A deep adaptive cycle generative adversarial neural network for inverse estimation of groundwater contaminated source and model parameter

Zidong Pan[1], Wenxi Lu[1], Yaning Xu[1], Chengming Luo[1], and Yukun Bai[1]

[1]Jilin University

August 3, 2023

Abstract

In light of the challenges posed by groundwater contamination and the urgent need for accurate and efficient groundwater contaminated source estimation (GCSE), the present study proposes a novel approach for GCSE using a deep adaptive cycle generative adversarial neural network (DA-CGAN). Given the equifinality from different parameters (EFDP) often associated with GCSE, we leveraged a bidirectional adversarial training pattern involving a forward process and a recovery process to supervise the inverse mapping relationship. Once trained, the forward process can be utilized to provide estimation for GSCE. This bidirectional-training strategy mitigates EFDP, thereby effectively enhancing the reliability of GCSE. Moreover, the performance of DA-CGAN is closely related to the quality of the training samples. To address this, we introduced a significant enhancement through an adaptive sampling strategy. This substantially improves the quality of training samples and consequently increases the accuracy of the GCSE. Furthermore, the inherent data-driven attribute of the deep cycle GAN considerably reduces computational costs when conducting GCSE. The research unfolds in the contexts of both hypothetical and real-world scenarios, with the goal of providing an efficient, precise, and cost-effective solution for GCSE. The results demonstrate that the DA-CGAN, an innovative model in the hydrogeological domain, exhibits superior performance in both estimation accuracy (Average Relative Error (ARE) of 4.91% and R of 0.998) and computational efficiency (0.17 seconds per run). This is particularly notable when compared with typical inverse methods such as the genetic algorithm (GA) and the ensemble kalman filter (ENKF).

Hosted file

`968993_0_art_file_11196065_ry2s73.docx` available at https://authorea.com/users/641765/articles/655882-a-deep-adaptive-cycle-generative-adversarial-neural-network-for-inverse-estimation-of-groundwater-contaminated-source-and-model-parameter

A deep adaptive cycle generative adversarial neural network for inverse estimation of groundwater contaminated source and model parameter

Zidong Pan[1,2,3] Wenxi Lu[1,2,3*] Yaning Xu[1,2,3] Chengming Luo[1,2,3] Yukun Bai[1,2,3]

1. *Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, China*

2. *Jilin Provincial Key Laboratory of Water Resources and Environment, Jilin University, Changchun 130021, China*

3. College of New Energy and Environment, Jilin University, Changchun 130021, China

Corresponding author: Wenxi Lu

Email: luwx999@163.com

Address: *College of New Energy and Environment, Jilin University, Changchun 130021, China.*

Abstract

In light of the challenges posed by groundwater contamination and the urgent need for accurate and efficient groundwater contaminated source estimation (GCSE), the present study proposes a novel approach for GCSE using a deep adaptive cycle generative adversarial neural network (DA-CGAN). Given the equifinality from different parameters (EFDP) often associated with GCSE, we leveraged a bidirectional adversarial training pattern involving a forward process and a recovery process to supervise the inverse mapping relationship. Once trained, the forward process can be utilized to provide estimation for GSCE. This bidirectional-training strategy mitigates EFDP, thereby effectively enhancing the reliability of GCSE. Moreover, the performance of DA-CGAN is closely related to the quality of the training samples. To address this, we introduced a significant enhancement through an adaptive sampling strategy. This substantially improves the quality of training samples and consequently increases the accuracy of the GCSE. Furthermore, the inherent data-driven attribute of the deep cycle GAN considerably reduces computational costs when conducting GCSE. The research unfolds in the contexts of both hypothetical and real-world scenarios, with the goal of providing an efficient, precise, and cost-effective solution for GCSE. The results demonstrate that the DA-CGAN, an innovative model in the hydrogeological domain, exhibits superior performance in both estimation accuracy (Average Relative Error (ARE) of 4.91% and R of 0.998) and computational efficiency (0.17 seconds per run). This is particularly notable when compared with typical inverse methods such as the genetic algorithm (GA) and the ensemble kalman filter (ENKF).

Key words: Inverse estimation; groundwater contamination; cycle generative neural network; adaptive sampling; deep learning; bidirectional adversarial training

Key points

- First attempt to employ a DA-CGAN as a direct framework, rather than as a surrogate model, for conducting GCSE
- The bidirectional adversarial design of the DA-CGAN to mitigate equifinality from different parameters, enhancing the accuracy of GCSE.
- The adaptive sampling strategy improves the quality of training samples fed to the DA-CGAN, further increasing the accuracy of GCSE.

1.Introduction

The issue of groundwater contamination has severe ramifications for both drinking water quality and the broader ecological environment (Yang et al., 2020; Zhang et al., 2022; Zhao et al., 2023). The clandestine nature of groundwater contamination, often discovered with significant delay, complicates the process of revealing the contamination source (Luo et al., 2022). Groundwater contaminated source estimation (GCSE) is a pivotal process in both assessing the risk posed by contamination and implementing remediation measures (Moghaddam et al., 2021). GCSE involves matching simulated outputs from a contaminant transport model with actual observations from monitoring wells (Zhou et al., 2014). Over recent decades, various methods have emerged to conduct GCSE, which can be summarized as three categories: simulation-optimization methods (Ayvaz, 2016; Yeh, 2015), simulation-statistics methods (Chang et al., 2021; Jiang et al., 2021; Zhou et al., 2018) and simulation- data assimilation methods (Chen et al., 2018; Jiang et al., 2018).

The simulation-optimization methods focus on establishing an optimization model, which aims to minimize the discrepancy between simulated outputs and observed data by adjusting decision variables such as contamination source information or model parameters (Xing et al., 2019; Zhao et al., 2020). Jiang et al. (2013) proposed an almost-parameter-free harmony search algorithm for groundwater pollution source identification and achieved a robust estimation under conditions of irregular geometry and erroneous monitoring data. Li et al. (2020) proposed a hybrid particle swarm optimization-extreme learning machine to estimate the contaminated source considering the uncertainty of random hydraulic parameters.

The simulation-statistics methods update the state of unknown variables (including contaminated source information or model parameters) to maximum the likelihood function which can evaluate the bias between the simulated outputs and observed data (Wang & Jin, 2013). Zanini and Woodbury (2016) proposed a Bayesian framework to reconstruct the release history of a contaminated source. Zhang et al. (2017) utilized a two-stage Monte Carlo method to evaluate the small failure probability analysis in groundwater contaminant modelling. An et al. (2022) utilized an improved Markov Chain Monte Carlo (MCMC) as a promising solution to characterize groundwater contaminated sources.

The simulation-data assimilation methods using the covariance matrix between the unknown variables and the observed data to update the estimated values of unknown variables (Kurtz et al., 2014; Li et al., 2018). Xu et al. (2021) used an ensemble smoother with multiple data assimilation to simultaneously estimate a contaminant source and hydraulic conductivity, presenting superior performance than the restart ensemble Kalman filter.

While these methods have proven to be effective, they necessitate multiple iterations of simulation models, resulting in considerable time consumption, particularly when multiple GCSEs are required. Furthermore, the accuracy of these approaches may face limitations when tackling highly nonlinear and intricate groundwater inverse problems, particularly in the establishment of the inverse mapping relationship. In light of these limitations, this paper introduces a novel approach that employs a deep cycle generative adversarial network (CGAN) to rapidly and accurately conduct GCSE.

Recently, deep learning methods, particularly generative neural network (GAN), have demonstrated remarkable capabilities in image recognition and translation tasks (Bond-Taylor et al., 2022; Yinka-Banjo & Ugot, 2020). GANs, a form of deep learning model, are known for their ability to generate data that mimic the input data (Goodfellow et al., 2014). They consist of two neural networks, a generator and a discriminator, that work in tandem to improve the generalization capacity for complex system. A myriad of studies demonstrated the potential of GANs in capturing complex geological input-output relationship. In the domain of hydrogeology, Laloy et al. (2018) used GANs for high-dimensional inverse modeling in hydrogeology. The researchers employed a Wasserstein GAN with a gradient penalty to generate plausible hydrogeological models that respected the observed data, which significantly improve the efficiency and reliability of the inversion process. Sun (2018) proposed a state-parameter identification GAN for estimating the spatial structure of the hydraulic conductivity and achieved satisfactory inverse results. Dagasan et al. (2020) applied a conditional GAN as a forward operator surrogate to characterize the spatial distribution of the hydraulic conductivity. Our previous work Pan et al. (2022) has explored the potential of deep convolutional-generative adversarial neural network for estimating high-dimensional hydraulic conductivity field. Zheng et al. (2023) utilized a GAN to generate the training samples for a convolutional neural network surrogate to efficiently provide estimation of groundwater contaminant source and hydraulic conductivity.

While numerous past studies have examined the utility of generative adversarial networks (GANs) for surrogate tasks within the hydrology field, the potent capacity of GANs to capture relationships also presents a promising opportunity for the direct implementation of the GCSE, rather than solely being deployed for surrogate purposes.

Theoretically, a GAN can realize GCSE via establishing a single-directional mapping relationship between simulated outputs (SO) and the groundwater contamination sources and parameters (GCSP). However, GCSE often exhibits ill-posedness, leading to a scenario where different combinations of GCSP can produce similar observations, a phenomenon known as equifinality from different parameters (EFDP) (Zhao et al., 2020). Given this circumstance, it becomes evident that a bidirectional mapping pattern is more suitable for conducting GCSE, compared to the single-directional mapping.

Therefore, in the present study, a variant of the traditional GAN, known as a cycle GAN (Zhu et al., 2017), was employed to conduct GCSE. This model incorporates two interconnected GANs working together, each consisting of a generator and a discriminator (Wang et al., 2022). These GANs work in a cyclical process where one GAN learns to translate from one data domain to another, and the other GAN learns to reverse this translation (Liang et al., 2022). This *cycle consistency* ensures that the data retains its original characteristics after translation and re-translation, making it an ideal tool for GCSE. To the best of our knowledge, no studies to date have implemented a cycle GAN as a direct framework, rather than as a surrogate model, for conducting GCSE.

In the context of GCSE, we used one GANs of the deep cycle GAN to translate the domain of SO derived from the transport model into the domain of GCSP—a process referred to as "forward mapping". The other GAN then reverts the translated GCSP domain back into its original SO domain—termed as "recovery mapping". The recovered SO domain closely resembles the simulated outputs from the transport model. The unique cycle adversarial training design of the deep cycle GAN can supervise the mapping from SO to GCSP, thereby mitigating EFDP. This provides an efficient and precise way to estimate groundwater contamination sources and parameters, offering a significant improvement over traditional GCSE methods.

However, the efficacy of deep learning methods also hinges on the quality of training samples (Sun et al., 2017; Van Horn et al., 2018). In light of this, an adaptive sampling strategy was implemented to enhance the quality of training samples for the cycle GAN. This strategy concentrates computational resources on areas of the GCSP space that yield more significant information, potentially obtaining more accurate results with a reduced number of total samples. In particular, we added one new sample at a time, utilizing all the accumulated information from updated training samples to determine more informative locations for generating subsequent samples. This adaptive-sampling strategy can effectively enhance the performance of the deep cycle GAN, thereby further improving the accuracy of GCSE. Furthermore, the inherent data-driven nature of deep cycle GAN results in a notably faster computation time compared to commonly used ensemble-based (GCSE) methods.

In the present study, we proposed a novel deep adaptive cycle GAN (DA-CGAN) for the estimation of contaminated groundwater sources using observed concentration data. Unlike conventional standard GANs typically employed for surrogate purposes, our proposed DA-CGAN employs a bi-directional training pattern and an adaptive sampling strategy. This innovative approach significantly improves both the accuracy and efficiency of GCSE. The performance of this method is evaluated in two scenarios: a hypothetical scenario and a real-world site scenario. The key contributions of the proposed method are as follows:

- First attempt to employ a DA-CGAN as a direct framework, rather than as a surrogate model, for conducting GCSE
- The bidirectional adversarial design of the DA-CGAN to mitigate equifinality from different parameters, thereby enhancing the accuracy of GCSE.
- The implementation of an adaptive sampling strategy improves the quality of training samples fed to the deep cycle GAN, further increasing the accuracy of GCSE.

- The inherent data-driven attribute of the deep cycle GAN considerably reduces computational costs when conducting GCSE.

2.Methodology

2.1 Numerical simulation model

The transportation of contaminant can be described by two sub models: a groundwater flow model and solute transport model. The governing equation of groundwater flow can be expressed as:

$$\frac{\partial}{\partial x_i}\left[K(H-z_0)\frac{\partial}{\partial x_i}\right]+\frac{\partial}{\partial x_j}\left[K(H-z_0)\frac{\partial}{\partial x_j}\right]+W(x,y,t)=0 \quad (1)$$

Where x_i and x_j denote the location distances along the respective Cartesian coordinate axis, K represents the hydraulic conductivity, W denotes the volumetric flux per unit volume, H represents the water level above the sea level. z_0 represents the elevation of the aquifer bottom above the sea level.

The governing equation of solute transport model can be expressed as:

$$\frac{\partial C}{\partial t}=\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x_j}\right)-\frac{\partial}{\partial x_i}(u_i C)+\frac{R}{\theta} \quad (2)$$

$$u_i=\frac{K_{ij}}{\theta}\frac{\partial H}{\partial x_i} \quad (3)$$

Where C represents the solute concentration, D_{ij} denotes the hydrodynamic dispersion tensor, u_i represents the average pore groundwater velocity that satisfies Darcy’s Law, θ denotes the effective porosity, R represents the source or sink term. For non-aqueous phase liquids (NAPLs) transportation, R can be expressed as:

$$R = R_{\text{source}}^{\text{NAPL}} - R_{\text{sink}}^{\text{Bio}} \quad (4)$$

Where $R_{\text{source}}^{\text{NAPL}}$ represents the rate of hydrocarbon from NAPL to aqueous phase, $R_{\text{sink}}^{\text{Bio}}$ represents the rate of hydrocarbon removal by biodegradation. The numerical simulation models of two scenarios were calculated using MODFLOW and MT3D/SEAM3D module of groundwater modeling system.

2.2 Generative adversarial neural network

Generative Adversarial Networks (GANs) constitute a subcategory of artificial intelligence algorithms designed to discern data distributions via an adversarial interaction between two unique neural networks: the generator and the discriminator. The generator strives to formulate data instances indistinguishable from authentic data, whereas the discriminator's role involves distinguishing real data instances from those manufactured by the generator. Both constituents are usually realized as various forms of neural networks, including but not limited to fully connected and convolutional neural networks.

The generator G utilizes a prior random noise variable, $p_z(\boldsymbol{z})$, to convert it into a data distribution, $\boldsymbol{m}$. The notation $G(\boldsymbol{z};\, \theta_g)$ signifies a generative/mapping operator to the data space of $\boldsymbol{m}$, where θ_g are the parameters of a neural network. In contrast, the discriminator D serves the function $D(\boldsymbol{m};\ \theta_d)$, signifying the probability of the generated samples, m, originating from real samples. θ_d are the parameters of the other neural network. The fundamental goal of a GAN is to simultaneously minimize the generator loss $log(1 - D(G(\boldsymbol{z})))$ and maximize the discriminator loss $log(D(\boldsymbol{m}))$. This objective can be represented as a two-player minmax game, formulated with the following value function as described by Goodfellow et al. (2014):

$$\min_G \max_D V(D,G) = \mathbf{E}_{\boldsymbol{x} \sim p_{data(x)}}[\log D(\boldsymbol{m})] + \mathbf{E}_{z \sim p_{z(z)}}[1 - \log D(G(\boldsymbol{z}))] \quad (5)$$

In other words, the θ_g of the generator and the θ_d of the discriminator must be alternately trained with the same objective function until the adversarial process between them reaches Nash equilibrium, which means the generator G can generate the perfect imitation of m that the discriminator D cannot distinguish. For GCSE, the m_o^r represents the real samples of observation data domain, m_t^g represents the fake (generated) samples of GCSP data domain derived by the generator and m_r^t represents the real samples of GCSP domain (Fig.1).

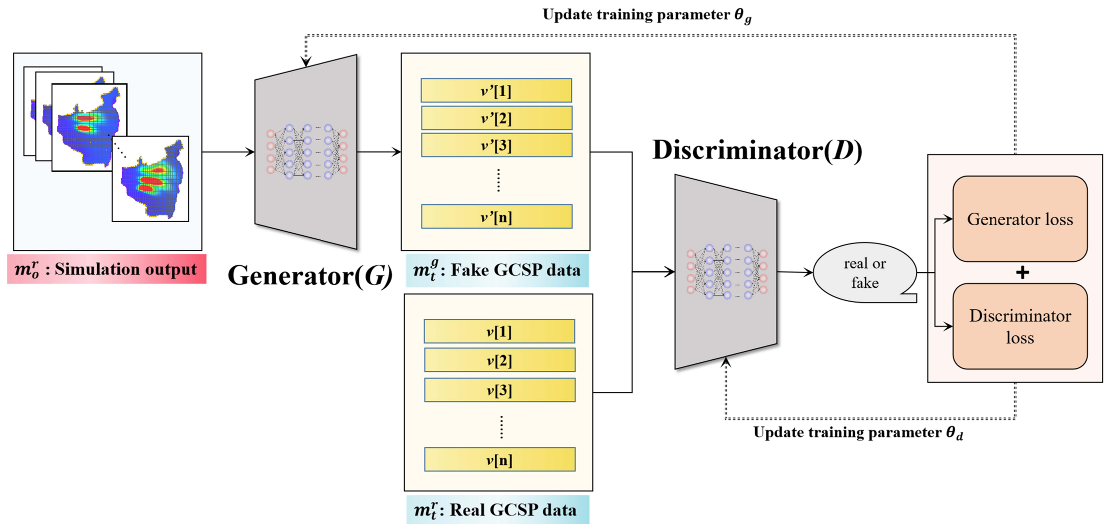


Fig.1 The basic topological structure of traditional GAN for GCSE

2.3 Deep adaptive cycle generative adversarial network

Cycle generative adversarial network

However, the standard GANs might suffer from the equifinality from different parameters (EFDP). In other words, the generator network starts producing similar samples of observation data, despite being given different inputs of GCSP. Thus, a cycle GAN with a bi-directional mapping strategy was proposed to mitigate EFDP. This method involves training two interconnected GANs in a cyclic manner. Each GAN consists of a generator and a discriminator, with one GAN (consists of G_p and D_p) translating from SO domain (O) to GCSP domain (P), and the second GAN ((consists of G_o and D_0)) reversing this process (Fig.2). In particular, the GCSP-SO transformation loss used in the cycle GAN encourages the generators to create estimated results of GCSP from SO and then recover the GCSP back to SO, which can be expressed as:

$$L_{trans}(G_p, G_o) = \mathrm{E}_{P \sim p_{data}(P)}\left[\left\| G_p(G_o(P)) - P \right\|_1\right] + \mathrm{E}_{O \sim p_{data}(O)}\left[\left\| G_o(G_p(O)) - O \right\|_1\right] \quad (6)$$

The total training loss consists of three components, namely, adversarial loss of two GANs and the transformation loss of the generated O and P, which can be expressed as:

$$L(G_o, G_p, D_o, D_p) = L_{\mathrm{GAN}}(G_o, D_o, O, P) + L_{\mathrm{GAN}}(G_p, D_p, P, O) + \lambda L_{trans}(G_o, G_p) \quad (7)$$

Where λ represents the relative importance of the GCSP-SO transformation loss towards the adversarial loss L_{GAN}, which is set to “0.5” in the present study. The G_p and G_o aim to minimize the total training loss whereas the D_p and D_0 aim to maximize the loss. Once trained, the G_p can be utilized to estimate the GCSP from the given observation data.

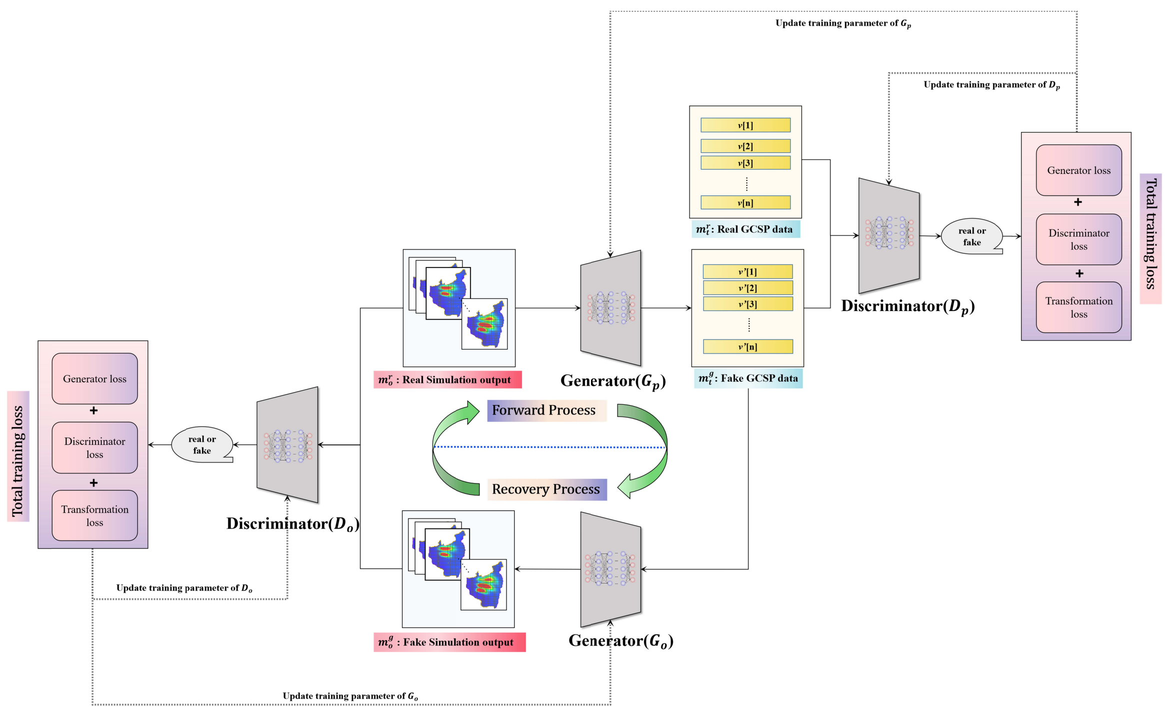


Fig.2 The basic topological structure of cycle GAN for GCSE

Adaptive sampling-generated strategy

Deep learning algorithms, such as a cycle GAN (CGAN), irrespective of the input, will invariably provide an output. Nonetheless, to elicit high-quality estimation results, the algorithm must be trained with superior quality data samples (Xiao et al., 2018). Moreover, adaptive sampling allows for more focused and efficient use of computational resources by prioritizing data points that provide the most information or learning potential (Li et al., 2021; Liu et al., 2018). Therefore, an adaptive sampling-generated strategy was proposed to provide high-quality samples for the CGAN. In particular, at the initial step, we used the pre-trained CGAN to estimate GCSP from the observation data and get an estimated result, thereby obtaining an initial estimate. Subsequently, by executing a forward run of the numerical simulation model, an updated sample was adaptively generated and incorporated into the pre-existing dataset. The CGAN was then retrained using this augmented dataset. The iterative process continued until the bias (B) between the current estimation result and that of the previous step reached a tolerance value (δ).

Table 1 Flow of adaptive sampling-generated strategy

1 INITIALIZATION STEP

1.1 Set the tolerance δ and max iteration, define numbers of unknown variable: n_v, observation: n_o and number of training samples: n_s, lower boundary of **GCSP**: lb and upper boundary of **GCSP**: ub

$$n_{tr} = n_v + n_o$$

1.2 Generate the initial training dataset of **GCSP** v ($n_s \times n_v$ matrix) and the corresponding **SO** o ($n_o \times n_v$ matrix), "*lhs*"means Latin hypercube sampling.

$$v_{initial} = \mathrm{lb} + lhs(n_s, n_v) \cdot (\mathrm{ub} - \mathrm{lb})$$

The total training dataset Tr consists of v and o.

2 ITERATIVE LOOP

While $B >$ δ and iteration< max iteration **do**

2.1 Training the cycle GAN with the initial dataset Tr.

2.2 Execute **GCSE**, obtain an estimation result of **GCSP**.

2.3 Adaptive Sample generated: forward run the simulation model with the estimation result, update the prior dataset.

2.4 Retrain cycle GAN with the updated dataset.

***loop* = *loop* + 1**

3 TERMINATION Obtain the optimal estimated results of GCSP.

3.Application

3.1 case overview

In this section, the effectiveness and applicability of the proposed deep adaptive cycle GAN (DA-CGAN) for GCSE were assessed using two scenarios: a hypothetical scenario and a real-world scenario. The hypothetical scenario provides reference values, which enable the comparison of estimated results and actual values, specifically in terms of the unknown variables (GCSP). The observational data at the monitoring wells were produced by conducting a forward run of the simulation model with the reference values of GCSP. Meanwhile, for the real-world scenario, the actual observational data serve as the sole criterion for evaluating the proposed DA-CGAN.

3.1.1 Hypothetical scenario

The site of the hypothetical scenario encompasses an unconfined aquifer, with groundwater flow directed from west to east ($2000\mathrm{m} \times 2500\mathrm{m}$). In term of the groundwater flow boundary, the west and east boundaries are specific head boundaries while the north and south boundaries are no-flow boundaries (Fig.3). In terms of solute boundaries, only the west boundary holds a specific concentration, with other boundaries manifesting no-flow. The hydraulic conductivity can be divided into four zones: $k(\mathrm{I})$, $k(\mathrm{II}), k(\mathrm{III}), k(\mathrm{IV})$ Table 2 provides detailed information regarding the aquifer. Three potential contamination sources, situated on the west side of the site, release contaminants into the aquifer. The release histories of these sources are conceptualized into five stress periods: T_1, T_2, T_3, T_4, T_5.

The estimation involves three types of unknown variables: features of the contamination source, boundary conditions, and hydraulic parameters. Specifically, contamination source features include the release intensities of the three potential sources during five stress periods, labeled as, $S_i T_j, i = 1,2,3, j =$1,2,3,4,5 (15 dimensions). The boundary conditions incorporate the contaminant recharge flux on the west boundary, denoted as C_b. The hydraulic parameters involve the hydraulic conductivities in four zones (4 dimensions) and the longitudinal dispersivity D_l (1 dimension). In total, the hypothetical scenario targets the estimation of 21-dimensional unknown variables. For the calculation of numerical simulation model, the domain has been discretized by the grids with the size of $20m \times 20m$.

Table 2 Prior values/ranges of the aquifer and the contaminated source (hypothetical scenario)

Parameter	Value/Range
Hydraulic conductivity(m/d)	(30,50)
Contaminant recharge flux C_b (mg/l)	(30,90)
Specific yield	0.24
Longitudinal dispersivity D_l (m)	(20,60)
Ratio of transverse dispersivity to longitudinal dispersivity	0.1
Saturated thickness(m)	40
Grid spacing in x-direction(m)	20
Grid spacing in y-direction(m)	20
Stress periods(year)	5
Fluxes of pollution source during stress period (g/d)	(0,52)

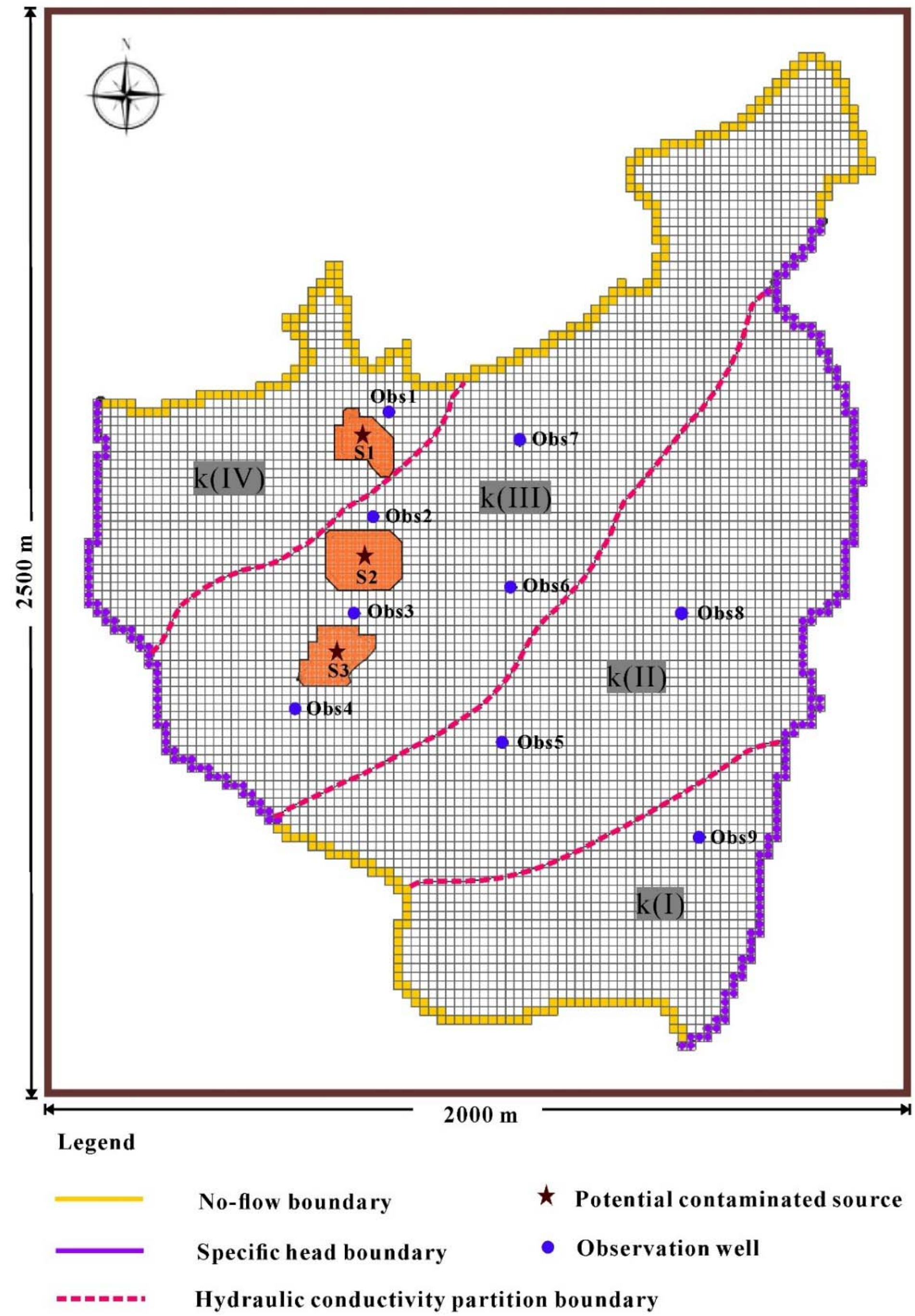


Fig.3 Boundary conditions, hydrogeology conditions and potential contaminated sources

3.1.2 Real world scenario

The contaminated site is a chemical plant located in Jilin Province, China, with a width of 560 m and length of 620 m. According to field investigation, the plant released Benzene into the aquifer and ten monitoring wells were set to trace the contaminant. The chemical reaction and biodegradation reaction were considered. According to the observed groundwater head, Γ_1 is generalized as a specific head boundary. Γ_3 is the Songhua River, which is conceptualized as the specific boundary. Γ_1 and Γ_1 are parallel to the groundwater flow line, thus are generalized as no-flow boundary. Table 3 provides detailed prior information regarding the aquifer and the contaminated source. We set ten wells to monitoring the solute transport of the groundwater. In particular, $\#1, \#2, \#3, \#4, \#5, \#6, \#7$ were allocated for observing both the water level and contaminant concentration, while wells $\#8, \#9, \#10$ were reserved exclusively for water level monitoring.

It must be noted that, some model parameters were selected as unknown variables through sensitivity analysis. The estimation involves three types of unknown variables: spatial-temporal features of the contaminated source, hydraulic parameter and reaction parameter. In particular, the spatial-temporal features of contaminated source include the position (x, y), the initial release concentration (C_r) and dissolved rate (D_r). The hydraulic parameter involves the hydraulic conductivity (K_c), the porosity (P), longitudinal dispersivity (L_d) and the ratio of horizontal transverse dispersivity to longitudinal dispersivity (α). The reaction parameter includes the initial concentration of dissolved oxygen (D_o). In total, the real-world scenario targets the estimation of 9-dimensional unknown variables. For the calculation of numerical simulation model, the domain has been discretized by the grids with the size of $5\mathrm{m} \times 5\mathrm{m}$.

Table 3 Prior values/ranges of the aquifer and the contaminated source (real world scenario)

Parameter	Value/Range
Position x (m)	(20,200)
Position y (m)	(0,140)
Initial release concentration C_r (*10E-3 mg/l)	(0.8,1.2)
Dissolve rate (1/d)	(0.5,0.8)
Hydraulic conductivity(m/d)	(40,60)
Porosity P	(0.2,0.3)
Longitudinal dispersivity L_d (m)	(20,60)
Ratio of transverse dispersivity to longitudinal dispersivity α	(0.3,0.5)
Initial concentration of dissolved oxygen D_o (mg/l)	(1.4,3)
Initial concentration of Fe(II) (mg/l)	0.003
Microcolony minimum (mg/m^3)	0.0055
Grid spacing in x-direction(m)	5
Grid spacing in y-direction(m)	5

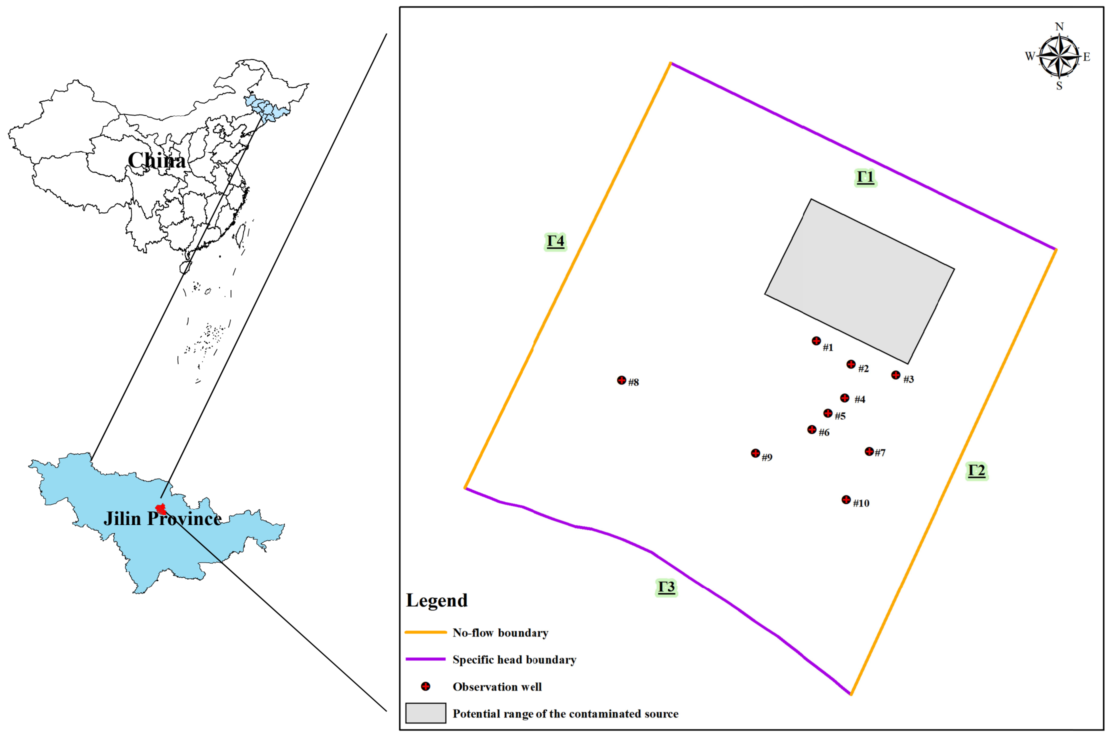


Fig.4 Overview of boundary conditions, potential contaminated sources and observation wells

Training dataset

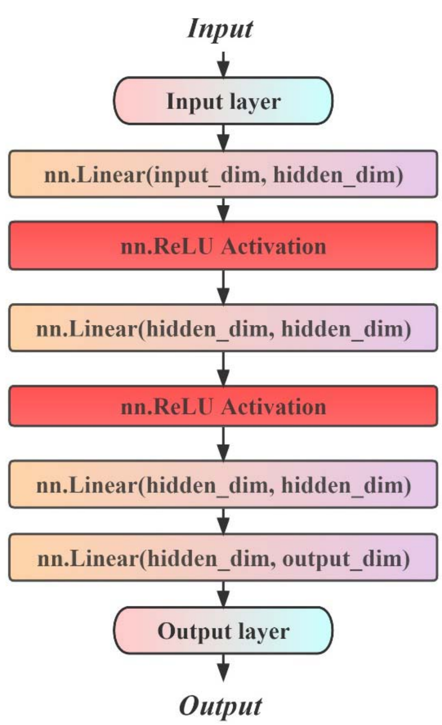
For DA-CGAN, the initial training dataset is derived by forward-running the simulation model using the provided samples of GCSP. These samples can be generated within their upper and lower boundaries using the Latin hypercube sampling method, as outlined in Table 1. In the present study, the quantity of training samples for the two scenarios is 500 and 400, respectively. In particular, 80% of these samples are utilized for training purposes, while the remaining 20% serve as the validation dataset. As the estimation loop initiates (Table 1), the adaptively generated samples are sequentially fed into the DA-CGAN to train until the termination of the loop.

Training details

The DA-CGAN has been trained on a PC with Intel Core i7-12700H CPU i7-12700H processor, GTX3060 GPU, and 16.0 GB RAM. The important part of DA-CGAN is the design of the two GANs which involve their own generator and discriminator. For the purpose of generated data of numerical form, the generators of two GANs were designed as a designed-friendly fully-connected structure. Figure 5 presents the topological structure of generator, where input_dim represents the dimensions of input of generator and output_dim represents the dimensions of output of generator, hidden_dim represents the dimensions of neurons in the hidden layers.

It must be noted that, the generator G_p and generator G_o possess identical structures, with hidden_dim of values of “100”. Despite their analogous structures, differences exist in their weights and biases. This discrepancy arises from the unique mapping relationships established by each generator: G_p (SO → GCSP) versus G_0 (GCSP → SO). The experiment was conducted in a Python environment, leveraging the torch package to construct the network structure.

Functions such as “nn.Linear()” and “nn.ReLU” were invoked from the torch package (Paszke et al., 2019). With regard to the optimization hyperparameters, “Adam” was selected as the optimizer, the “initial learning rate” is set to “0.002”. This rate linearly declines from 0.002 to zero over the course of the final 500 epochs. The “batch size” equals the size of the training dataset, thus accelerating the training process. Additional details regarding the optimization hyperparameters can be located in our attached source code.



.

Fig.5 The topological structure of generator G_p and G_o

4. Result and discussion

This section assesses the performance of DA-CGAN in terms of estimation accuracy and computational time cost. For a comprehensive comparison, we contrast DA-CGAN with three typical indirect methods: the Genetic Algorithm, Markov Chain Monte Carlo (MCMC), and the Ensemble Kalman Filter. It must be noted that, the typical indirect methods required massive realizations of numerical simulation model, which is high-time cost. The model generalization ability/ estimation accuracy (can be evaluated through the criterion of the correlation coefficient (R) and the average relative error (ARE), which can be expressed as:

$$R = 1 - \sum_{i=1}^{n} (v_{tru}(i) - v_{est}(i))^2 / (v_{tru}(i) - m_{tru})^2 \quad (8)$$

$$\text{ARE} = 100\% \times \frac{1}{n} \sum_{i=1}^{n} (v_{tru}(i) - v_{est}(i)) / v_{tru}(i) \quad (9)$$

Where v_{tru} represents the true values of GCSP, v_{est} represents the estimation values of GCSP, m_{tru} represents the mean values of v_{tru}. It must be noted that, the generalization ability was assessed using the validation dataset, while the estimation accuracy was assessed based on the reference values of GCSP. Theoretically, the reference values of GCSP can be a subset of the validation dataset.

4.1 Hypothetical scenario

Fig.6 shows the trace plot of training loss of $L_{GAN}(G_p, D_p, P, O)$ (fig.6(a)) $L_{GAN}(G_o, D_o, O, P)$ (fig.6(b)) and $L_{trans}(G_O, G_P)$ (fig.6(c)) as mentioned before in equation (7). It can be indicated that the CGAN reached a stable training process till 10,000 th epoch. Figure 6(d) shows that the implementation of the CGAN model resulted in accurate and stable estimations of GCSP with the ARE 4.9% and R of 0.9856 when compared with the validation data. Fig.6(e) illustrates the loss of discriminators D_o and D_p, which reveals that the discriminators also improved the ability to distinguish real data and generated data. In general, both the generators' and discriminators' capabilities have been enhanced through the adversarial training process.

Moreover, an adaptive-sampling strategy was implemented to enhance the accuracy of the CGAN. Figure 7 shows that the ARE was improved from 8.86% to 4.91% whereas the R was improved from 0.948 to 0.998. It was evident that the estimation accuracy of DA-CGAN for GCSE increased as new training samples were adaptively generated and used to retrain the DA-CGAN (fig.7). Table 4 presents the comparison of estimated values and reference values of GCSP. The AREs of the GCSP were all found to be below 10%, reaching an average value of 4.91%. In terms of SO, figure 9 presents the comparison between the observed and simulated contaminant concentrations corresponding to the estimated

GCSP. DC-CGAN achieved a mean ARE of 4.62% between the observed and simulated outputs at the monitoring wells. It further substantiates that the bi-directional strategy ensures the accuracy of GCSP and the corresponding SO. This suggests that the proposed DA-CGAN achieved promising accuracy in GCSE.

Furthermore, the performance of DC-CGAN was compared with traditional methods such as genetic algorithm (GA) and ensemble Kalman filter (ENKF), where DA-CGAN outperformed these techniques in terms of estimation accuracy (ARE) and calculated time cost (fig.8). With regard to the estimation accuracy, the notable performance of the DA-CGAN (ARE of 4.9%) can be primarily attributed to three techniques: the unique bi-directional design (BD), the deep generative-adversarial learning structure (DGAL), and an adaptive-sampling strategy (AS), respectively. In particular, BD and DGAL enhance the learning capacity of DA-CGAN, while AS ameliorates the quality of the training samples used for DA-CGAN. In terms of computational time, the data-driven nature of the DA-CGAN enables it to execute GCSE rapidly in 0.17 seconds, which is markedly faster than both the ENKF at 10.62 seconds, and the GA at 40.30 seconds.

It should be emphasized that the Genetic Algorithm (GA) and the Ensemble Kalman Filter (ENKF) both incorporate a surrogate. This surrogate role can be served by the recovery process embedded within our

DA-CGAN. In other words, this recovery process (surrogate) establishes a mapping relationship from GSCP to SO. This demonstrates that the DA-CGAN can serve not only as an inverse estimation framework but also as a surrogate model.

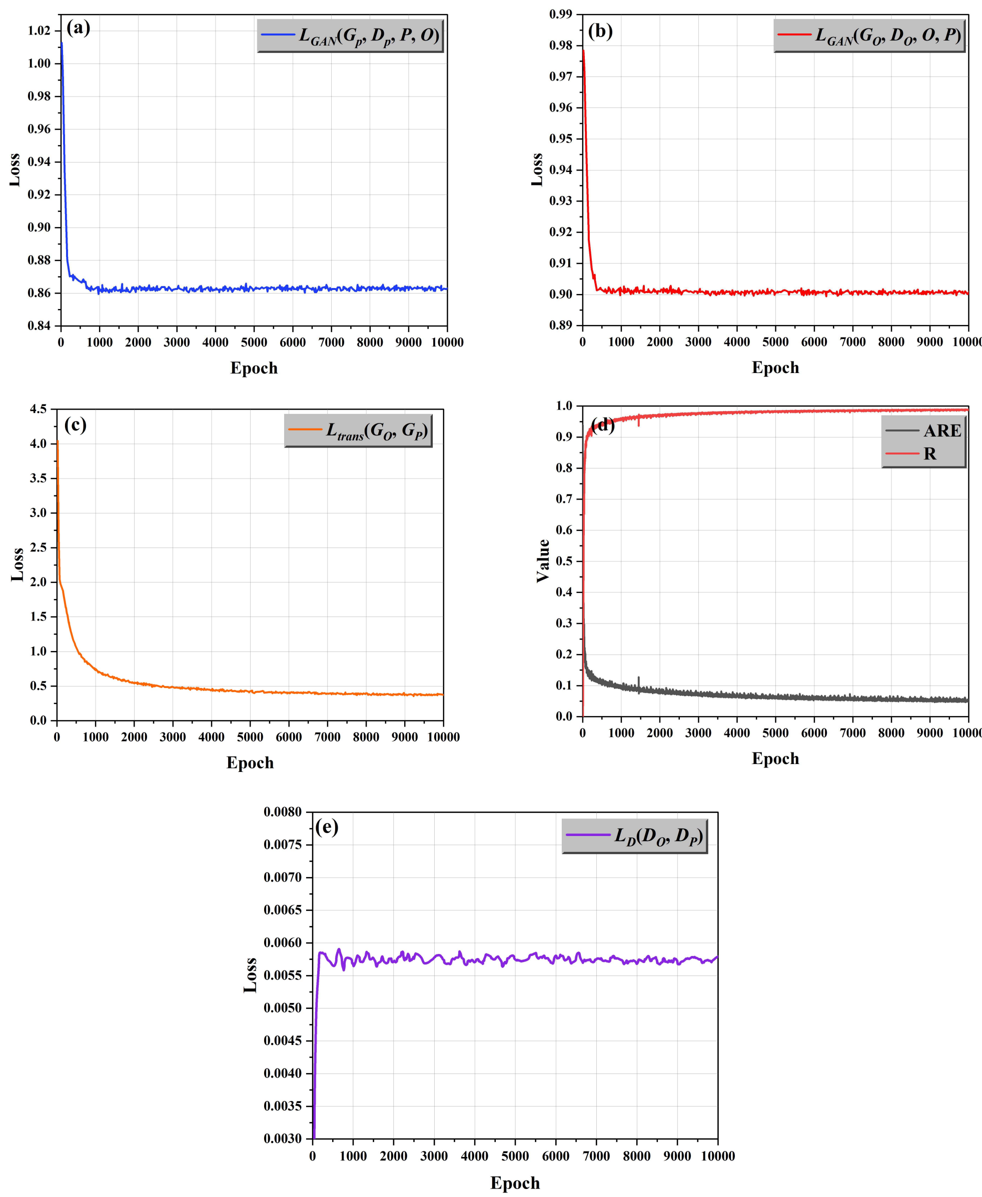


Fig.6 Trace of training loss and ARE and R of the CGAN (Hypothetical scenario)

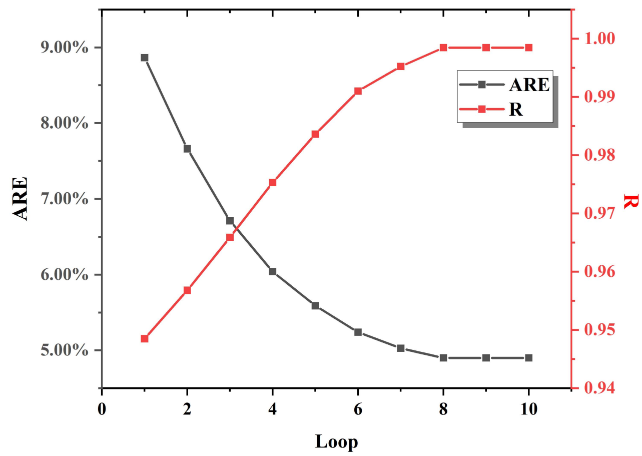


Fig.7 Trace plot of ARE and R of DC-CGAN using adaptive-sampling strategy

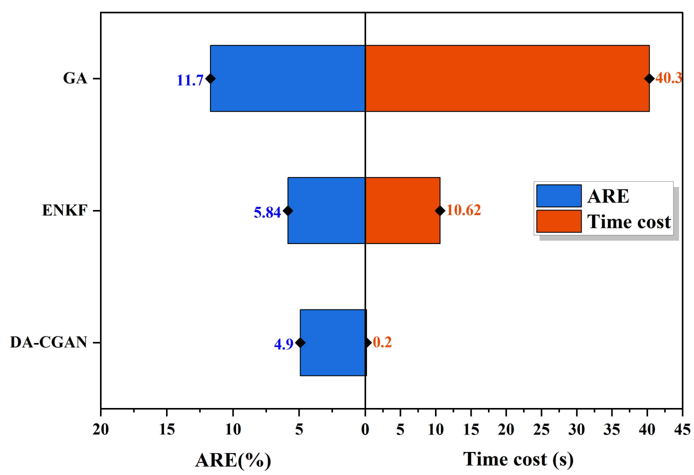
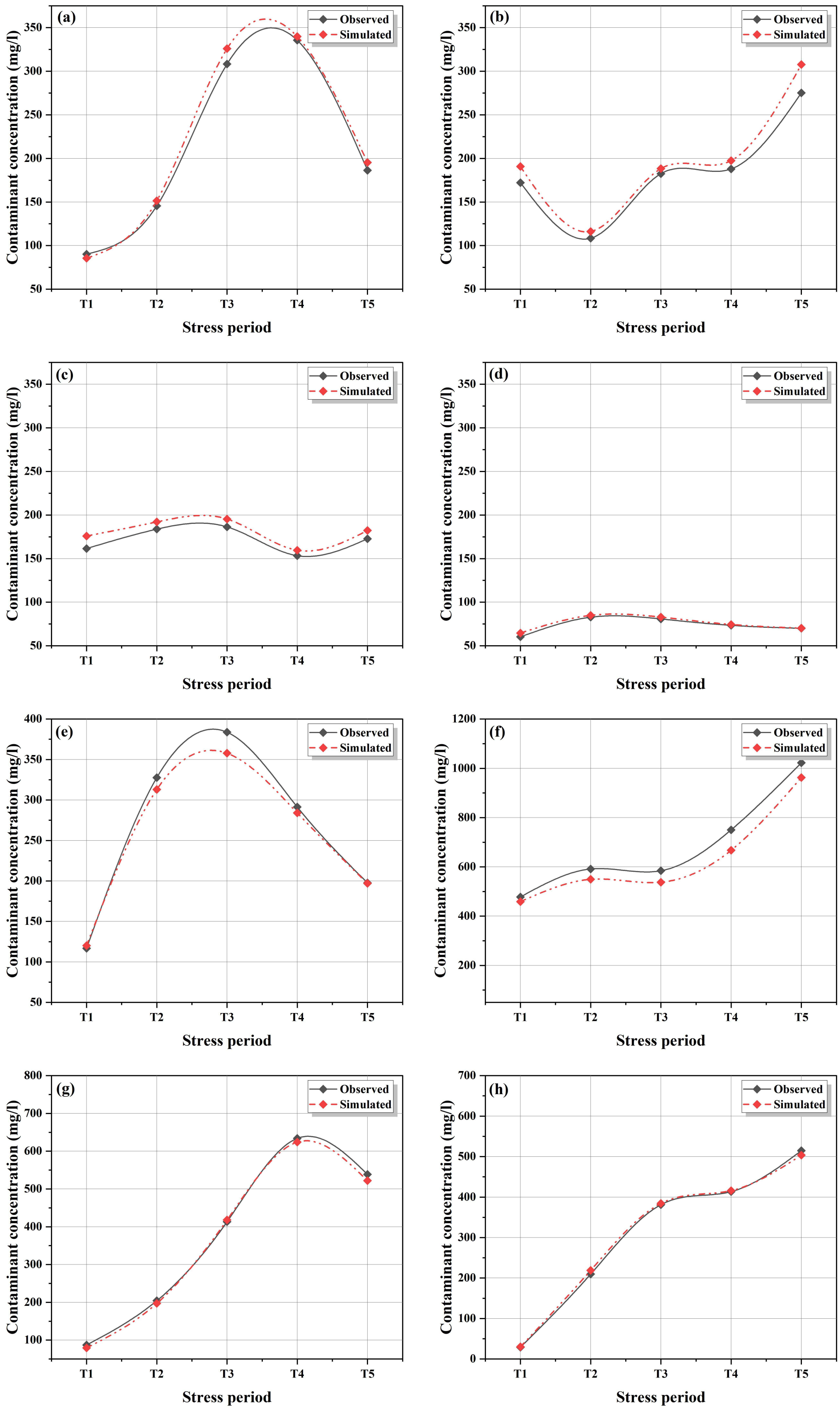


Fig.8 Comparison of performance of GA, ENKF and DA-CGAN

Table 4 Comparison of estimated values and reference values of GCSP (hypothetical scenario)

GCSP	Reference values	Estimated values	ARE(%)
C_b	63.36	62.20	1.83
S_1T_1	14.14	12.86	9.07
S_1T_2	16.78	17.87	6.46
S_1T_3	47.70	49.10	2.94
S_1T_4	42.72	38.68	9.45
S_1T_5	11.07	9.98	9.88
S_2T_1	39.40	39.62	0.56
S_2T_2	5.05	5.20	2.99
S_2T_3	28.99	26.54	8.45
S_2T_4	23.44	22.26	5.04
S_2T_5	47.89	49.41	3.17
S_3T_1	32.57	33.69	3.41
S_3T_2	36.95	34.71	6.05
S_3T_3	21.93	23.56	7.44
S_3T_4	9.85	10.62	7.90
S_3T_5	8.08	8.33	3.12
$k(I)$	34.98	31.90	8.81
$k(II)$	43.68	46.20	1.19
$k(III)$	45.08	42.42	0.75
$k(IV)$	48.82	53.21	0.80
D_l	55.12	60.12	3.63


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Observed
Simulated
Contaminant concentration (mg/l)
Stress period
T1
T2
T3
T4
T5

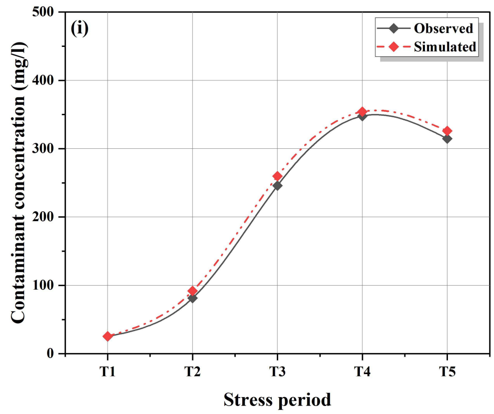


Fig.9 The comparison between observed and simulated contaminant concentration corresponding to the estimated GCSP

4.2 Real world scenario

The effectiveness of the DA-CGAN was evaluated in the previous section using a hypothetical scenario. In this section, we applied the DA-CGAN to perform GCSE in a real-world scenario. Fig.10 shows the trace plot of training loss and ARE and R of the CGAN in a real-world scenario.

At the start of training, this loss of $L_{GAN}(G_p, D_p, P, O)$ (fig.10(a)) and $L_{GAN}(G_o, D_o, O, P)$ (fig.10(b)) were high, given that the generator initially produces data easily distinguishable from real data. As training progresses, the generator loss decreased, implying that the generators of G_p and G_o were improving their ability to produce data closely resembling the real data. That is to say, the generators of G_p and G_o can provide more accurate and stable estimation results of GCSP and SO, respectively. The The decreasing $L_{trans}(G_O, G_P)$ (fig.10(c)) further proved that the accuracy of bi-transformation from SO to GCSP is valid. Fig.10(e) illustrates the loss of discriminators D_o and D_p. It is the measure of how well the discriminator is able to correctly classify real and generated data. A higher loss signifies a better ability of the discriminator to correctly differentiate between real and generated data. An upward trend in the loss can be observed, indicating that the discriminators' capacity to distinguish training samples has consistently improved throughout the adversarial training process. After training, the

G_p can be utilized to perform the GCSE by transforming SO into GCSP. Fig.10(d) shows that after 20,000 epochs, the CGAN achieved a stable and reliable estimations of GCSP with ARE of 7.5% and R of 0.95.

It must be noted that, when dealing with a real-world scenario, it is essential to compare the corresponding SO of the estimated GCSP with the real observation data. Figure 11 illustrates the trace plot of SO ARE of DC-CGAN using adaptive-sampling strategy. It demonstrates a distinct decrease the ARE of the DA-CGAN, stabilizing at 23.06%, as adaptive samples were sequentially generated and incorporated into the network. This result validates the effectiveness of the adaptive-sampling strategy. Figure 12 visualized the comparison of the simulated and observed contaminated concentrations. Table 5 presents the estimated values of unknown GCSP in a real-world scenario. The visualization of Estimated position of the contaminated source can be found in fig.13.

It must be noted that a SO ARE of 23.06% in a real-world scenario is notably higher than that of 4.62% in the hypothetical scenario. But the mean GCSP ARE of 7.5% (validation dataset) presents not much difference form that of 8.86% (validation dataset) in the hypothetical scenario. This discrepancy of SO may be attributable to the noise present in the measurement of contaminant concentrations at monitoring wells. Consequently, exploring denoising techniques would be a potential direction for future research.

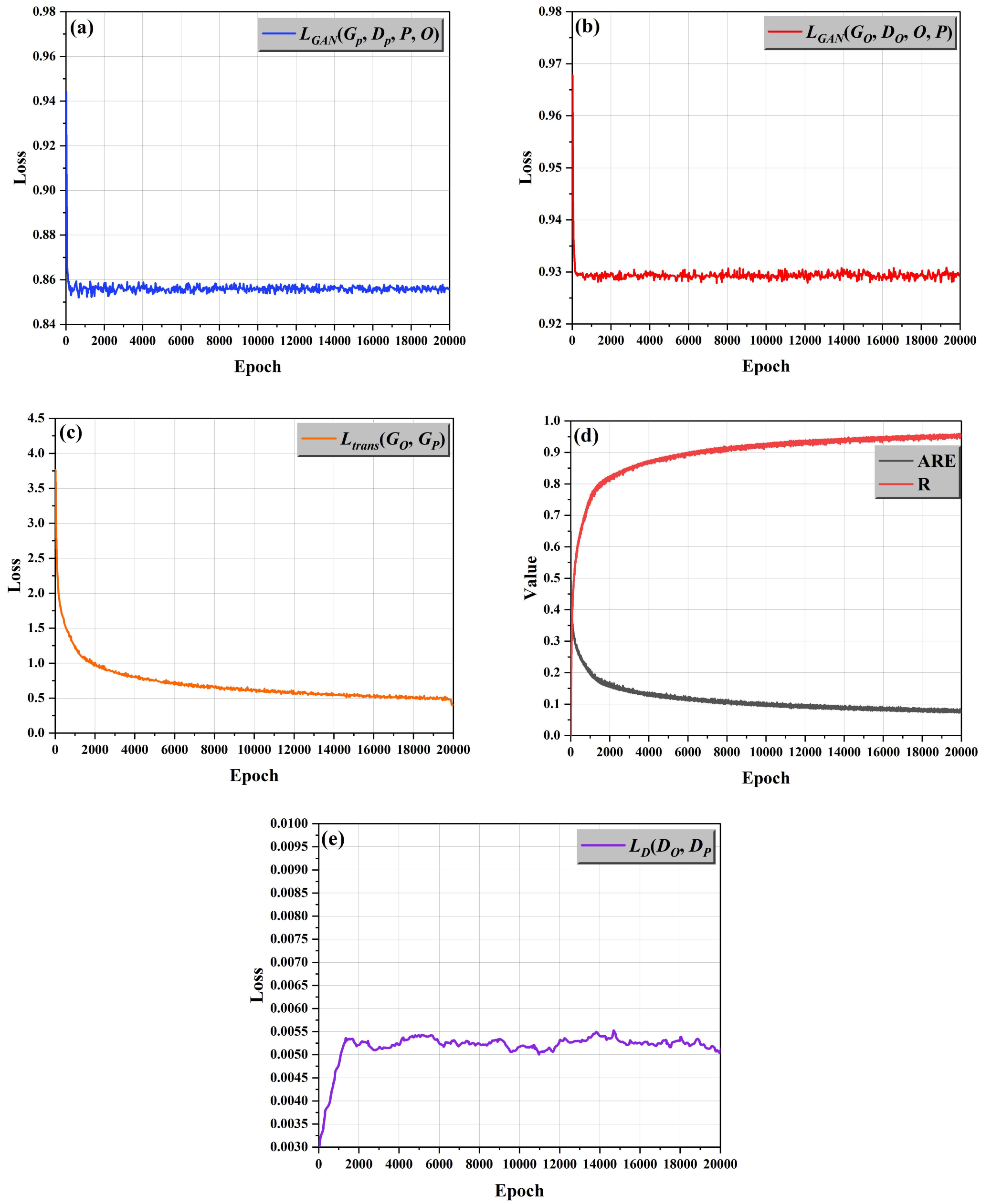


Fig.10 Trace of training loss and ARE and R of the CGAN (real world scenario)

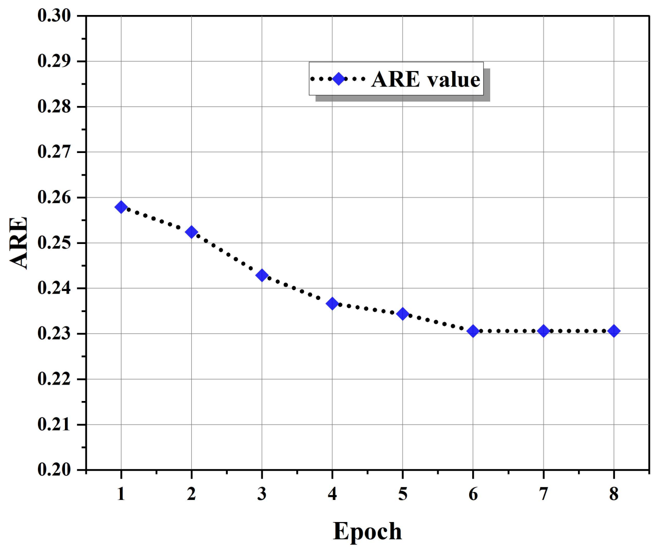


Fig.11 Trace plot of SO ARE of DC-CGAN using adaptive-sampling strategy

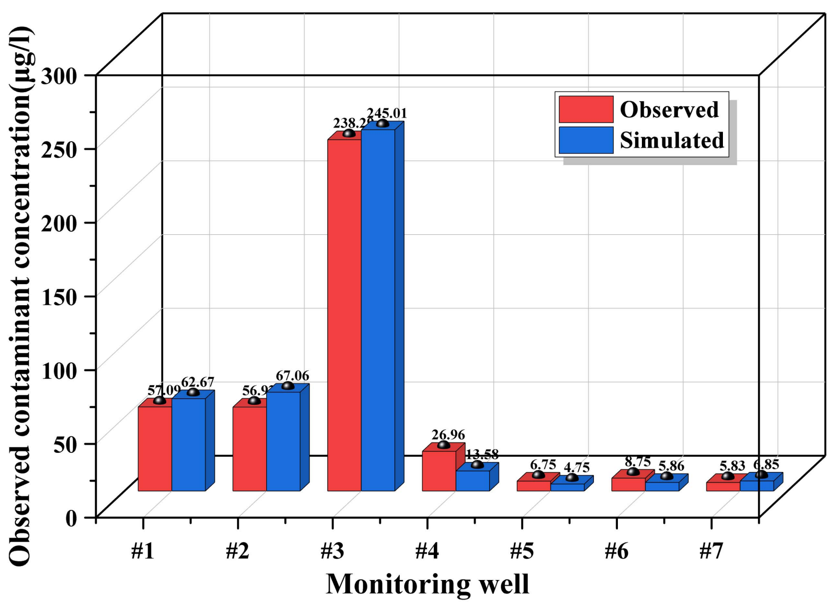


Fig.12 Comparison of the simulated and observed contaminated concentrations

Table 5 Estimated values of unknown GCSP (real world scenario)

Unknown GCSP	Prior Range	Estimated value
Position x (m)	(20,200)	183.637
Position y (m)	(0,140)	105.584
Initial release concentration C_r (*10E-3 mg/l)	(0.8,1.2)	1.0
Dissolve rate (1/d)	(0.5,0.8)	0.528
Hydraulic conductivity(m/d)	(40,60)	47.32
Porosity P	(0.2,0.3)	0.244
Longitudinal dispersivity L_d (m)	(20,60)	28.898
Ratio of transverse dispersivity to longitudinal dispersivity α	(0.3,0.5)	0.341
Initial concentration of dissolved oxygen D_o (mg/l)	(1.4,3)	2.401

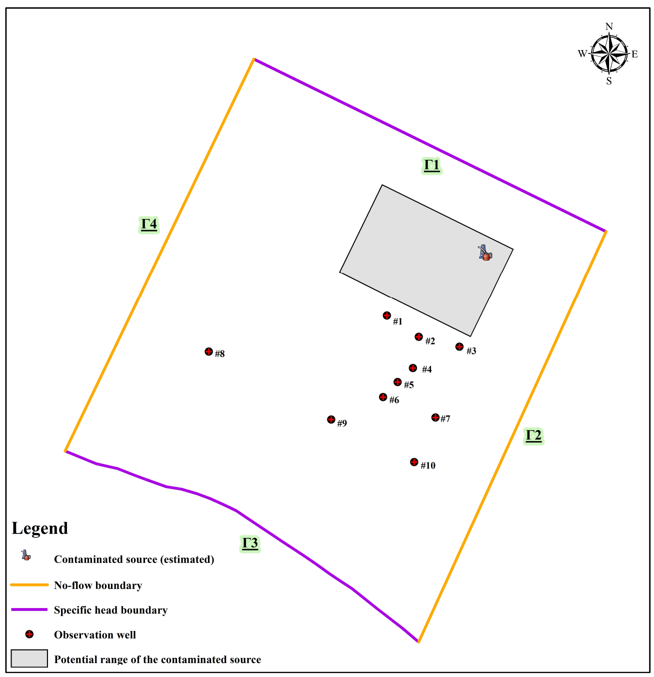


Fig.13 Estimated position of the contaminated source

5. Conclusion

In the present study, we proposed a deep adaptive cycle generative adversarial network (DA-CGAN) for the task of groundwater contaminated source estimation (GCSE). The efficiency and effectiveness of this DA-CGAN were assessed in both hypothetical and real-world scenarios. The following conclusions have been drawn from this study:

1. The proposed DA-CGAN proved to be a powerful tool for GCSE. This model, built on deep learning and adversarial training concepts, have provided reliable estimations of various parameters, such as boundary conditions, hydraulic conductivities, and release intensity and position of contaminated source across diverse GCSE scenarios.
2. The bidirectional design, deep generative-adversarial learning structure, and adaptive-sampling strategy employed in DA-CGAN were integral to its performance. In particular, the unique bidirectional design supervised the mapping from SO to GCSP, mitigating the phenomenon of EFDP. Moreover, the deep learning structure enhanced the capacity of DA-CGAN to learn complex mapping relationships from SO to GCSP. Furthermore, the adaptive-sampling strategy improved the quality of training samples, leading to better estimation accuracy for GCSE.
3. Comparisons with traditional methods such as the Genetic Algorithm (GA) and the Ensemble Kalman Filter (ENKF) showed that

DA-CGAN outperformed these methods in both estimation accuracy and computational efficiency. This superiority in performance underscores the potential of DA-CGAN as a robust and efficient solution for GCSE.

4. The data-driven nature of DA-CGAN enabled it to rapidly estimate GCSE, drastically reducing computational time. This time efficiency, combined with its high accuracy, makes DA-CGAN a promising framework for real-world applications.

In conclusion, the proposed DA-CGAN has demonstrated promising potential for accurate and efficient GCSE, exploring a novel potential of deep generative neural network for advanced applications in the field of hydrogeology. Our Future work will focus on improving the ability of model to handle real-world data noise and further refining its adaptive learning capabilities.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No.42272283), the National Natural Science Foundation of China (No.41972252) and the Graduate Innovation Fund of Jilin University (No. 2022060).

Declarations

Authors contributions Zidong Pan: Conceptualization, Writing - original draft, Software, Methodology. Wenxi Lu: Writing - review & editing, Methodology, Software. Yaning Xu: Supervision; Validation. Chengming Luo: Supervision; Validation. Yukun Bai: Supervision; Validation.

Funding: See the Acknowledgments section

Competing interest:None

Ethics Approval:Not applicable.

Consent to Participate:Not applicable.

Consent for Publication:Not applicable.

Open research

Data Availability statement:

The DA-CGAN was trained in a Python environment of version 3.7.0. The training data and code (DA-CGAN for GCSE) are available at 10.6084/m9.figshare.23714010.

Reference

An, Y., Yan, X., Lu, W., Qian, H., & Zhang, Z. (2022). An improved Bayesian approach linked to a surrogate model for identifying groundwater pollution sources [Article]. *Hydrogeology Journal*, *30*(2), 601-616. https://doi.org/10.1007/s10040-021-02411-2

Ayvaz, M. T. (2016). A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems. *Journal of Hydrology*, *538*, 161-176. https://doi.org/10.1016/j.jhydrol.2016.04.008

Bond-Taylor, S., Leach, A., Long, Y., & Willcocks, C. G. (2022). Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models [Review]. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7327-7347. https://doi.org/10.1109/tpami.2021.3116668

Chang, Z., Lu, W., & Wang, Z. (2021). A differential evolutionary Markov chain algorithm with ensemble smoother initial point selection for the identification of groundwater contaminant sources [Article]. *Journal of Hydrology*, *603*, Article 126918. https://doi.org/10.1016/j.jhydrol.2021.126918

Chen, Z., Gomez-Hernandez, J. J., Xu, T., & Zanini, A. (2018). Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter [Article]. *Journal of Hydrology*, *564*, 1074-1084. https://doi.org/10.1016/j.jhydrol.2018.07.073

Dagasan, Y., Juda, P., & Renard, P. (2020). Using Generative Adversarial Networks as a Fast Forward Operator for Hydrogeological Inverse Problems [Article]. *Groundwater*, *58*(6), 938-950. https://doi.org/10.1111/gwat.13005

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, 2014

Dec 08-13). Generative Adversarial Nets.*Advances in Neural Information Processing Systems* [Advances in neural information processing systems 27 (nips 2014)]. 28th Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA.

Jiang, S., Fan, J., Xia, X., Li, X., & Zhang, R. (2018). An Effective Kalman Filter-Based Method for Groundwater Pollution Source Identification and Plume Morphology Characterization [Article]. *Water*, *10*(8), Article 1063. https://doi.org/10.3390/w10081063

Jiang, S. M., Zhang, Y. L., Wang, P., & Zheng, M. H. (2013). An almost-parameter-free harmony search algorithm for groundwater pollution source identification. *Water Science and Technology*, *68*(11), 2359-2366. https://doi.org/10.2166/wst.2013.499

Jiang, X., Ma, R., Wang, Y., Gu, W., Lu, W., & Na, J. (2021). Two-stage surrogate model-assisted Bayesian framework for groundwater contaminant source identification [Article]. *Journal of Hydrology*, *594*, Article 125955. https://doi.org/10.1016/j.jhydrol.2021.125955

Kurtz, W., Hendricks Franssen, H.-J., Kaiser, H.-P., & Vereecken, H. (2014). Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions [Article]. *Water Resources Research*, *50*(2), 1665-1688. https://doi.org/10.1002/2013wr014823

Laloy, E., Herault, R., Jacques, D., & Linde, N. (2018). Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network [Article]. *Water Resources Research*, *54*(1), 381-406. https://doi.org/10.1002/2017wr022148

Li, J., Lu, W., Wang, H., Fan, Y., & Chang, Z. (2020). Groundwater contamination source identification based on a hybrid particle swarm optimization-extreme learning machine [Article]. *Journal of Hydrology*, *584*, Article 124657. https://doi.org/10.1016/j.jhydrol.2020.124657

Li, L., Puzel, R., & Davis, A. (2018). Data assimilation in groundwater modelling: ensemble Kalman filter versus ensemble smoothers [Article]. *Hydrological Processes*, *32*(13), 2020-2029. https://doi.org/10.1002/hyp.13127

Li, S., Ahn, C. K., Guo, J., & Xiang, Z. (2021). Neural-Network Approximation-Based Adaptive Periodic Event-Triggered Output-Feedback Control of Switched Nonlinear Systems [Article]. *Ieee Transactions on Cybernetics*, *51*(8), 4011-4020. https://doi.org/10.1109/tcyb.2020.3022270

Liang, Z. H., Huang, J. X., & Antani, S. (2022). Image Translation by Ad CycleGAN for COVID-19 X-Ray Images: A New Approach for Controllable GAN. *Sensors*, *22*(24), Article 9628. https://doi.org/10.3390/s22249628

Liu, H., Ong, Y.-S., & Cai, J. (2018). A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design [Review]. *Structural and Multidisciplinary Optimization*, *57*(1), 393-416. https://doi.org/10.1007/s00158-017-1739-8

Luo, J., Liu, Y., Li, X., Xin, X., & Lu, W. (2022). Inversion of groundwater contamination source based on a two-stage adaptive surrogate model-assisted trust region genetic algorithm framework [Article]. *Applied Mathematical Modelling*, *112*, 262-281. https://doi.org/10.1016/j.apm.2022.07.035

Moghaddam, M. B., Mazaheri, M., & Samani, J. M. V. (2021). Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: a review [Review]. *Groundwater for Sustainable Development*, *15*, Article 100651. https://doi.org/10.1016/j.gsd.2021.100651

Pan, Z., Lu, W., Wang, H., & Bai, Y. (2022). Fast inverse estimation of hydraulic conductivity field based on a deep convolutional-cycle generative adversarial neural network [Article]. *Journal of Hydrology*, *613*, Article 128420. https://doi.org/10.1016/j.jhydrol.2022.128420

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019, 2019

Dec 08-14). PyTorch: An Imperative Style, High-Performance Deep Learning Library.*Advances in Neural Information Processing Systems* [Advances in neural information processing systems 32 (nips 2019)]. 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, CANADA.

Sun, A. Y. (2018). Discovering State-Parameter Mappings in Subsurface Models Using Generative Adversarial Networks [Article]. *Geophysical Research Letters*, *45*(20), 11137-11146. https://doi.org/10.1029/2018gl080404

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *IEEE Computer Society*.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S., & Ieee. (2018, 2018

Jun 18-23). The iNaturalist Species Classification and Detection Dataset.*IEEE Conference on Computer Vision and Pattern Recognition* [2018 ieee/cvf conference on computer vision and pattern recognition (cvpr)]. 31st IEEE/CVF Conference on Computer Vision and Pattern

Recognition (CVPR), Salt Lake City, UT.

Wang, H., & Jin, X. (2013). Characterization of groundwater contaminant source using Bayesian method [Article]. *Stochastic Environmental Research and Risk Assessment*, *27*(4), 867-876. https://doi.org/10.1007/s00477-012-0622-9

Wang, Y.-Q., Wang, Q., Lu, W.-K., Ge, Q., & Yan, X.-F. (2022). Seismic impedance inversion based on cycle-consistent generative adversarial network [Article]. *Petroleum Science*, *19*(1), 147-161. https://doi.org/10.1016/j.petsci.2021.09.038

Xiao, N.-C., Zuo, M. J., & Zhou, C. (2018). A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis [Article]. *Reliability Engineering & System Safety*, *169*, 330-338. https://doi.org/10.1016/j.ress.2017.09.008

Xing, Z., Qu, R., Zhao, Y., Fu, Q., Ji, Y., & Lu, W. (2019). Identifying the release history of a groundwater contaminant source based on an ensemble surrogate model [Article]. *Journal of Hydrology*, *572*, 501-516. https://doi.org/10.1016/j.jhydrol.2019.03.020

Xu, T., Jaime Gomez-Hernandez, J., Chen, Z., & Lu, C. (2021). A comparison between ES-MDA and restart EnKF for the purpose of the simultaneous identification of a contaminant source and hydraulic conductivity [Article]. *Journal of Hydrology*, *595*, Article 125681. https://doi.org/10.1016/j.jhydrol.2020.125681

Yang, Q. C., Li, Z. J., Xie, C., Liang, J., & Ma, H. Y. (2020). Risk assessment of groundwater hydrochemistry for irrigation suitability in Ordos Basin, China. *Natural Hazards*, *101*(2), 309-325. https://doi.org/10.1007/s11069-018-3451-4

Yeh, W. W. G. (2015). Review: Optimization methods for groundwater modeling and management [Review]. *Hydrogeology Journal*, *23*(6), 1051-1065. https://doi.org/10.1007/s10040-015-1260-3

Yinka-Banjo, C., & Ugot, O.-A. (2020). A review of generative adversarial networks and its application in cybersecurity [Review]. *Artificial Intelligence Review*, *53*(3), 1721-1736. https://doi.org/10.1007/s10462-019-09717-4

Zanini, A., & Woodbury, A. D. (2016). Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion [Article]. *Journal of Contaminant Hydrology*, *185*, 74-86. https://doi.org/10.1016/j.jconhyd.2016.01.006

Zhang, J., Li, W., Lin, G., Zeng, L., & Wu, L. (2017). Efficient evaluation of small failure probability in high-dimensional groundwater contaminant transport modeling via a two-stage Monte Carlo method [Article]. *Water Resources Research*, *53*(3), 1948-1962. https://doi.org/10.1002/2016wr019518

Zhang, Q., Li, P., Lyu, Q., Ren, X., & He, S. (2022). Groundwater contamination risk assessment using a modified DRATICL model and pollution loading: A case study in the Guanzhong Basin of China [Article]. *Chemosphere*, *291*, Article 132695. https://doi.org/10.1016/j.chemosphere.2021.132695

Zhao, Y., Fan, D., Cao, S. H., Lu, W. X., & Yang, F. (2023). Visualization of biochar colloids transport and retention in two-dimensional porous media. *Journal of Hydrology*, *619*, Article 129266. https://doi.org/10.1016/j.jhydrol.2023.129266

Zhao, Y., Qu, R., Xing, Z., & Lu, W. (2020). Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms [Article]. *Advances in Water Resources*, *138*, Article 103540. https://doi.org/10.1016/j.advwatres.2020.103540

Zheng, N., Jiang, S., Xia, X., Kong, W., Li, Z., Gu, S., & Wu, Z. (2023). Efficient Estimation of Groundwater Contaminant Source and Hydraulic Conductivity by an ILUES Framework Combining GAN and CNN. *Journal of Hydrology*, 129677. https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129677

Zhou, H., Gomez-Hernandez, J. J., & Li, L. (2014). Inverse methods in hydrogeology: Evolution and recent trends [Review]. *Advances in Water Resources*, *63*, 22-37. https://doi.org/10.1016/j.advwatres.2013.10.014

Zhou, J., Su, X., & Cui, G. (2018). An adaptive Kriging surrogate method for efficient joint estimation of hydraulic and biochemical parameters in reactive transport modeling [Article]. *Journal of Contaminant Hydrology*, *216*, 50-57. https://doi.org/10.1016/j.jconhyd.2018.08.005

Zhu, J. Y., Park, T., Isola, P., Efros, A. A., & Ieee. (2017, Oct 22-29). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.*IEEE International Conference on Computer Vision* [2017 ieee international conference on computer vision (iccv)]. 16th IEEE International Conference on Computer Vision (ICCV), Venice, ITALY.