixed bin width conditions, the convergence values of marginal entropies differ. For example, in the calculation of summed solute concentrations (see Fig. 12a), as Δx increases, the convergence value of M_{mean} rises from 1.75 to 3.0, while M_{std} increases from 0.48 to 0.67. This indicates that while the overall data information captured by CMLs increases, the variability among different locations also grows. A similar trend can be observed in the solute concentration data at $T = 30$ d (Fig. 12b) and the hydraulic head data (Fig. 12c). Additionally, in terms of M_{mean} convergence, the marginal entropies of the two solute concentration datasets are significantly higher than those of the hydraulic head data. The marginal entropies of solute concentration at $T =$

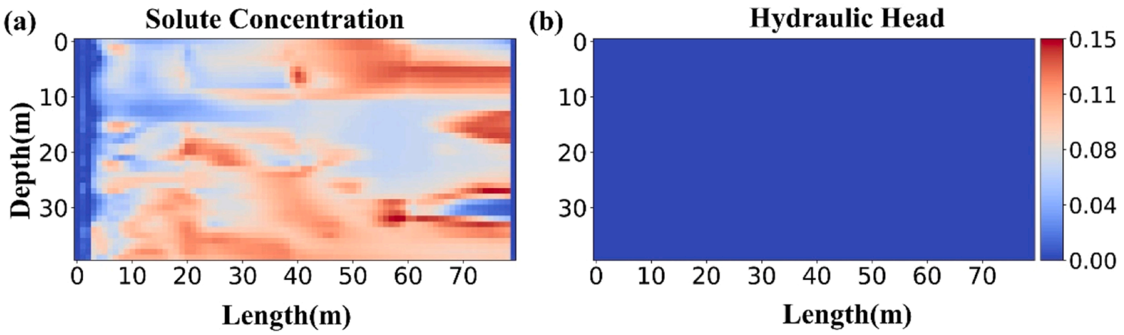


Fig. 8. Temporal dynamics of solute concentration and hydraulic head.

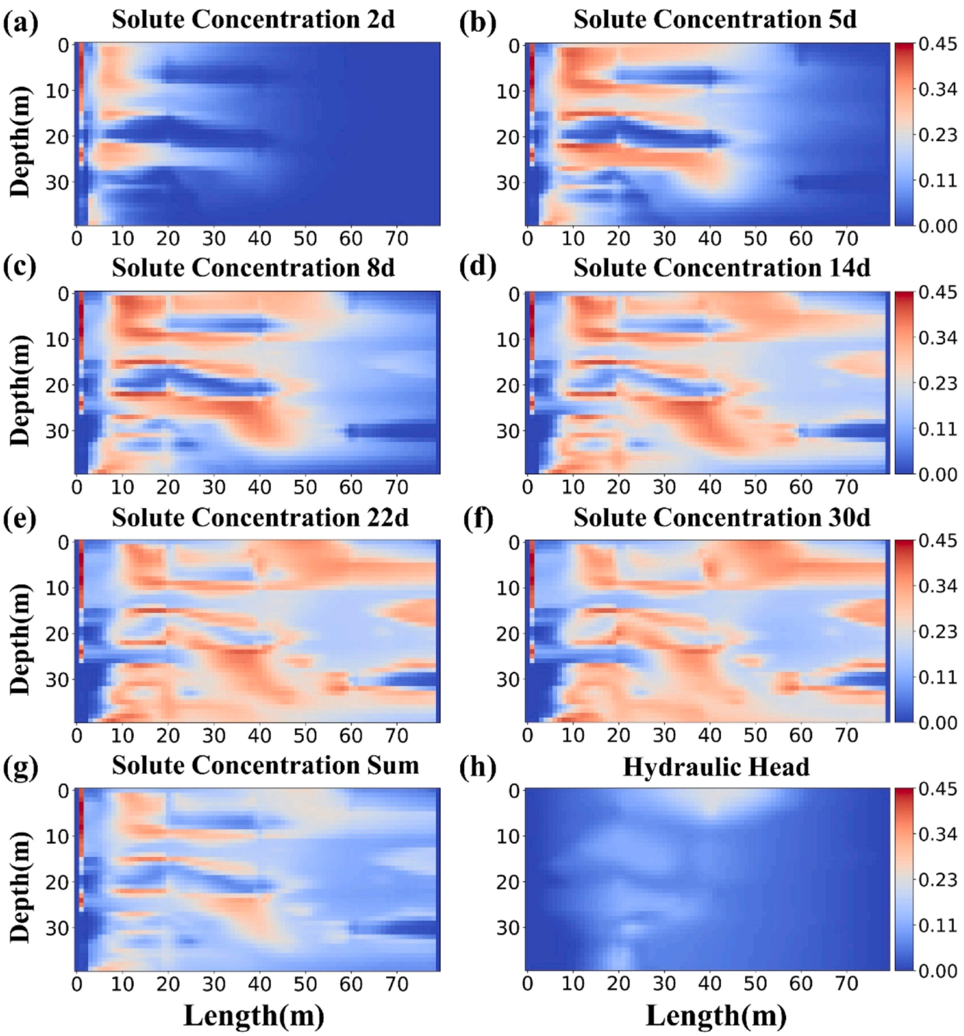


Fig. 9. Spatial dynamics of solute concentration and hydraulic head.

Table 4
Definition of various surrogate models and their corresponding predicted data.

Name	Output mode	Output data	Number of output data
DOCRP-ST	Point-information	$c_{sum} + h_T$	390
DOCRI-ST	Image-information	$c_{sum} + h_T$	6400
DOCRP-TT	Point-information	$c_T + h_T$	390
DOCRI-AA	Image-information	$c_{all} + h_{all}$	38,400

Abbreviations: c_{sum}, Sum of solute concentrations over all observation times; c_T, solute concentration at the observation time $T = 30$ d; c_{all}, solute concentration at observation times; h_T, hydraulic head at the observation time $T = 30$ d; h_{all}, hydraulic head at all observation times.

30 d are greater than those of summed solute concentration, indicating that the former contains more data worth. Ultimately, this study selects Δx=0.05 and a data size of 2000, and monitoring network optimization is conducted based on the solute concentration and hydraulic head data at $T = 30$ d

The *Pr* in Eq. (6) affects the number of selected monitoring locations. A larger *Pr* makes the optimized monitoring network contain information closer to CMLs, but it also significantly increases the number of monitoring points, which is unfavorable for controlling survey costs. Fig. 13 shows the variation in the number of selected monitoring locations with respect to *Pr*. When $Pr > 0.92$, the number of monitoring points for solute concentration stabilizes, indicating that the high-worth monitoring points have been selected. When *Pr* reaches 0.95 and 0.97, the number of monitoring points for hydraulic head increases rapidly. This suggests that, to meet the predefined information threshold, additional monitoring points with relatively lower data worth are included, and there is a significant amount of redundant information between newly added and previously selected points. At this stage, adding more monitoring points is no longer economically justified. The subsequent analysis will examine in detail the scenarios where *Pr* is set to 0.95 and 0.97.

TP and entropy-based monitoring network optimizations are conducted under different *Pr* conditions. Details are provided in Text S1 in Supporting Information. The selection frequency of each CMLs can be obtained from 100 stochastic optimizations. A higher selection frequency for a specific monitoring location indicates a higher data worth of the response data. Fig. 14 visually presents the selection frequency of all CMLs for solute concentration and hydraulic head data across 100 optimization runs under different *Pr* conditions. It is evident that as *Pr* increases, the number of newly selected monitoring locations for hydraulic head data is significantly higher than for solute concentration data. This is related to the fact that high-worth monitoring locations for solute concentration data contain more information. The impact of structure parameter uncertainty on optimization results is mainly reflected in cases where some monitoring locations are never selected at *Pr*=0.95 but are selected a few times (less than five) at *Pr*=0.97. Additionally, certain monitoring locations that are selected once at *Pr*=0.95 are not selected at all at *Pr*=0.97. However, the three monitoring wells

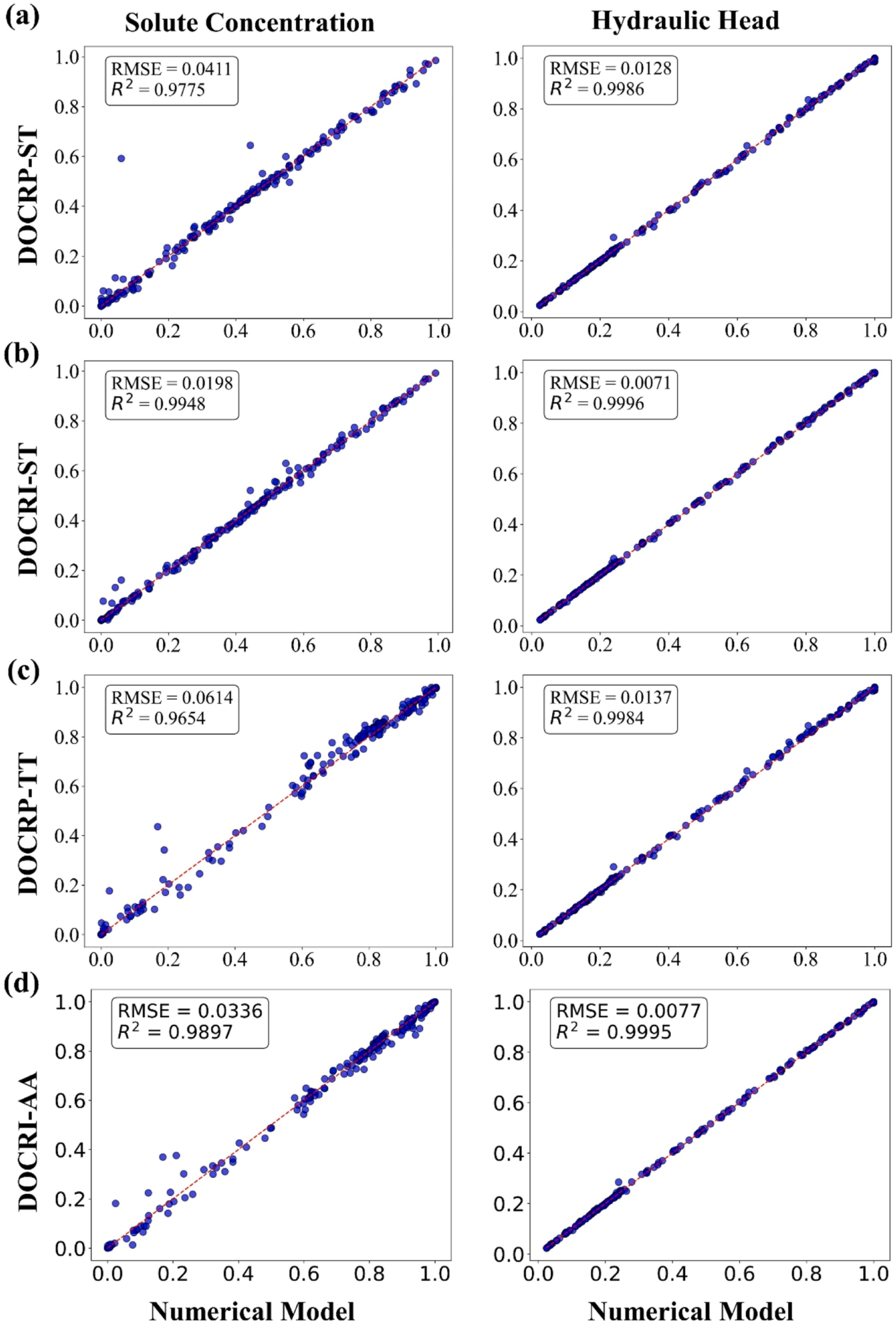


Fig. 10. Comparison of dynamic responses on CMLs between DL surrogate models and high-fidelity numerical models.

selected based on the monitoring well prioritization strategy (indicated by yellow triangles in Fig. 14) remained consistent across different *Pr* conditions, indicating that the proposed method helps mitigate the impact of structure parameter uncertainty.

The optimized monitoring networks for solute concentration (Fig. 15a) and hydraulic head (Fig. 15b) data are formed by selecting the most frequently chosen monitoring locations from the identified monitoring wells. Since one monitoring well overlapped, the fusion monitoring network ultimately included five wells (Fig. 15c). The stochastic monitoring network (Fig. 15d) is designed to compare its information content with that of the fusion monitoring network while maintaining the same economic cost. The identification of heterogeneous structures is conducted based on the identification scenarios described in Table 2. First, 2000 sets of prior parameters are stochastically sampled within the structure parameter ranges specified in Table 1 to ensure the same prior parameters are used across all identification scenarios. For each scenario, the observations are generated using a high-fidelity numerical model, including hydraulic head and solute concentration data at six observation times. ILUES is used to assimilate the corresponding observations and obtain 2000 sets of posterior structure parameters. During the inversion process, DOCRI-AA is employed to accelerate forward simulations, with the iteration process terminating after 40 iterations.

First, a comparative analysis is conducted on the posterior ensembles

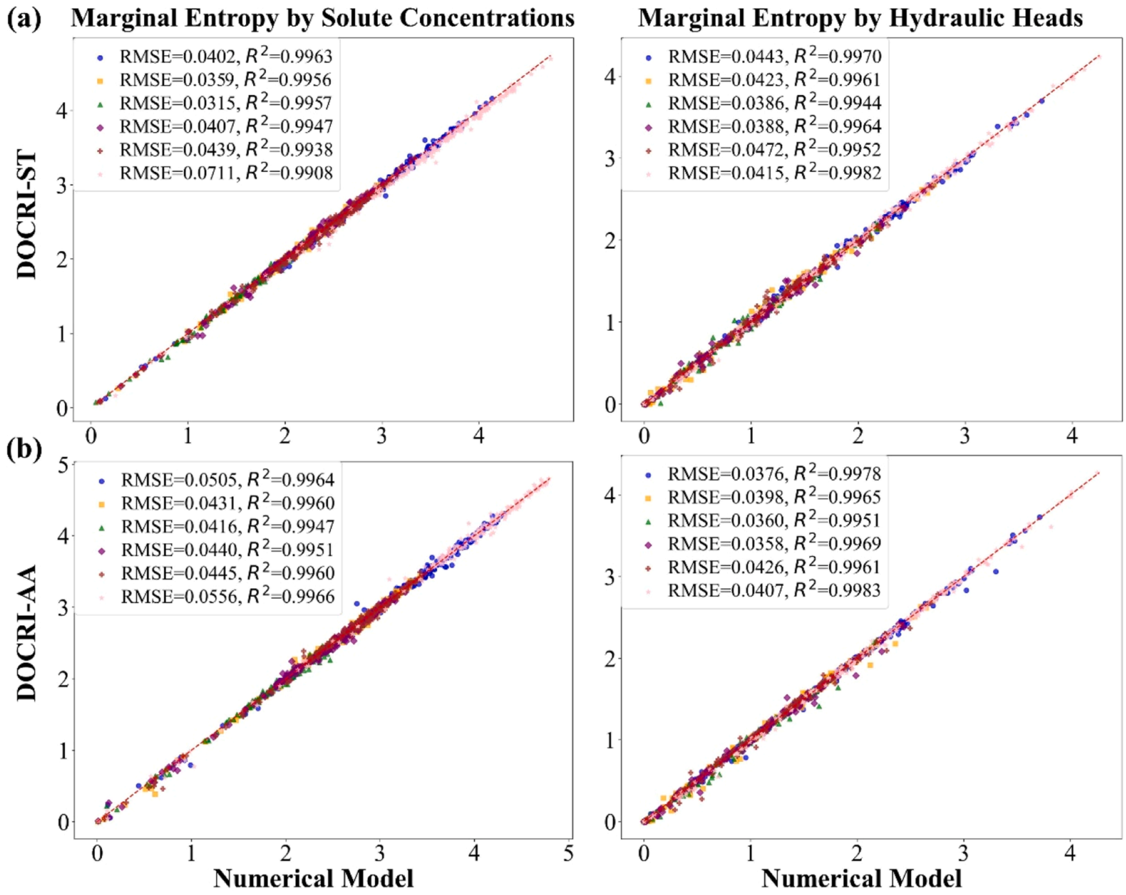


Fig. 11. Performance of DOCRI-ST and DOCRI-AA for entropy analysis. The blue circle, yellow rectangle, and green triangle represent fixed bin width Δx=0.05, 0.1, 0.15; the purple diamond, brown plus sign, and pink star represent the optimal bin width Δx_{sc}, Δx_{st}, Δx_{pb}, respectively.

of structure parameters obtained from inversion under different scenarios (see Fig. 16a-c). The posterior ensemble of structure parameters obtained from at least one type of observations inversion is significantly closer to the true values compared to the prior ensemble, with a notable reduction in uncertainty. The effectiveness of different types of observations in identifying structure parameters varies depending on the monitoring network. Hydraulic head data (OS1–3) yield relatively lower identification accuracy but has the least uncertainty. The identification results based on solute concentration data (OS4–6) are closest to the true values but have relatively higher uncertainty. For a given type of observations, the inversion results from the fusion monitoring network outperform those from the single monitoring network, while the stochastic monitoring network yields the worst performance. In the single inversion of hydraulic head data, the stochastic monitoring network achieves higher accuracy in identifying p. However, for L_x and L_z, the fusion monitoring network performs better. In scenarios combining solute concentration and hydraulic head data (OS7–8), all monitoring networks achieve precise and reliable parameter identification. The fusion monitoring network, in particular, significantly outperforms the stochastic monitoring network in identifying L_z and p.

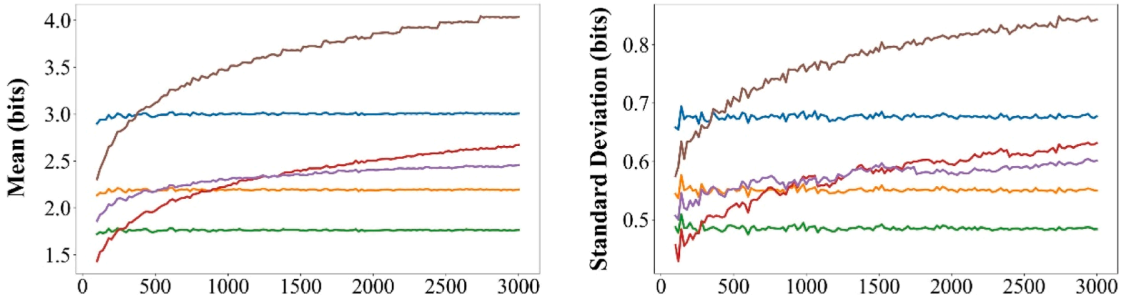
The accuracy of identifying heterogeneous structures is shown in Fig. 16d. The posterior structure ensemble obtained from at least one type of observations inversion exhibits significantly higher accuracy and lower uncertainty compared to the prior structure ensemble. The identification results vary significantly depending on the type of data used. Solute concentration data provide higher structure identification accuracy than hydraulic head data, but the latter exhibits the lowest uncertainty. The accurate reconstruction of structure requires considering all three structure parameters together, as the identification accuracy of any single parameter does not determine the final structure identification accuracy. For example, in identifying p, the stochastic monitoring network performs better than the other two networks when using hydraulic head data alone. However, it performs poorly in identifying L_x and L_z. As a result, the fusion monitoring network achieves a significantly higher structure identification accuracy than the stochastic monitoring network. For any given type of observations, including the data fusion scenario, the fusion monitoring network consistently achieves higher identification accuracy. This indicates that the fusion monitoring network effectively integrates the dynamic variability of different data types in relation to structure heterogeneity, ensuring that the selected monitoring locations contain the most valuable information.

3.2. Coupled hydrogeophysical inversion results

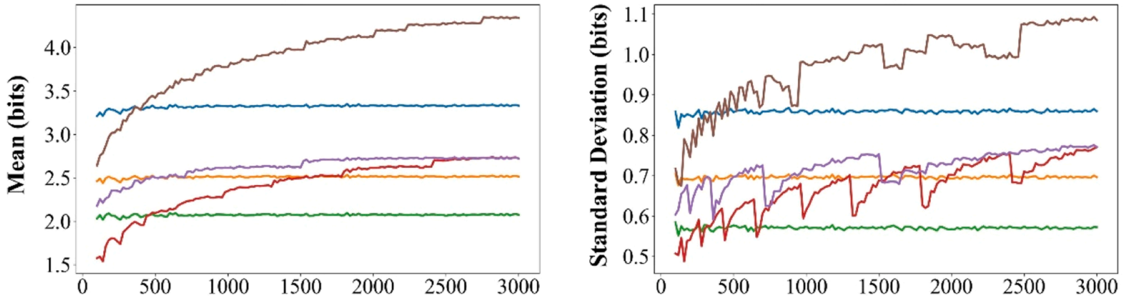
3.2.1. Structure parameters

2000 sets of prior structural parameters are stochastically selected from the predefined range in Table 1 for single-data inversion and fusion-data inversion in the coupled hydrogeophysical inversion (see Table 3). DOCRI-AA is used to accelerate the forward simulation of hydraulic head and solute concentration dynamic responses, while HROCN is employed to speed up the forward simulation of time-lapse ERT data. The observations required for different scenarios are obtained from the fusion monitoring network. To investigate whether integrating different types of data always benefits structure parameter identification, a parameterization method based on the TP generation model is used, with 1 % observation noise and 0 % model noise. The identification results of structure parameters under different scenarios are shown in Fig. 17. The assimilation of various observations types significantly improved the accuracy of posterior estimates while

(a) Sum of Solute Concentrations Marginal Entropy



(b) T = 30d Solute Concentrations Marginal Entropy



(c) T = 30d Hydraulic Heads Marginal Entropy

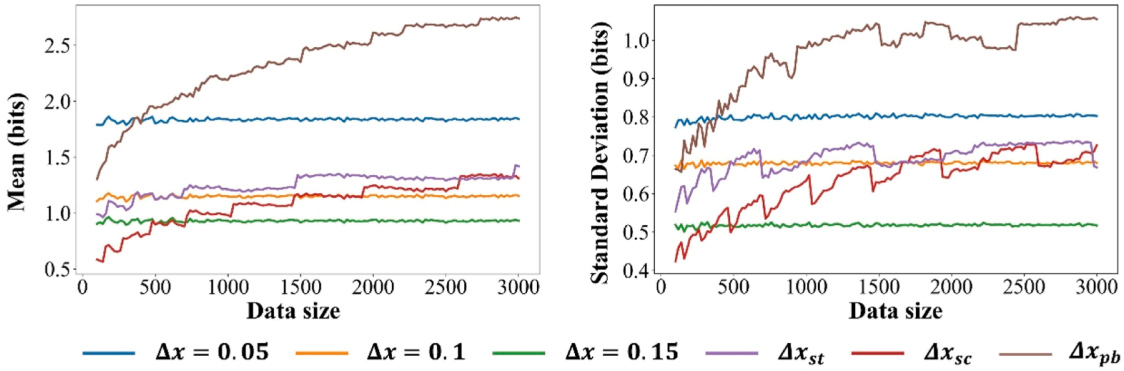


Fig. 12. Mean and standard deviation marginal entropy values of dynamic responses on CMLs under different bin widths with increasing data size.

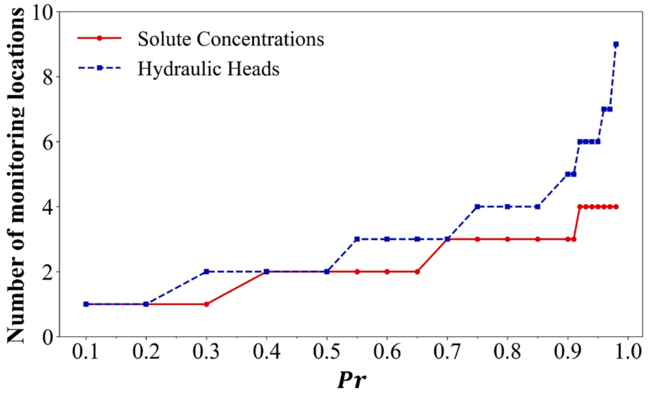


Fig. 13. Variation of the number of monitoring locations under different predefined ratio (***Pr***).

reducing uncertainty. Different types of observations have varying impacts on structure parameter identification. For single-data inversion, solute concentration data (scenarios C_1, C_2, and C) provide significantly higher estimation accuracy for structure parameters compared to other types of data. Hydraulic head data (scenario H) has the lowest estimation accuracy, while ERT data (scenarios E_1, E_2 and E) exhibit the lowest uncertainty. This suggests that solute concentration data from an optimized monitoring network hold higher data worth than hydraulic head data, while the lower uncertainty in ERT estimates is attributed to its extensive coverage and large data volume. Fusion-data inversion outperformed single-data inversion. As shown in Fig. 17a and b, the fused data of hydraulic head and ERT (scenarios HE_1, HE_2 and HE) provide more accurate estimations with lower uncertainty compared to hydraulic head. The estimation results of the fused data of hydraulic head and solute concentration (scenarios C_1E_1, C_1E_2, C_1E, C_2E_1, C_2E_2, C_2E, CE_1, CE_2 and CE) show that fusing solute concentration and ERT data does not further improve accuracy but significantly reduces uncertainty. However, fusion-data inversion does not always lead to improvement. For example, in Fig. 17c, for the identification of L_z, a comparison between hydraulic head and the fused data of hydraulic head and ERT (scenarios H and HE) shows little difference in accuracy and uncertainty. This is because ERT data have limited inversion coverage in the depth direction, making it difficult to provide reliable depth information. Similarly, ERT data exhibit higher uncertainty in estimating L_z compared to p and L_x. Additionally, compared to the fused data of hydraulic head and solute concentration (scenario HC), further integrating ERT data in scenario HCE positively impacts both the accuracy and uncertainty of structure parameter estimation.

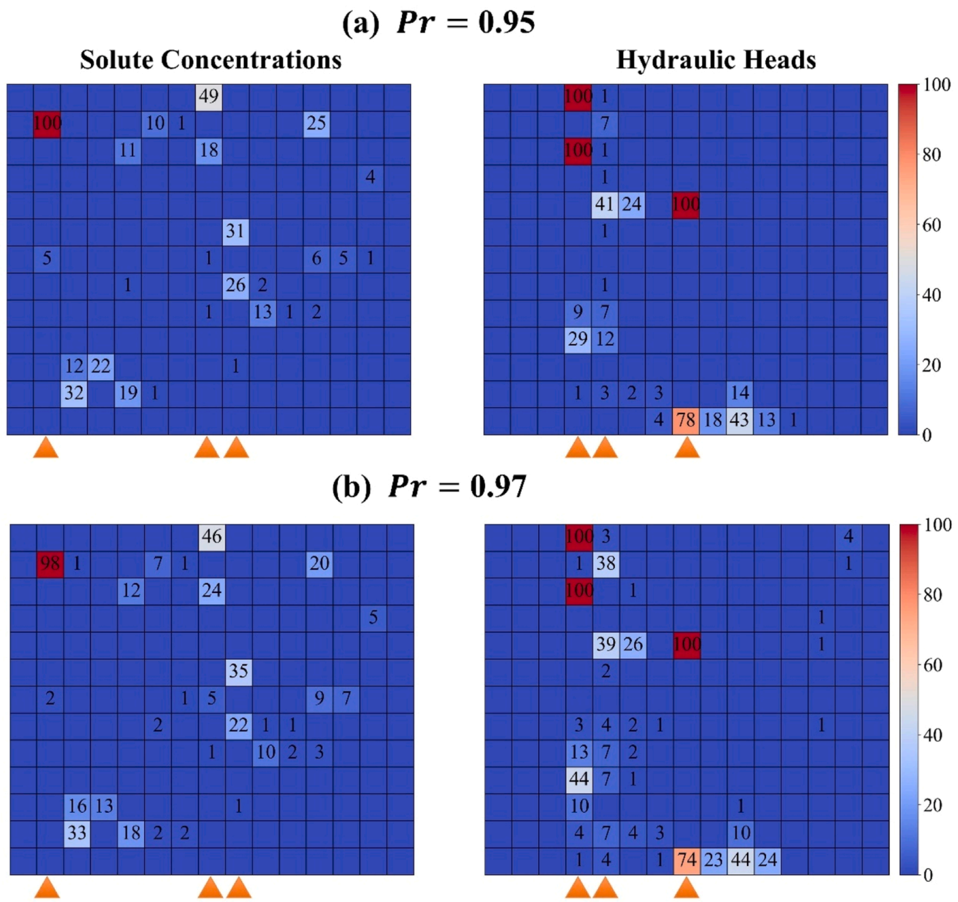


Fig. 14. The selection frequency of CMLs after 100 stochastic optimizations of solute concentration and hydraulic head data (locations with a frequency of 0 are not shown): (a) ***pr***=0.95; (b) ***pr***=0.97. Yellow triangles represent the selected monitoring well locations.

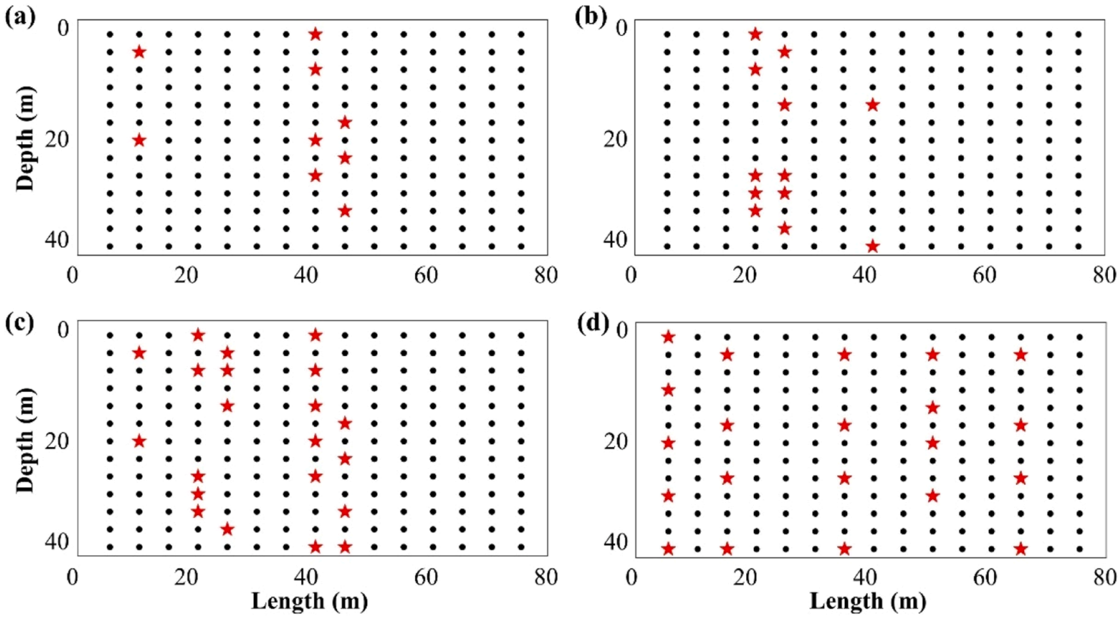


Fig. 15. Different types of optimized monitoring networks: (a) single monitoring network based on solute concentration; (b) single monitoring network based on hydraulic head; (c) fusion monitoring network; (d) stochastic monitoring network. The black dots represent CMLs, and the red stars represent the selected monitoring locations.

3.2.2. Structure identification accuracy

The optimal structure parameters obtained under different scenarios in Section 3.2.1 (i.e., the parameters that minimize the error between the dynamic responses and observations) are selected, and the

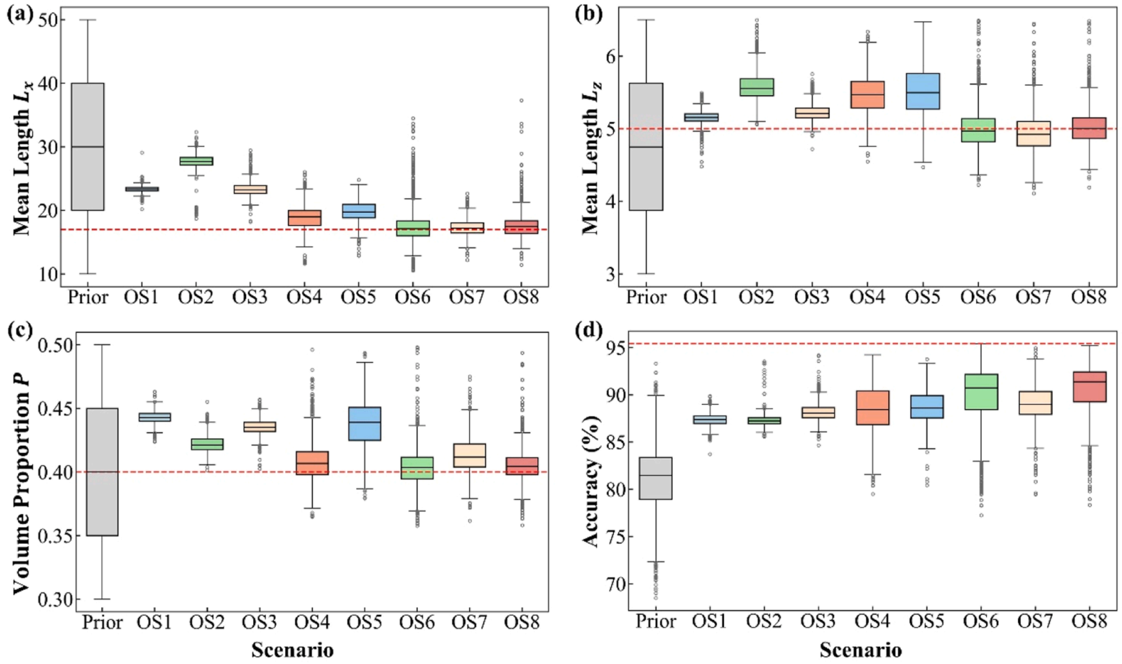


Fig. 16. (a-c) Identification results of structural parameters under different scenarios, where the red dashed lines represent the true values; (d) Structure identification accuracy, where the red dashed line denotes the maximum accuracy across all scenarios.

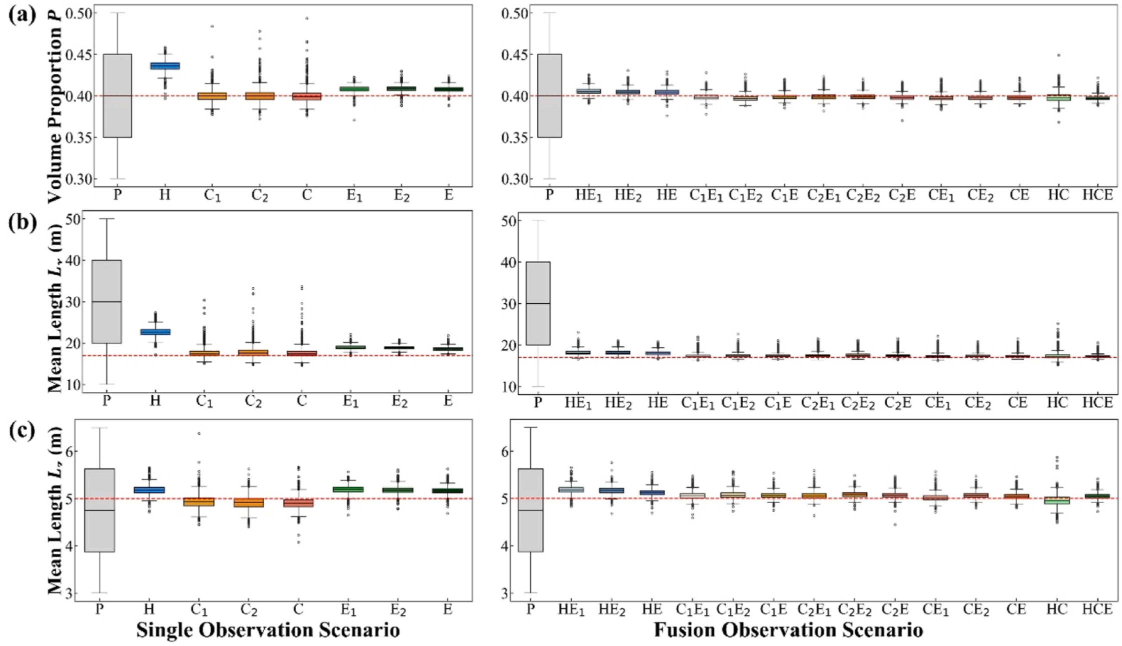


Fig. 17. Structure parameters identification results based on the TP generation model under different scenarios. The red dashed line represents the true structure parameters.

corresponding interval ranges are determined based on these parameters. Detailed information is provided in Table S1 in Supporting Information. 10,000 sets of structure parameters are stochastically sampled within the interval and input into the TP generation model to generate stochastic heterogeneous structures, which serve as the training dataset for the DL generative model. A total of 2000 latent vectors are stochastically generated from a standard normal distribution to serve as prior latent vectors in coupled hydrogeophysical inversion under different scenarios. Perform the same inversion process using the DL generation model for the scenarios described in Table 3 and obtain the posterior latent vectors. Compare the posterior structures generated by both generation methods under different scenarios with the target structure shown in Fig. 7a, and the structure identification accuracy is computed accordingly. Fig. 18a and b present the accuracy results based on the TP generation model, where P_1 represents the prior structures generated from the prior structure parameters. The results clearly show that posterior structures exhibit higher accuracy than prior ones across various identification scenarios. In single-data inversion scenarios, different data types have varying effects on structure identification: in Fig. 18a, hydraulic head data (scenario H) has the lowest accuracy,

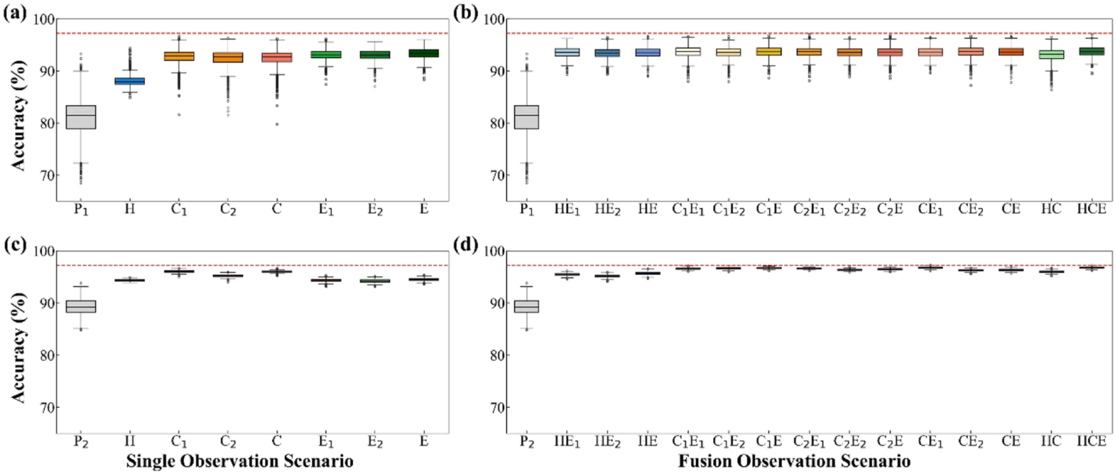


Fig. 18. Structure identification accuracy based on the TP generation model (a, b) and DL generation model (c, d) under different scenarios. The red dashed line represents the maximum accuracy.

while solute concentration data (scenarios C_1, C_2 and C), and ERT data (scenarios E_1, E_2 and E) exhibit similar accuracy levels but with significantly higher uncertainty. Moreover, the accuracy of structure identification is correlated with structure parameters estimation. Larger estimation errors in structure parameters lead to lower structure identification accuracy. When structure parameters estimation errors are within a smaller range, uncertainty becomes a key factor influencing structure identification accuracy. Data fusion enhances accuracy over single-data scenarios, particularly in the fusion of hydraulic head and ERT (scenarios HE_1, HE_2, and HE). As shown in the comparison, Scenario HE produces more precise structures than hydraulic head data and also outperforms the single ERT scenario in overall accuracy. This structural improvement aligns with the parameter estimation trends analyzed in Fig. 17. It is crucial to recognize that the accuracy of structure identification is not determined by any single structure parameter. As previously discussed, due to the limited depth resolution of ERT, the fusion scenario (HE) did not yield a notable improvement in estimating the L_z compared to the hydraulic head-only scenario (H). However, the positive refinements in estimating the L_x and p effectively compensated for this limitation. Consequently, the collective optimization of these key parameters translated into a definitive advantage for Scenario HE in terms of overall structural identification accuracy. Furthermore, the structures obtained from the fused data of solute concentration and ERT (scenarios C_1E_1, C_1E_2, C_1E, C_2E_1, C_2E_2, C_2E, CE_1, CE_2 and CE) have lower uncertainty compared to solute concentration data. The fused data of hydraulic head and solute concentration (scenario HC) produce more accurate structures than hydraulic head, and after integrating ERT data (scenario HCE), accuracy further improves while uncertainty significantly decreases.

Fig. 18c and d show the accuracy results based on the DL generation model, where P_2 represents the prior structures generated from the prior latent vectors. The results clearly indicate that both prior and posterior structures obtained from different scenarios using the DL generation model have higher identification accuracy and lower uncertainty compared to the posterior structures generated by the TP generation model. The conclusion for the single-data inversion scenario differs from the previous findings. The accuracy of hydraulic head (scenario H) improves to the same level as ERT data (scenarios E_1, E_2 and E), while the accuracy of solute concentration surpasses that of ERT data (scenarios C_1, C_2 and C). In addition, it can be clearly observed that the accuracy of scenario C_2 is lower than that of scenarios C_1 and C, indicating that the data worth of solute concentration data at different time points is different. Compared to the TP generation method, the DL generation method can further exploit the data worth, thereby amplifying this difference. Specifically, the latent vectors have 200 dimensions, while the structure parameters have only 3 dimensions. Increasing the dimensionality of the estimated parameters enhances the detailed characterization of heterogeneous structures. Similar to previous findings, the fusion-data inversion scenarios show improved identification accuracy across all datasets after integrating ERT data. Additionally, the scenarios combining solute concentration data at different time points with ERT data exhibit consistently high accuracy. The previously raised concern about redundant information from the fusion of data from the same source does not negatively impact the coupled inversion results.

To further assess the differences in solute transport simulation between the posterior structures obtained from coupled hydrogeophysical inversion and the true structure, the posterior structure ensembles from different scenarios are input into a high-fidelity numerical model to obtain solute concentration fields every other day over 30 days. Breakthrough curves (BTC) are plotted for three stochastically selected observation wells, which differ from those in the fusion monitoring network shown in Fig. 15c. The specific locations of these wells are provided in Figure S8 of the Supplementary Information. Fig. 19 presents the BTC corresponding to the inversion results for hydraulic head and ERT data (scenarios H, HE_1, HE_2, and HE). Each row represents BTC results for different posterior structures at the same location, while each column represents BTC results at different locations for the same posterior structure. The BTC for other scenarios are provided in Figure S9-S13 of the Supplementary Information. Compared to the prior structures, the posterior structures in all scenarios produce reliable BTC predictions. The prediction performance of posterior structures varies across observation locations within the same scenario. For hydraulic head data (scenario H), the posterior BTC at the second observation location exhibits lower uncertainty than at the other two points, but its accuracy is noticeably lower. This is related to the complexity of the heterogeneous structure around the selected location. The BTC predictions from data fusion improve upon those from single data. For instance, at the first observation point, the posterior BTC for the fused data of hydraulic head and ERT (scenarios HE_1, HE_2, and HE) are closer to the true values and exhibit lower uncertainty.

3.2.3. Observation noise and model noise

Coupled hydrogeophysical inversion are conducted under different data noise conditions for scenarios H, C, E, HE, CE and HCE (See Fig. 20, where observation noise and model noise are set to 1 % and 15 %, and 0 % and 15 %, respectively). Structure identification results for other noise conditions are shown in Figure S14-S19 of the Supplementary Information. The accuracy varies across different scenarios. As shown in

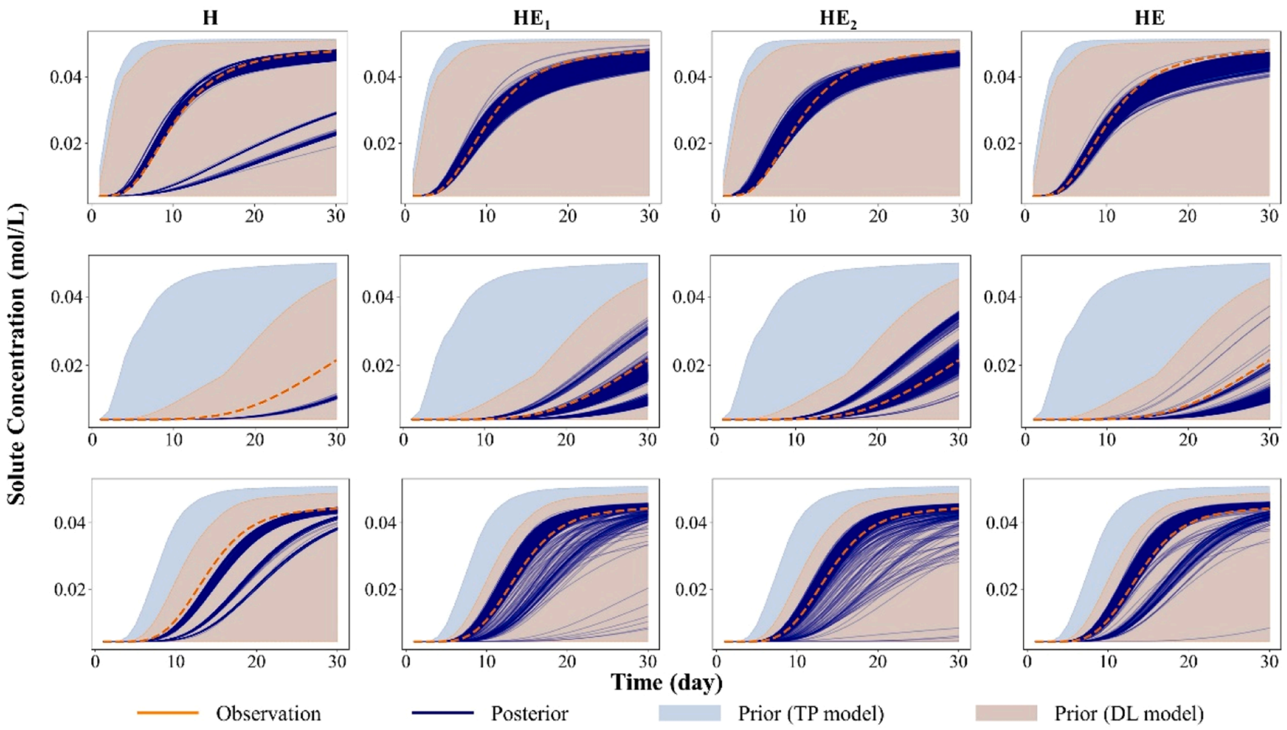


Fig. 19. BTC calculation results for scenarios H, HE$_1$, HE$_2$ and HE. Each row of subplots represents data from a monitoring location, with a total of three monitoring locations.

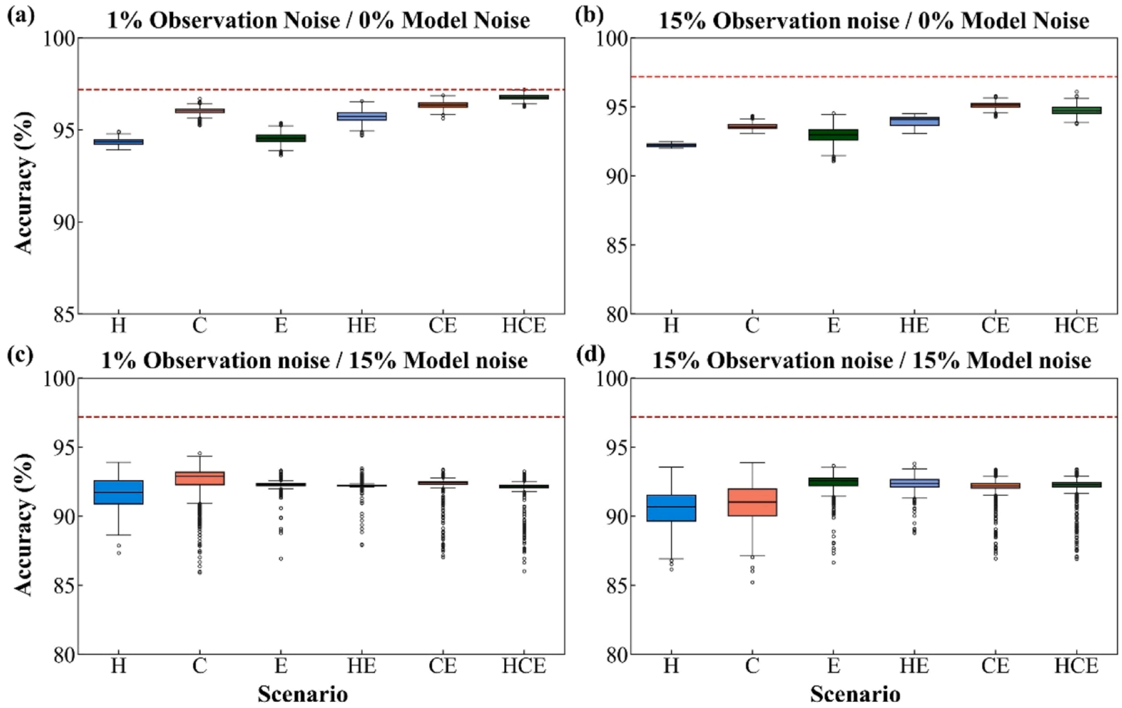


Fig. 20. Structure identification accuracy under different noise conditions across various scenarios. The red dashed line represents the maximum accuracy.

Fig. 21a, when observation noise increases to 15 %, the average accuracy of scenarios H, C, E, HE, CE and HCE decreases by 2.35 %, 2.44 %, 1.59 %, 1.78 %, 1.24 %, and 2.03 %, respectively. Scenarios integrating ERT data experience a smaller decline in average accuracy than other scenarios. This indicates that under 0 % model noise conditions, integrating ERT data helps mitigate the negative impact of observation noise. However, in Fig. 21b, the standard deviation of the ERT data (scenario E) slightly increases compared to other scenarios. This suggests that in the absence of model noise, observation noise affects the uncertainty of ERT data inversion but has little impact on the uncertainty of hydraulic head and solute concentration data inversion. This effect is notably reduced when ERT data is integrated with hydraulic head or solute concentration data. Under the 15 % model noise condition, as observation noise increases (Fig. 20c and d), the accuracy of

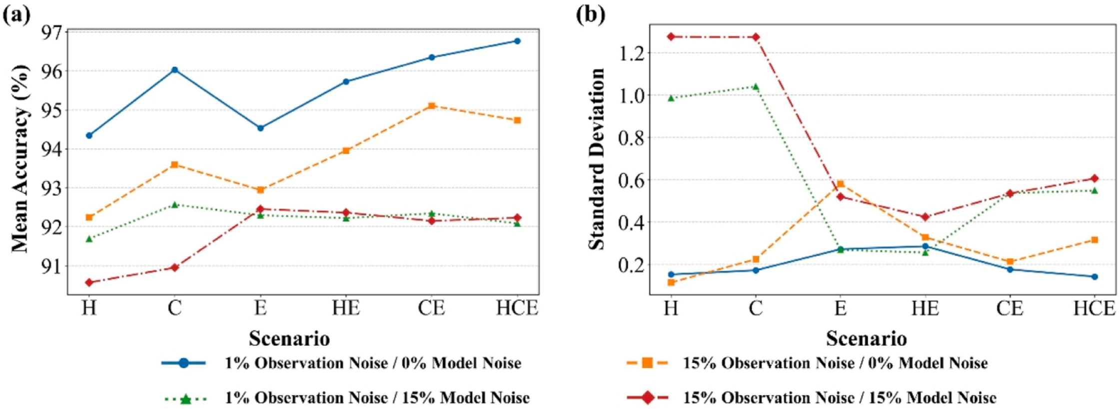


Fig. 21. The mean and standard deviation of structure identification accuracy under different noise conditions across various scenarios.

hydraulic head and solute concentration data (scenarios H and C) decreases. However, the accuracy of scenarios E, HE, CE, and HCE, which incorporate ERT data, remains almost unchanged, further demonstrating the ability of ERT to resist the negative impact of observation noise. Furthermore, Fig. 21a shows that under 1 % observation noise and 15 % model noise conditions, the average accuracy of fusion-data scenarios is already comparable to that of single-data scenarios. With a further increase in observation noise, fusion-data scenarios achieve higher accuracy than single-data scenarios. In terms of uncertainty, as shown in Fig. 21b, under high model noise levels, the uncertainty in scenarios incorporating ERT data is significantly lower than in other scenarios.

Comparing Fig. 20a and c, under the 1 % observation noise condition, an increase in model noise levels leads to variations in both the accuracy and uncertainty of the posterior structure ensembles across different scenarios. As shown in Fig. 21a, the average accuracy of scenarios H, C, E, HE, CE and HCE decreases by 2.65 %, 3.46 %, 2.24 %, 3.51 %, 4.01 %, and 4.68 %. The ERT data experiences the smallest drop in average accuracy. The accuracy decline is more pronounced in fusion-data scenarios than in single-data scenarios, with the largest decrease observed in the scenario HCE, which fuses all the data. The uncertainty of hydraulic head and solute concentration data (scenarios H and C) increases significantly, but in scenarios incorporating ERT data, the uncertainty is effectively controlled, particularly in scenarios E and HE, where it remains almost unchanged. Comparing Fig. 20b and d, a similar trend is observed under the 15 % observation noise condition. For example, as shown in Fig. 21a, the average accuracy across scenarios H, C, E, HE, CE and HCE decreases by 1.43 %, 2.64 %, 0.49 %, 1.59 %, 2.96 %, and 2.51 %, with ERT data experiencing the smallest reduction.

To quantify the sources of posterior uncertainty and elucidate the mechanisms underlying data fusion performance, a formal decomposition analysis based on gradient noise experiments was conducted. The posterior standard deviation, denoted as σ, is defined as the metric for uncertainty spread, and a differential sensitivity method was adopted to isolate the marginal contribution of each noise source. In this decomposition analysis, the explainable total spread (σ_{total}) is conceptualized as the summation of intrinsic uncertainty and the marginal contributions from observation and model noise. It is defined as follows:

$$\sigma_{total} = \sigma_{base} + \Delta_{\sigma_{obs}} + \Delta_{\sigma_{mod}} \tag{13}$$

where σ_{base} represents the intrinsic uncertainty (or the limit of data worth) measured under baseline conditions (1 % observation noise and 0 % model noise). $\Delta_{\sigma_{obs}}$ denotes the marginal contribution of observation noise, calculated as $\max\left(0, \sigma_{high\_obs} - \sigma_{base}\right)$, where σ_{high_obs} corresponds to the scenario with 15 % observation noise and 0 % model noise. Similarly, $\Delta_{\sigma_{mod}}$ represents the marginal contribution of model noise, calculated as $\max\left(0, \sigma_{high\_mod} - \sigma_{base}\right)$, where σ_{high_mod} corresponds to the scenario with 1 % observation noise and 15 % model noise. It should be noted that σ_{total} represents the sum of independent main effects. Although the simultaneous presence of high observation and model noise might induce high-order interaction effects (resulting in an actual spread larger than σ_{total}), this linear decomposition provides a robust first-order approximation sufficient to identify the dominant drivers of uncertainty across different scenarios.

The decomposition results (as shown in Fig. 22) reveal distinct uncertainty characteristics among different data types. Hydrological scenarios (H and C) exhibit low intrinsic uncertainty, yet their posterior spread is overwhelmingly dominated by model noise ($\Delta_{\sigma_{mod}}$ reaches 0.84 and 0.87, respectively, accounting for approximately 80–85 %). Conversely, the contribution of observation noise is negligible. This indicates that while single-source hydrological inversion is robust to observation noise, it is highly sensitive to the prediction accuracy of the DL surrogate model. In contrast, ERT demonstrates strong resistance to model noise, whereas observation noise ($\Delta_{\sigma_{obs}} = 0.31$) constitutes the primary source of its increased uncertainty and accounts for over 53 % of the total spread. Upon the introduction of ERT into the hydraulic head inversion (forming the HE scenario), it acts as a powerful source of complementary information, effectively eliminating the high model noise contribution of 0.84 observed in the H scenario. This enables the HE scenario to maintain stability comparable to the ideal ERT case, even under adverse model conditions. Similarly, in the CE scenario, the inclusion of ERT reduces the massive model noise contribution faced by solute concentration data (0.87) by nearly 60 % (down to 0.36). Notably, this correction is bi-directional. The high-precision H and C

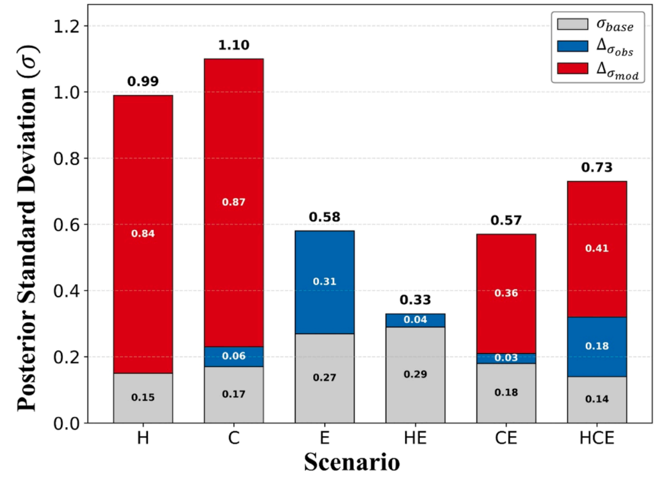


Fig. 22. Quantitative decomposition of posterior uncertainty sources across various scenarios.

data, in turn, almost completely suppress the significant observation noise sensitivity originally inherent in ERT (reducing $\Delta_{\sigma_{obs}}$ from 0.31 to 0.04 and 0.03, respectively). Ultimately, the fully coupled HCE scenario integrates all the aforementioned advantages. It not only maintains the lowest intrinsic uncertainty ($\sigma_{base} = 0.14$), but also achieves an optimal balance of errors globally. Compared to the standalone C scenario, HCE suppresses the dominant model noise contribution by 53 % (down to 0.41) while maintaining observation noise at a controllable level. This demonstrates that multi-source fusion is not merely a simple weighted average; rather, by introducing geophysical constraints, it fundamentally reshapes the error propagation mechanism of hydrogeological inversion.

The prediction error analysis of dynamic response data for posterior structure sets in different scenarios is conducted. This paper uses Mean Absolute Error (MAE) for evaluation:

$$MAE = \frac{\sum_{i=1}^{N_s}\sum_{j=1}^{N_o}\left|r_{i,j} - o_{i,j}\right|}{N_s N_o} \tag{14}$$

where $r_{i,j}$ and $o_{i,j}$ represent the normalized jth dynamic response data and the actual observation data corresponding to the ith posterior structure, respectively. N_s and N_o denote the number of posterior structures and dynamic observation data.

Fig. 23 presents the MAE calculations for hydraulic head, solute concentration, and ERT data under different noise conditions. Overall, the MAE of hydraulic head data remains relatively low across all scenarios and stays stable even under high-noise conditions. The solute concentration data generally have the highest MAE and are more susceptible to noise compared to hydraulic head and ERT data. Assimilating a specific type of data improves the prediction accuracy of that dynamic responses but has limited constraint on other types. For example, the MAE of hydraulic head is the lowest, while the errors for solute concentration and ERT are higher. In scenario E, the MAE of ERT is the lowest, but the errors for hydraulic head and solute concentration are higher. Data fusion enhances the prediction accuracy of the assimilated data types but still cannot effectively constrain unassimilated data types. For example, the MAE of solute concentration data is significantly higher than in scenarios HC and HCE. Comparing Fig. 23a and b, under low model noise conditions, the MAE of solute concentration data increases significantly as observation noise increases, while the MAE of hydraulic head data shows a slight increase. Meanwhile, the MAE of ERT data remains stable at a low level. Comparing Fig. 23c and d, under high model noise conditions, no clear increasing trend in MAE is observed as observation noise increases. Furthermore, regardless of the observation noise level, an increase in model noise has a significant impact on the MAE of all data types. This indicates that model noise has a greater impact on the prediction error of dynamic response data than observation noise. When model noise is high, the parameter update direction in ILUES is predominantly influenced by model noise, meaning that nearly all errors between the dynamic response and observation data originate from model noise.

4. Summary and conclusions

This study first proposes the transition probability and entropy-based monitoring network optimization method to identify monitoring locations with higher data worth, effectively addressing the issue of insufficient data worth caused by the randomness of monitoring locations in traditional intrusive hydrological sampling methods. On this basis, the study explores the impact mechanism of data noise on data worth within the coupled hydrogeophysical inversion framework. Specifically, 21 different data assimilation scenarios are designed for hydraulic head, solute concentration, and ERT data to perform structure identification. Subsequently, six optimal data fusion strategies are selected, and a total of 96 structure identification scenarios are considered under varying levels of observation noise and model noise. The following are the insights gained from this study.

4.1. Method performance

4.1.1. Monitoring network optimization

The proposed TP and entropy-based monitoring network optimization effectively selects valuable monitoring locations from densely distributed CMLs. Furthermore, the monitoring well prioritization

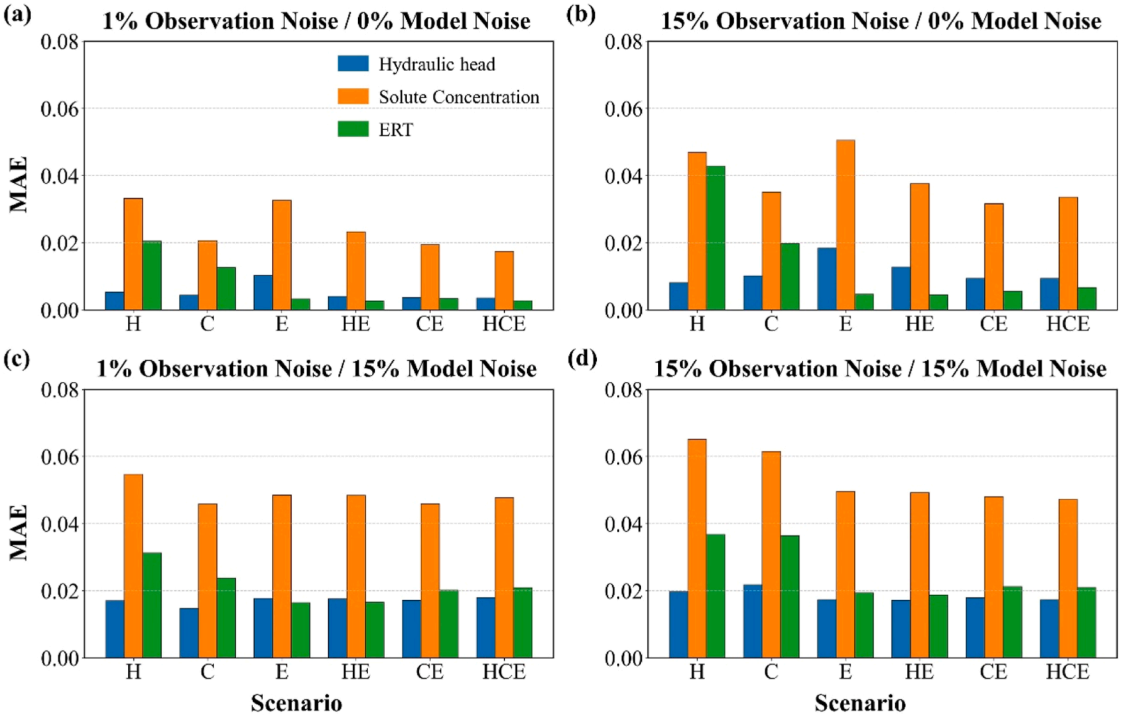


Fig. 23. Comparison of the MAE between dynamic responses and observations under different noise conditions in different scenarios.

strategy helps balance monitoring costs and data worth in practical applications. TP stochastic simulation generates numerous stochastic structures using limited borehole data. Combined with Monte Carlo simulation, it enhances the diversity of the dynamic response dataset, effectively addressing the uncertainty associated with traditional monitoring network optimization. In this process, the DL surrogate model efficiently obtains dynamic response datasets. To mitigate potential entropy errors induced by surrogate bias, a robust optimization framework based on selection frequency is adopted. Specifically, by constructing a frequency matrix through Monte Carlo iterations, the method functions as a statistical filter to prioritize monitoring locations that are consistently selected. This mechanism reduces the sensitivity of optimization results to model noise. Finally, by quantifying the data worth of each monitoring location, the optimized network consists of only a small number of valuable monitoring locations, achieving satisfactory results in structure inversion.

In the process of data worth assessment, the discretization constant and data size of dynamic response data are key factors affecting the accuracy of entropy analysis. An excessively large or small discretization constant can result in incorrect characterization of data variability, while insufficient data size may cause the frequency distribution to deviate from the true probability density function, affecting the stability of optimization results. Therefore, in practical applications, these parameters should be reasonably set based on the hydrogeological characteristics of the target area.

4.1.2. Coupled hydrogeophysical inversion

Under the general framework for coupled hydrogeophysical inversion, different types of observations can be flexibly assimilated for precise structure identification. The introduction of DL methods further enhances the performance of this framework, primarily because DL enables the parameterization of heterogeneous structures using a latent vector with customizable dimensions, allowing for higher-dimensional information representation. However, the dimensionality of the latent vector must balance computational cost and structure refinement needs to suit specific application scenarios.

The results of this study indicate that assimilating at least one type of observation data significantly improves identification accuracy. However, a single data source has limitations in spatial resolution and coverage, and the characteristics of the information it provides also differ. For example, the structure identification accuracy and uncertainty of solute concentration data are relatively high, and the corresponding dynamic response prediction accuracy is poor. ERT data help reduce uncertainty in identification results and exhibit lower dynamic response prediction errors. Additionally, assimilating a specific type of data improves the prediction accuracy of its corresponding dynamic response but has limited constraints on other types of dynamic responses. For instance, assimilating ERT data significantly reduces ERT prediction errors; however, compared to assimilating solute concentration data, the prediction errors for concentration increase.

The advantage of data fusion lies in improving identification accuracy and reducing uncertainty through complementary constraints. However, when the sensitivity of the integrated data to the target parameters is insufficient, its enhancement effect will be limited. For example, ERT data provide weak constraints on the average extension length parameter in the depth direction, mainly because their resolution decreases rapidly with depth, limiting their ability to capture deep structure variations. This attenuation effect originates from energy loss in the propagation of the electric field signal, which not only reduces the reliability of deep inversion results but also increases the uncertainty in the mapping relationships between resistivity and parameters such as water content and solute concentration (Cassiani et al., 2012). It is important to emphasize that the proposed framework is not intrinsically dependent on a fixed resistivity-concentration relationship. When the petrophysical relationship is unknown or spatially variable, it may be addressed through different modeling strategies, for example by estimating petrophysical parameters from site-specific measurements or by incorporating theoretical relationships, each with their respective advantages and limitations. Alternatively, petrophysical uncertainty can be handled by treating relevant parameters as uncertain inputs or by absorbing their effects into observation error or model discrepancy terms, without altering the overall inversion strategy. When designing data fusion schemes, the sensitivity and complementarity of different data types should be carefully considered to maximize data worth. Crucially, to avoid detrimental fusion outcomes, it is necessary to recognize that fusion is beneficial only when the information provided by the new data outweighs the introduced noise. Future applications should prioritize a preliminary sensitivity analysis, such as evaluating the depth of investigation, to ensure that the assimilated data possess sufficient physical constraints on the target parameters. This step is vital to prevent performance degradation caused by incorporating data with low sensitivity but high uncertainty.

4.2. Observation noise and model noise

This study further explores the impacts of observation noise and model noise on data worth. When the prediction of dynamic responses is sufficiently accurate (i.e., model noise is low), observation noise tends to reduce the accuracy of the structure identification while having a limited effect on uncertainty. The prediction accuracy of solute concentration data is particularly sensitive to observation noise, whereas the accuracy of hydraulic head data remains relatively stable. This difference is likely influenced by the specific boundary conditions and the simplified synthetic domain adopted in this work. Under the conditions investigated in this study, solute concentration data exhibit significantly higher spatiotemporal variability than hydraulic head data, indicating that solute concentration is more sensitive to localized heterogeneities in the structure. Additionally, ERT data appear to help mitigate the negative effects of observation noise. In scenarios incorporating ERT data, the decline in accuracy due to increased observation noise is significantly smaller than in other scenarios, and the dynamic response prediction error of ERT data remains consistently low.

As model noise increases, the accuracy of structure identification decreases significantly across all scenarios, while uncertainty rises simultaneously, leading to higher dynamic response prediction errors. Assimilating only ERT data helps mitigate the impact of model noise on structure identification. However, this capability weakens when ERT data are integrated with other data, and the decline in accuracy becomes more pronounced as the number of fused data types increases. This phenomenon suggests that model noise may be further amplified in data fusion scenarios. Additionally, under high model noise conditions, the impact of increasing observation noise on identification results and dynamic response predictions diminishes, suggesting that model noise exerts a stronger influence than observation noise.

In fact, data assimilation algorithms typically attribute deviations between predicted responses and observations to differences between the simulated and actual structures. However, disturbances caused by data noise can lead the data assimilation algorithm to update parameters in the wrong direction. High-worth data are often characterized by strong spatiotemporal variability. Under low-noise conditions, data assimilation algorithms can effectively extract useful information from such data but are also more susceptible to noise interference, potentially resulting in erroneous representations of localized heterogeneities. Observation noise, which arises from monitoring conditions or inherent sensor precision, is difficult to quantify in practice. It is typically mitigated through preprocessing techniques such as data cleaning. While in this study model noise originates from prediction errors in the DL model, it also highlights the overlooked impact of numerical model simulation errors on identification results. Crucially, although the model noise analysis framework proposed in this study is not tied to a specific surrogate architecture and is primarily intended to facilitate a quantitative understanding of the relationship between surrogate model accuracy and inversion uncertainty, our analysis clearly highlights the necessity of improving baseline model accuracy. Acknowledging that the current data-driven surrogate does not explicitly enforce governing equations, future work will focus on integrating physics-informed architectures that incorporate physical conservation laws into the loss function as soft penalties, thereby ensuring physical consistency and reducing model error at its source. Additionally, the ERT data used in this study demonstrate strong noise resistance, highlighting the potential practical value of the coupled hydrogeophysical inversion framework. However, it is important to note that these findings are based on a synthetic study with specific assumptions regarding noise distribution and geological stationarity. Real-world applications may present more complex challenges. Future studies could explore integrating additional reliable observations into this framework to better adapt to these complex and challenging application scenarios. Practically, these findings suggest a strategic shift in site characterization. In field environments characterized by high measurement uncertainty engineers should prioritize the deployment of surface ERT over reliance on sparse solute concentration sampling. While solute data theoretically offers high information content, its sensitivity to noise renders it less reliable for structure identification in practical, noisy settings compared to the robust volumetric coverage of ERT. Furthermore, our results serve as a cautionary note for data fusion: simply increasing data types is not a panacea. Engineers must ensure the fidelity of the baseline physical model, as model errors can be amplified in fusion scenarios, potentially negating the benefits of multi-source monitoring.

4.3. Limitations and future prospects

Although this study validates the effectiveness of the proposed method through synthetic experiments, applying this framework to practical engineering problems requires addressing challenges related to dimensionality and environmental complexity.

First, regarding scalability to realistic 3D scenarios, the transition from 2D to 3D significantly increases the degrees of freedom for structure identification and the computational burden of physical simulations (e.g., TOUGHREACT and SimPEG) used for training data generation. However, the DL-based workflow proposed in this study offers a scalable solution to this challenge. Future work will focus on extending the framework to 3D domains by leveraging high-performance computing for offline dataset generation and adopting 3D Convolutional Neural Networks or Graph Neural Networks as surrogate models. Since the heavy computational load is shifted to the offline training phase, the high efficiency of the DL surrogate model during the online inference phase ensures the feasibility of the method for real-time 3D monitoring network optimization and inversion. In 2D models, flow is often forced to traverse low-permeability heterogeneities, whereas 3D environments permit fluid to bypass such obstacles. While this implies that 2D models impose stronger hydraulic constraints, it theoretically underscores the necessity of the proposed data fusion strategy for 3D applications. The bypass phenomenon in 3D attenuates hydraulic signals induced by heterogeneities, rendering the inverse problem more ill-posed when relying solely on hydraulic data. Consequently, incorporating spatially sensitive geophysical data (like ERT), which maintains volumetric sensitivity independent of flow topology, becomes even more critical. Thus, the conclusions regarding data worth drawn from this 2D study serve as a baseline; the advantages of multi-source fusion are likely amplified in the more challenging 3D context.

Second, concerning geological realism and uncertainty characterization, this study adopts stationary transition probability-based models and Gaussian observation noise as controlled test cases to systematically evaluate data worth and noise sensitivity. In practical field applications, subsurface heterogeneity is often non-stationary (e.g., fining-upward sequences), boundary conditions are transient (e.g., time-varying pumping or recharge), and measurement errors may exhibit heavy-tailed distributions or systematic biases due to sensor drift and calibration uncertainty. These factors represent genuine challenges for any inversion framework rather than shortcomings unique to the method. From a methodological perspective, the proposed framework is not restricted to stationary geological priors or Gaussian noise assumptions. Non-stationary geological patterns can be incorporated by enriching the training dataset using advanced geostatistical techniques such as Multiple-Point Statistics. Likewise, non-Gaussian and biased noise can be explicitly introduced during the training phase through noise-aware data augmentation, enabling the surrogate model to learn more realistic error structures. Moreover, transfer learning and domain adaptation strategies provide practical pathways to bridge the gap between synthetic training data and field observations, thereby improving robustness against model errors, systematic bias, and complex boundary conditions. Overall, while the present study focuses on a controlled synthetic setting to isolate fundamental behaviors, the framework is designed with sufficient flexibility to accommodate the key sources of complexity encountered in real-world hydrogeophysical inversion, which will be the focus of future field-scale validation.

Open research

The data and code used in this study are available at https://doi.org/10.5281/zenodo.18082442 and https://doi.org/10.5281/zenodo.7088956.

CRediT authorship contribution statement

Yuzhou Xia: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **Chuanjun Zhan:** Writing – review & editing, Visualization, Methodology, Conceptualization. **Zhenxue Dai:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Jichun Wu:** Writing – review & editing. **Xiaoying Zhang:** Writing – review & editing. **Huichao Yin:** Writing – review & editing. **Jiahe Yan:** Writing – review & editing, Methodology. **Junjun Chen:** Writing – review & editing, Methodology. **Zihao Wang:** Writing – review & editing. **Mohamad Reza Soltanian:** Writing – review & editing. **Kenneth C. Carroll:** Writing – review & editing.

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 is funded by the National Key R&D Program of China (2024YFC3713800), the National Natural Science Foundation of China (NSFC: U2267217, 42141011, 42502245, 42402241), Shandong Key Water Conservancy Science and Technology Project (2024370203001957) and Shandong Provincial Natural Science Foundation (ZR2024QD095).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2025.105204.

Appendix A:List of acronyms

Acronyms	Full Term
DA	data assimilation
CMLs	candidate monitoring locations
MIMR	maximum information and minimum redundancy
ERT	Electrical Resistivity Tomography
DL	deep learning
Pr	predefined ratio
TP	transition probability
DOCRP	Deep Octave Convolution Residual Network for Point-Information
DOCRI	Deep Octave Convolution Residual Network for Image-Information
DOCRP-ST	DOCRP predicting sum of solute concentrations and single-time hydraulic head
DOCRI-ST	DOCRI predicting sum of solute concentrations and single-time hydraulic head
DOCRP-TT	DOCRP predicting single-time solute concentration and hydraulic head
DOCRI-AA	DOCRI predicting single-time solute concentration and hydraulic head
ILUES	Iterative Local Updating Ensemble Smoother

Appendix B: Primer on key concepts for cross-disciplinary readers

To bridge the gap between machine learning and hydrogeophysics, this appendix provides brief explanations of the core concepts and terminologies used in this study.

B.1. Hydrogeology and geophysics (for readers with an machine learning background)

Hydraulic Head: In simple terms, this represents the energy level of groundwater (combining pressure and elevation). Water flows from high hydraulic head to low hydraulic head.

Contaminant Transport (Solute Concentration): This refers to how pollutants move through groundwater. The solute concentration is the amount of pollutant in the water. Predicting this is the ultimate goal of environmental monitoring.

Electrical Resistivity Tomography (ERT): A geophysical imaging technique similar to a medical CT scan. It injects electrical current into the ground and measures the voltage to calculate resistivity. Since different materials conduct electricity differently, ERT helps us see the subsurface geological structure without digging.

Transition Probability: A geostatistical method used to generate 2D/3D models of subsurface structures. It uses Markov Chains to simulate the probability of one soil type transitioning into another (e.g., sand changing to clay) over a certain distance. It is used here to create realistic geological patterns for training the DL model.

B.2. Machine learning and optimization (for readers with a geoscience background)

Surrogate Model: A substitute model created using deep learning. Physical simulations (solving differential equations for groundwater flow) are often very slow. We train a DL model to mimic the physical simulator. Once trained, this surrogate model can predict results almost instantly, significantly speeding up the optimization process.

Inversion (Structure Identification): The process of working backward from data to cause. Here, we observe data (hydraulic head, solute concentration, resistivity) and use algorithms to estimate the unknown subsurface structure (the input parameters). In machine learning terms, this is an "inverse problem" or "parameter estimation."

Entropy: A measure of information content from information theory. In this study, entropy quantifies the variability of predicted state variables across different geological realizations. A location with high entropy indicates that the observations there are highly sensitive to the subsurface structure, representing high data worth for distinguishing between different geological models.

Monitoring Network Optimization: The process of deciding the best locations to place sensors. It involves identifying locations with the highest data worth to maximize the information gained from observations. The goal is to obtain the most useful information for characterizing the subsurface using the minimum number of sensors.

Data availability

Data will be made available on request.

References

Aeschbach-Hertig, W., Gleeson, T., 2012. Regional strategies for the accelerating global problem of groundwater depletion. Nat. Geosci. 5 (12), 853–861. https://doi.org/10.1038/ngeo1617.

Aghasi, A., Mendoza-Sanchez, I., Miller, E.L., Ramsburg, C.A., Abriola, L.M., 2013. A geometric approach to joint inversion with applications to contaminant source zone characterization. Inverse Probl. 29 (11), 115014.

Alfonso, L., Lobbrecht, A., Price, R., 2010a. Information theory–based approach for location of monitoring water level gauges in polders. Water Resour. Res. 46 (3). https://doi.org/10.1029/2009WR008101.

Alfonso, L., Lobbrecht, A., Price, R., 2010b. Optimization of water level monitoring network in polder systems using information theory. Water Resour. Res. 46 (12). https://doi.org/10.1029/2009WR008953.

Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. Trans. AIME 146 (01), 54–62. https://doi.org/10.2118/942054-g.

Azadi, S., Amiri, H., Ataei, P., Javadpour, S., 2020. Optimal design of groundwater monitoring networks using gamma test theory. Hydrogeol. J. 28 (4), 1389–1402. https://doi.org/10.1007/s10040-020-02115-z.

Barati Moghaddam, M., Mazaheri, M., Mohammad Vali Samani, J., 2021. Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: a review. Groundw. Sustain. Dev. 15, 100651. https://doi.org/10.1016/j.gsd.2021.100651.

Berg, S.J., Illman, W.A., 2011. Capturing aquifer heterogeneity: comparison of approaches through controlled sandbox experiments. Water Resour. Res. 47 (9). https://doi.org/10.1029/2011WR010429.

Binley, A., 2015. 11.08 - Tools and techniques: electrical methods. In: Schubert, G. (Ed.), Treatise on Geophysics (Second Edition). Elsevier, Oxford, pp. 233–259.

Binley, A., Cassiani, G., Middleton, R., Winship, P., 2002. Vadose zone flow model parameterisation using cross-borehole radar and resistivity imaging. J. Hydrol. 267 (3), 147–159. https://doi.org/10.1016/S0022-1694(02)00146-4.

Binley, A., et al., 2015. The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. Water Resour. Res. 51 (6), 3837–3866. https://doi.org/10.1002/2015wr017016.

Camporese, M., Cassiani, G., Deiana, R., Salandin, P., Binley, A., 2015. Coupled and uncoupled hydrogeophysical inversions using ensemble Kalman filter assimilation of ERT-monitored tracer test data. Water Resour. Res. 51 (5), 3277–3291. https://doi.org/10.1002/2014wr016017.

Canchumuni, S.A., Emerick, A.A., Pacheco, M.A., 2017. Integration of Ensemble Data Assimilation and Deep Learning for History Matching Facies Models. OTC Brasil.

Cassiani, G., et al., 2012. Noninvasive monitoring of soil static characteristics and dynamic states: a case study highlighting vegetation effects on agricultural land. Vadose Zone J. 11 (3). https://doi.org/10.2136/vzj2011.0195.

Chan, S., Elsheikh, A.H., 2019. Parametric generation of conditional geological realizations using generative neural networks. Comput. Geosci. 23 (5), 925–952. https://doi.org/10.1007/s10596-019-09850-7.

Chang, Z., et al., 2024. A comparison of inversion methods for surrogate-based groundwater contamination source identification with varying degrees of model complexity. Water Resour. Res. 60 (4). https://doi.org/10.1029/2023wr036051.

Chen, J., et al., 2022. Integration of deep learning and information theory for designing monitoring networks in heterogeneous aquifer systems. Water Resour. Res. 58 (10). https://doi.org/10.1029/2022wr032429.

Cheng, Q., Chen, X., Tao, M., Binley, A., 2019. Characterization of karst structures using quasi-3D electrical resistivity tomography. Environ. Earth Sci. 78, 1–12. https://doi.org/10.1007/s12665-019-8284-2.

Cockett, R., Kang, S., Heagy, L.J., Pidlisecky, A., Oldenburg, D.W., 2015. SimPEG: an open source framework for simulation and gradient based parameter estimation in geophysical applications. Comput. Geosci. 85, 142–154. https://doi.org/10.1016/j.cageo.2015.09.015.

Dai, Z., et al., 2022. An integrated experimental design framework for optimizing solute transport monitoring locations in heterogeneous sedimentary media. J. Hydrol. 614. https://doi.org/10.1016/j.jhydrol.2022.128541.

Dai, Z., et al., 2014. An integrated framework for optimizing CO2 sequestration and enhanced oil recovery. Environ. Sci. Technol. Lett. 1 (1), 49–54. https://doi.org/10.1021/ez4001033.

Danquigny, C., Ackerer, P., Carlier, J.P., 2004. Laboratory tracer tests on three-dimensional reconstructed heterogeneous porous media. J. Hydrol. 294 (1), 196–212. https://doi.org/10.1016/j.jhydrol.2004.02.008.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010. Quantifying data worth toward reducing predictive uncertainty. Groundwater 48 (5), 729–740. https://doi.org/10.1111/j.1745-6584.2010.00679.x.

de Barros, F.P.J., Ezzedine, S., Rubin, Y., 2012. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. Adv. Water Resour. 36, 51–63. https://doi.org/10.1016/j.advwatres.2011.05.004.

Deng, Y., et al., 2024. Characterization of discrete fracture networks with deep-learning based hydrogeophysical inversion. J. Hydrol. 631. https://doi.org/10.1016/j.jhydrol.2024.130819.

Emery, X., 2008. Uncertainty modeling and spatial prediction by multi-Gaussian kriging: accounting for an unknown mean value. Comput. Geosci. 34 (11), 1431–1442. https://doi.org/10.1016/j.cageo.2007.12.011.

Fahle, M., Hohenbrink, T.L., Dietrich, O., Lischeid, G., 2015. Temporal variability of the optimal monitoring setup assessed using information theory. Water Resour. Res. 51 (9), 7723–7743. https://doi.org/10.1002/2015WR017137.

Fernández-Muñiz, Z., Khaniani, H., Fernández-Martínez, J.L., 2019. Data kit inversion and uncertainty analysis. J. Appl. Geophy. 161, 228–238. https://doi.org/10.1016/j.jappgeo.2018.12.022.

Ferrari, A., Jimenez-Martinez, J., Borgne, T.L., Méheust, Y., Lunati, I., 2015. Challenges in modeling unstable two-phase flow experiments in porous micromodels. Water Resour. Res. 51 (3), 1381–1400. https://doi.org/10.1002/2014WR016384.

Gharasoo, M., et al., 2019. A robust optimization technique for analysis of multi-tracer experiments. J. Contam. Hydrol. 224, 103481. https://doi.org/10.1016/j.jconhyd.2019.04.004.

Gleeson, T., et al., 2020. Illuminating water cycle modifications and Earth system resilience in the Anthropocene. Water Resour. Res. 56 (4), e2019WR024957. https://doi.org/10.1029/2019WR024957.

Han, Z., Kang, X., Wu, J., Shi, X., 2022. Characterization of the non-Gaussian hydraulic conductivity field via deep learning-based inversion of hydraulic-head and self-potential data. J. Hydrol. 610. https://doi.org/10.1016/j.jhydrol.2022.127830.

Heagy, L.J., Cockett, R., Kang, S., Rosenkjaer, G.K., Oldenburg, D.W., 2017. A framework for simulation and inversion in electromagnetics. Comput. Geosci. 107, 1–19. https://doi.org/10.1016/j.cageo.2017.06.018.

Hinnell, A.C., et al., 2010. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. Water Resour. Res. 46 (4). https://doi.org/10.1029/2008wr007060.

Hosseini, M., Kerachian, R., 2017. A data fusion-based methodology for optimal redesign of groundwater monitoring networks. J. Hydrol. 552, 267–282. https://doi.org/10.1016/j.jhydrol.2017.06.046.

Huisman, J.A., Rings, J., Vrugt, J.A., Sorg, J., Vereecken, H., 2010. Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion. J. Hydrol. 380 (1), 62–73. https://doi.org/10.1016/j.jhydrol.2009.10.023.

Irving, J., Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resour. Res. 46 (11). https://doi.org/10.1029/2009wr008340.

Jackson, R., Jin, M., 2005. The measurement of DNAPL in low-permeability lenses within alluvial aquifers by partitioning tracers. Environ. Eng. Geosci. 11 (4), 405–412. https://doi.org/10.2113/11.4.405.

JafarGandomi, A., Binley, A., 2013. A Bayesian trans-dimensional approach for the fusion of multiple geophysical datasets. J. Appl. Geophy. 96, 38–54. https://doi.org/10.1016/j.jappgeo.2013.06.004.

Jamil, A., et al., 2024. Comparison of machine learning and electrical resistivity arrays to inverse modeling for locating and characterizing subsurface targets. J. Appl. Geophy. 229. https://doi.org/10.1016/j.jappgeo.2024.105493.

Johnson, T.C., et al., 2012. Characterization of a contaminated wellfield using 3D electrical resistivity tomography implemented with geostatistical, discontinuous boundary, and known conductivity constraints. Geophysics 77 (6). https://doi.org/10.1190/geo2012-0121.1. EN85-EN96.

Johnson, T.C., Versteeg, R.J., Ward, A., Day-Lewis, F.D., Revil, A., 2010. Improved hydrogeophysical characterization and monitoring through parallel modeling and inversion of time-domain resistivity andinduced-polarization data. Geophysics 75 (4), WA27–WA41. https://doi.org/10.1190/1.3475513.

Kang, X., et al., 2021a. Hydrogeophysical characterization of nonstationary DNAPL source zones by integrating a convolutional variational autoencoder and ensemble smoother. Water Resour. Res. 57 (2). https://doi.org/10.1029/2020wr028538.

Kang, X., et al., 2021b. Integrating deep learning-based data assimilation and hydrogeophysical data for improved monitoring of DNAPL source zones during remediation. J. Hydrol. 601. https://doi.org/10.1016/j.jhydrol.2021.126655.

Kang, X., et al., 2018. Coupled hydrogeophysical inversion of DNAPL source zone architecture and permeability field in a 3D heterogeneous sandbox by assimilation time-lapse cross-borehole electrical resistivity data via ensemble Kalman filtering. J. Hydrol. 567, 149–164. https://doi.org/10.1016/j.jhydrol.2018.10.019.

Keum, J., Coulibaly, P., 2017. Information theory-based decision support system for integrated design of multivariable hydrometric networks. Water Resour. Res. 53 (7), 6239–6259. https://doi.org/10.1002/2016WR019981.

Keum, J., et al., 2018. Application of SNODAS and hydrologic models to enhance entropy-based snow monitoring network design. J. Hydrol. 561, 688–701. https://doi.org/10.1016/j.jhydrol.2018.04.037.

Keum, J., Kornelsen, K.C., Leach, J.M., Coulibaly, P., 2017. Entropy applications to water monitoring network design: a review. Entropy 19 (11), 613. https://doi.org/10.3390/e19110613.

Kornelsen, K.C., Coulibaly, P., 2015. Design of an optimal soil moisture monitoring network using SMOS retrieved soil moisture. IEEE Trans. Geosci. Remote Sens. 53 (7), 3950–3959.

Kuang, X., et al., 2024. The changing nature of groundwater in the global water cycle. Science 383 (6686). https://doi.org/10.1126/science.adf0630.

Laloy, E., Hérault, R., Jacques, D., Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network. Water Resour. Res. 54 (1), 381–406. https://doi.org/10.1002/2017wr022148.

Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. Adv. Water Resour. 110, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029.

Li, C., Singh, V.P., Mishra, A.K., 2012. Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy. Water Resour. Res. 48 (5). https://doi.org/10.1029/2011wr011251.

Linde, N., Renard, P., Mukerji, T., Caers, J., 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review. Adv. Water Resour. 86, 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019.

Lopez-Alvis, J., Nguyen, F., Looms, M.C., Hermans, T., 2022. Geophysical inversion using a variational autoencoder to model an assembled spatial prior uncertainty. J. Geophys. Res.: Solid Earth 127 (3). https://doi.org/10.1029/2021jb022581.

Luo, X., Kwok, K.L., Liu, Y., Jiao, J., 2017. A permanent multilevel monitoring and sampling system in the coastal groundwater mixing zones. Groundwater 55 (4), 577–587. https://doi.org/10.1111/gwat.12510.

McClymont, A.F., et al., 2011. Investigating groundwater flow paths within proglacial moraine using multiple geophysical methods. J. Hydrol. 399 (1), 57–69. https://doi.org/10.1016/j.jhydrol.2010.12.036.

McLachlan, P., Chambers, J., Uhlemann, S., Sorensen, J., Binley, A., 2020. Electrical resistivity monitoring of river–groundwater interactions in a Chalk river and neighbouring riparian zone. Near. Surf. Geophys. 18, 385–398. https://doi.org/10.1002/nsg.12114. Geoelectrical Monitoring.

Mishra, A.K., Coulibaly, P., 2010. Hydrometric network evaluation for Canadian watersheds. J. Hydrol. 380 (3), 420–437. https://doi.org/10.1016/j.jhydrol.2009.11.015.

Mo, S., Zabaras, N., Shi, X., Wu, J., 2020. Integration of adversarial autoencoders with residual dense convolutional networks for estimation of Non-Gaussian hydraulic conductivities. Water Resour. Res. 56 (2). https://doi.org/10.1029/2019wr026082.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S., 2013. Data assimilation and parameter estimation via ensemble Kalman filter coupled with stochastic moment equations of transient groundwater flow. Water Resour. Res. 49 (3), 1334–1344. https://doi.org/10.1002/wrcr.20113.

Piersol, A., Bendat, J., 1966. Measurement and Analysis of Random Data. John Willey &Sons.

Pollock, D., Cirpka, O.A., 2010. Fully coupled hydrogeophysical inversion of synthetic salt tracer experiments. Water Resour. Res. 46 (7). https://doi.org/10.1029/2009wr008575.

Pollock, D., Cirpka, O.A., 2012. Fully coupled hydrogeophysical inversion of a laboratory salt tracer experiment monitored by electrical resistivity tomography. Water Resour. Res. 48 (1). https://doi.org/10.1029/2011wr010779.

Power, C., et al., 2015. Improved time-lapse electrical resistivity tomography monitoring of dense non-aqueous phase liquids with surface-to-horizontal borehole arrays. J. Appl. Geophy. 112, 1–13. https://doi.org/10.1016/j.jappgeo.2014.10.022.

Rakotonirina, N.C., Rasoanaivo, A., 2020. ESRGAN+ : further improving enhanced Super-resolution generative adversarial network. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3637–3641.

Ruddell, B.L., Kumar, P., 2009. Ecohydrologic process networks: 1. Identification. Water Resour. Res. 45 (3). https://doi.org/10.1029/2008WR007279.

Salman, R., et al., 2021. A novel Bayesian maximum entropy-based approach for optimal design of water quality monitoring networks in rivers. J. Hydrol. 603, 126822. https://doi.org/10.1016/j.jhydrol.2021.126822.

Samuel, J., Coulibaly, P., Kollat, J., 2013. CRDEMO: combined regionalization and dual entropy-multiobjective optimization for hydrometric network design. Water Resour. Res. 49 (12), 8070–8089. https://doi.org/10.1002/2013WR014058.

Schilling, O.S., Cook, P.G., Brunner, P., 2019. Beyond classical observations in hydrogeology: the advantages of including exchange flux, temperature, tracer concentration, residence time, and soil moisture observations in groundwater model calibration. Rev. Geophys. 57 (1), 146–182. https://doi.org/10.1029/2018RG000619.

Scott, D.W., 1979. On optimal and data-based histograms. Biometrika 66 (3), 605–610.

Sen, P.N., Goode, P.A., 1992. Influence of temperature on electrical conductivity on shaly sands. Geophysics 57 (1), 89–96. https://doi.org/10.1190/1.1443191.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Siade, A.J., Cui, T., Karelse, R.N., Hampton, C., 2020. Reduced-dimensional gaussian process machine learning for groundwater allocation planning using swarm theory. Water Resour. Res. 56 (3). https://doi.org/10.1029/2019wr026061.

Sturges, H.A., 1926. The choice of a class interval. J. Am. Stat. Assoc. 21 (153), 65–66. https://doi.org/10.1080/01621459.1926.10502161.

Tran, A.P., Vanclooster, M., Zupanski, M., Lambot, S., 2014. Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter. Water Resour. Res. 50 (4), 3131–3146. https://doi.org/10.1002/2013wr014583.

Tsai, C.H., Rucker, D.F., Brooks, S.C., Ginn, T., Carroll, K.C., 2022. Transient storage model parameter optimization using the simulated annealing method. Water Resour. Res. 58 (7). https://doi.org/10.1029/2022WR032018 e2022WR032018.

Tso, C.-H.M., Iglesias, M., Binley, A., 2024. Ensemble Kalman inversion of induced polarization data. Geophys. J. Int. 236 (3), 1877–1900. https://doi.org/10.1093/gji/ggae012.

Tso, C.-H.M., et al., 2021. Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion. Geophys. J. Int. 225 (2), 887–905. https://doi.org/10.1093/gji/ggab013.

Tso, C.-H.M., et al., 2017. Improved characterisation and modelling of measurement errors in electrical resistivity tomography (ERT) surveys. J. Appl. Geophy. 146, 103–119. https://doi.org/10.1016/j.jappgeo.2017.09.009.

Tso, C.M., et al., 2020. Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection. J. Contam. Hydrol. 234, 103679. https://doi.org/10.1016/j.jconhyd.2020.103679.

Wang, N., Chang, H., Zhang, D., 2023. Inverse modeling for subsurface flow based on deep learning surrogates and active learning strategies. Water Resour. Res. 59 (7). https://doi.org/10.1029/2022WR033644 e2022WR033644.

Wang, W., et al., 2018. Optimization of rainfall networks using information entropy and temporal variability analysis. J. Hydrol. 559, 136–155. https://doi.org/10.1016/j.jhydrol.2018.02.010.

Xia, Y., et al., 2026. Combining hydrologic, chemical, and geophysical deep learning-based inversion for heterogeneous aquifer structure identification. J. Hydrol. 665, 134701. https://doi.org/10.1016/j.jhydrol.2025.134701.

Xiao, C., Leeuwenburgh, O., Lin, H.-X., Heemink, A., 2021. Conditioning of deep-learning surrogate models to image data with application to reservoir characterization. Knowl. Based Syst. 220. https://doi.org/10.1016/j.knosys.2021.106956.

Xu, L., et al., 2025. Fractal and NMR based characterizations of multi-scale pore structure alterations in tight sandstones due to scCO2-water–rock interactions. Chem. Eng. J. 508, 160898. https://doi.org/10.1016/j.cej.2025.160898.

Xu, T., Gómez-Hernández, J.J., 2016. Characterization of non-Gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures. Water Resour. Res. 52 (8), 6111–6136. https://doi.org/10.1002/2016wr019011.

Xu, T., Sonnenthal, E., Spycher, N., Pruess, K., 2006. TOUGHREACT—a simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: applications to geothermal injectivity and CO2 geological sequestration. Comput. Geosci. 32 (2), 145–165. https://doi.org/10.1016/j.cageo.2005.06.014.

Yang, X., Chen, X., Carrigan, C.R., Ramirez, A.L., 2014. Uncertainty quantification of CO2 saturation estimated from electrical resistance tomography data at the Cranfield site. Int. J. Greenh. Gas Control 27, 59–68. https://doi.org/10.1016/j.ijggc.2014.05.006.

Yeh, T.-C.J., Liu, S., 2000. Hydraulic tomography: development of a new aquifer test method. Water Resour. Res. 36 (8), 2095–2105. https://doi.org/10.1029/2000WR900114.

Yeh, T.C.J., Zhu, J., 2007. Hydraulic/partitioning tracer tomography for characterization of dense nonaqueous phase liquid source zones. Water Resour. Res. 43 (6). https://doi.org/10.1029/2006WR004877.

Zhan, C., et al., 2022a. An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network. J. Hydrol. 610. https://doi.org/10.1016/j.jhydrol.2022.127844.

Zhan, C., Dai, Z., Soltanian, M.R., de Barros, F.P.J., 2022b. Data-worth analysis for heterogeneous subsurface structure identification with a stochastic deep learning framework. Water Resour. Res. 58 (11). https://doi.org/10.1029/2022wr033241.

Zhang, J., Lin, G., Li, W., Wu, L., Zeng, L., 2018. An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions. Water Resour. Res. 54 (3), 1716–1733. https://doi.org/10.1002/2017WR020906.

Zhang, J., Zheng, Q., Wu, L., Zeng, L., 2020. Using deep learning to improve ensemble smoother: applications to subsurface characterization. Water Resour. Res. 56 (12). https://doi.org/10.1029/2020WR027399 e2020WR027399.

Zhang, T.-F., et al., 2019. Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. Pet. Sci. 16 (3), 541–549. https://doi.org/10.1007/s12182-019-0328-4.

Zhu, J., Yeh, T.-C.J., 2005. Characterization of aquifer heterogeneity using transient hydraulic tomography. Water Resour. Res. 41 (7). https://doi.org/10.1029/2004WR003790.

Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366, 415–447. https://doi.org/10.1016/j.jcp.2018.04.018.