

Article

Informed Search Strategy for Synchronous Recognition of Groundwater Pollution Sources and Aquifer Parameters Based on an Improved DCN Substitute

Guanghua Li ^{1,2}, Han Wang ^{1,2,*}, Jiayuan Guo ¹, Jinping Zhang ^{1,2} and Wenxi Lu ³¹ Song-Liao River Water Resources Commission, Changchun 130000, China² River Basin Planning & Policy Research Center of Song-Liao River Water Resources Commission, Changchun 130000, China³ College of New Energy and Environment, Jilin University, Changchun 130000, China

* Correspondence: wangahan950505@163.com

Abstract: An informed search strategy based on random statistical analysis was developed for synchronous recognition of groundwater pollution source information and aquifer parameters. An informed search iterative course (ISIC) was accordingly designed, and each iteration included the determination of attempt point and state transition. In this paper, two improvement techniques were first adopted for choosing attempt points and judging state transition in ISIC to improve search efficiency and precision. The first improvement was that the variable radius free search method was applied to choosing the attempt point, and the size of the search radius was constantly adjusted in ISIC, taking the search ergodicity and efficiency into account. The second improvement technique was a *Tsallis* formula used for state transition judgment, and the controlled factor in the *Tsallis* formula was regulated continuously so that the search could consider ergodicity and efficiency simultaneously. Furthermore, frequent calls to the groundwater pollution numerical simulator to calculate the likelihood have inflicted a huge computational burden during ISIC. An effective way is to construct a substitute for emulating the simulator with a low calculating load. However, the mapping relation between the import and export of the numerical simulator was complex and had many variables. The precision of the substitute based on shallow learning is low sometimes. Therefore, we adopted the deep learning method and built an improved deep confidence network (DCN) substitute to emulate the highly nonlinear simulator. Finally, the synchronous recognition results for groundwater pollution source information and aquifer parameters were gained when ISIC ceased. The above-mentioned methods were verified in a case involving groundwater pollution. The consequence indicated that the ISIC with an improved DCN substitute can synchronously recognize groundwater pollution source information and aquifer parameters with a high degree of precision and efficiency.



Citation: Li, G.; Wang, H.; Guo, J.; Zhang, J.; Lu, W. Informed Search Strategy for Synchronous Recognition of Groundwater Pollution Sources and Aquifer Parameters Based on an Improved DCN Substitute. *Water* **2024**, *16*, 2143.
<https://doi.org/10.3390/w16152143>

Academic Editor: Lucila Candela

Received: 30 June 2024

Revised: 22 July 2024

Accepted: 25 July 2024

Published: 29 July 2024

Keywords: groundwater pollution; variable radius free search; *Tsallis* formula; substitute; deep confidence network



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Groundwater pollution source recognition (GPSR) has an important practical demand for the reasonable design of groundwater pollution remediation programs and accurate recognition of pollution responsibilities [1–4]. For practical GPSR cases, the actual situation of groundwater pollution is relatively complex, and there are several unknown factors [5–8]. The condition of groundwater pollution source and aquifer is unknown and should be recognized accurately. Therefore, partial groundwater pollution source information (release intensity in the first three periods) and aquifer parameters (the hydraulic conductivity of six zones in the research region and the longitudinal and transverse dispersity) were considered as unknown variables and synchronously recognized in this work [9–11].

The random statistical analysis based on the Bayesian formula is a popular method for GPSR [12–14]. Moreover, informed search refers to a search strategy that considers how to introduce relevant information in the search process to reduce the search scope and seek a solution as soon as possible. Based on the informed search theories and random statistical analysis, an informed search strategy for recognizing the unknown variables was first proposed in this paper. An informed search iterative course (ISIC) was accordingly arranged, and each iteration included attempt point selection and state transition judgment. The start point of ISIC (the first iteration) was mainly ascertained by professional experience. The start point of the following iterations was ascertained by the judgment of state transition. If the state transition happened, the attempt point was taken as the start point in the next iteration. Or the start point of the current iteration was taken as the start point in the next iteration. Thus, ISIC started from the original state (preliminary estimate of unknown variables) and ultimately reached the goal state (truth values of unknown variables).

Compared with previously informed search strategies, two improvement techniques were adopted in this work to attempt point selection and state transition judgment in ISIC to improve search efficiency and precision. The first improvement concerned the selection of attempt points. The attempt point was stochastically drawn from the neighborhood of the start point [15] in previous research, which generated a single attempt point with the poor ergodic property. Hence, the variable radius free search method was first exploited to acquire several attempt points, and the size of the search radius was constantly adjusted in ISIC considering the search ergodicity and efficiency. The second improvement was about the criterion for state transition. The common practice is to compare the ratio of the posterior probability of the attempt point and start point with a random number to judge whether to carry out a state transition [16]. To enhance the search ergodicity and efficiency, the *Tsallis* formula was applied to state transition judgment in each iteration, and the controlled factor in the formula was regulated continuously so that ISIC could take ergodicity and efficiency into account.

Furthermore, the numerical simulator needed to be called each iteration while calculating likelihood during ISIC, which resulted in a huge calculated load. Employing a substitute for the simulator can greatly reduce the calculated load. The frequently used modeling methods were basically shallow learning methods. Nevertheless, the shallow learning methods cannot fit the complex mapping relation between the import and export of the numerical simulator with a great many variables, and the approaching precision of substitute to simulator was sometimes low [17–19]. Recently, the deep learning method has shown great potential in fitting complex mapping relations [20–22]. Therefore, the deep confidence network (DCN) was introduced in this paper to build a substitute for the simulator. Two improvements for DCN were first proposed in this paper to further enhance the fitting capacity of complex mapping relations. Accordingly, an improved DCN substitute was established for the simulator to improve the approaching precision of the substitute to the simulator. The first improvement was to propose a hybrid training method combining unsupervised and supervised training in the pre-training phase. This practice could greatly enhance the feature extraction ability of input data. The second improvement was the use of the Levenberg–Marquardt (LM) algorithm after pre-training, which could particularly fine-adjust the network parameters of DCN globally. The validity of the proposed ISIC with an improved DCN substitute was tested by a hypothetical example of groundwater pollution.

Several novel approaches were taken in this study. (1) The variable radius free search method was first proposed for attempt point selection to substantially improve the search ergodicity. (2) The *Tsallis* formula was first applied to state transition judgment so that ISIC could take ergodicity and efficiency into account. (3) Furthermore, an improved DCN substitute was established by substituting the numerical simulator to decrease the computing load and enhance the approaching precision of the substitute to the simulator. (4) Finally, a mathematical analysis of groundwater pollution source information and aquifer parameters was conducted for GPSR. The analysis can provide interval estimation

and point estimation for unknown variables and help us better understand the real situation of unknown variables for practical GPSR. Figure 1 shows the flow chart of the novel approaches in this paper.

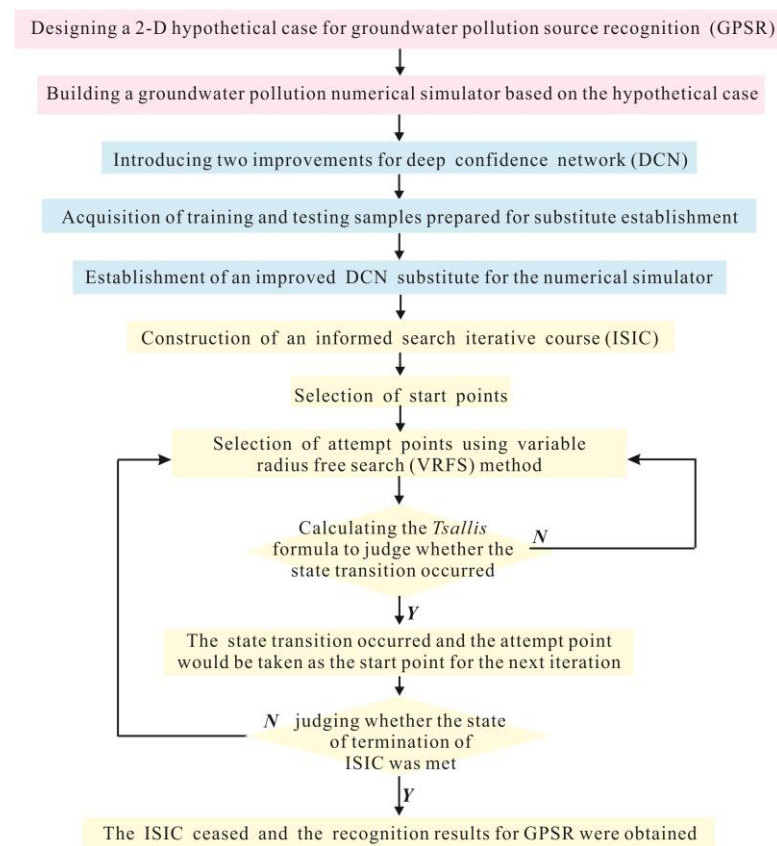


Figure 1. The flow chart of the novel approaches in this paper.

2. Methodology

2.1. The Numerical Simulator

The groundwater contaminant transportation is described by the groundwater flow model and solute transport model. The formula involving groundwater steady flow is described as follows:

$$\frac{\partial}{\partial x_i} \left(K_{ij} \frac{\partial h}{\partial x_j} \right) = 0 \quad i, j = 1, 2 \quad (1)$$

where $\frac{\partial}{\partial x}$ denotes gradient operator vector; h represents hydraulic head; and K_{ij} denotes hydraulic conductivity tensor.

The formula involving the solute transport model is described as follows:

$$\frac{\partial(\phi c)}{\partial t} = \frac{\partial}{\partial x_i} \left(\phi D_{ij} \frac{\partial c}{\partial x_j} \right) + - \frac{\partial(\phi c v_i)}{\partial x_i} + \frac{c_s Q}{d} \quad (2)$$

$$v_i = - \frac{K_{ij}}{\phi} \frac{\partial h}{\partial x_j} \quad i, j = 1, 2 \quad (3)$$

where ϕ denotes effective porosity; t denotes time; D_{ij} denotes hydrodynamic dispersion tensor; v_i represents velocity; c_s represents groundwater solute concentration; and d represents aquifer thickness. The MODFLOW and MT3DMS computing modules were exploited to resolve the mathematical equations controlling the groundwater flow and groundwater pollution transport in this work.

2.2. The DCN Method

The deep confidence network (DCN) was put forward by Hinton (2006) [23], which was a fusion of probabilistic statistics, machine learning and neural networks (Figure 2) (Hinton et al., 2014) [24]. DCN was a complex neural network structure consisting of multiple hidden layers designed to capture high-level abstract features in data. DCN was stacked with Restricted Boltzmann Machines (RBMs) or other types of generative models (Figure 3). Each RBM layer learned the distribution of the input data and attempted to capture features in the data. By stacking, each layer is further abstracted from the features captured in the previous layer, allowing the network to learn more complex data features. Accordingly, DCN was utilized to build a substitute for the numerical simulator. To further enhance the fitting capacity of the nonlinear mapping relation between the import and export of the simulator, two improvements for DCN were proposed in this paper, and an improved DCN substitute was established for the simulator. The first improvement was to propose a hybrid training method (also called partially supervised pre-training) combining unsupervised and supervised training in the pre-training phase. The second improvement was to use the Levenberg–Marquardt (LM) algorithm after the pre-training phase to fine-adjust the network parameters of DCN globally.

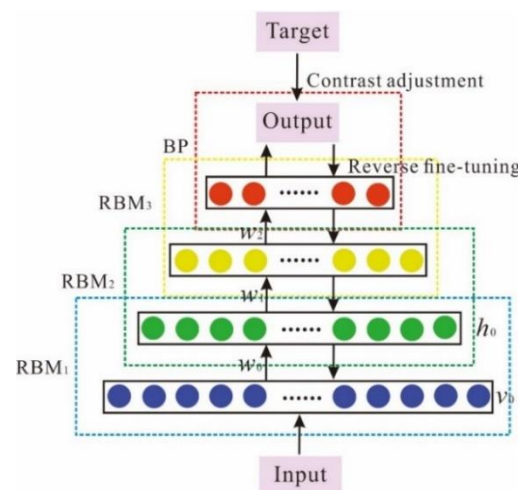


Figure 2. Structure diagram of DCN.

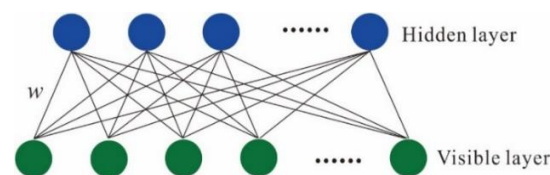


Figure 3. Schematic diagram of the RBM structure.

The two phases of improved DCN were described as follows:

(1) Partially supervised pre-training.

The pre-training of DCN refers to the layer-by-layer training of RBM. RBM is made up of visible and hidden layers, and the visible and hidden layers are interconnected. The export of the hidden unit can obtain the higher-order correlation of the visible unit.

RBM is a neural network model based on an energy model. v_i represents the visible unit, h_j represents the hidden unit, and E represents the energy function of RBM:

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j \quad (4)$$

where $\theta = \{\mathbf{w}, \mathbf{a}, \mathbf{b}\}$, w_{ij} represents the connection weight, a_i and b_j represent the bias, n and m represent the quantity of visible and hidden units, $E(\mathbf{v}, \mathbf{h}; \theta)$ represents energy function, $P(\mathbf{v}, \mathbf{h}; \theta)$ represents the joint probability distribution:

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{e^{-E(\mathbf{v}, \mathbf{h}; \theta)}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}; \theta)}} \quad (5)$$

The conditions of the hidden units in RBM are mutually independent, so the probability that h_j is equal to 1 when given the visible vector \mathbf{v} is:

$$P(h_j = 1 | \mathbf{v}) = \sigma \left(\sum_{i=1}^n w_{ij} v_i + b_j \right) \quad (6)$$

where $\sigma(x) = (1 + e^{-x})^{-1}$ represents the sigmoid activation function.

Given the condition of the hidden unit, the probability that v_i is equal to 1 is:

$$P(v_i = 1 | \mathbf{h}) = \sigma \left(\sum_{j=1}^m w_{ij} h_j + a_i \right) \quad (7)$$

For the above RBM model, the training dataset is considered a condition of the visible unit. Then, the condition of the hidden unit is computed based on Equation (6), and the updated condition of the visible unit is calculated according to Equation (7) $\mathbf{v}' = (v'_i)$. Finally, the updated condition of the hidden unit is computed based on Equation (6), and $\mathbf{h}' = (h'_j)$. The updated formula of the parameter is as follows:

$$\begin{cases} \Delta w_{ij} = \varepsilon_{CD} (\langle v_i h_j \rangle - \langle v'_i h'_j \rangle) \\ \Delta a_i = \varepsilon_{CD} (\langle v_i \rangle - \langle v'_i \rangle) \\ \Delta b_j = \varepsilon_{CD} (\langle h_j \rangle - \langle h'_j \rangle) \end{cases} \quad (8)$$

where ε_{CD} represents the study rate and $\langle \cdot \rangle$ represents mathematical expectation.

The import variable of the substitute for the numerical simulator is continuous data, so the Gaussian–Bernoulli RBM (GB-RBM), whose visible and hidden units are linear random units and binary random units, respectively, was taken as the first RBM in this paper. For GB-RBM, the energy function is described as follows:

$$E(\mathbf{v}, \mathbf{h}; \theta) = \sum_{i=1}^n \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{i=1}^n \sum_{j=1}^m w_{ij} \frac{v_i}{\sigma_i} h_j - \sum_{j=1}^m b_j h_j \quad (9)$$

where σ_i represents the Gaussian noise standard deviation of the visible unit. To simplify the calculation, $\sigma_i = 1$, Equation (7) can be expressed as:

$$P(v_i | \mathbf{h}) = N \left(\sum_{j=1}^m w_{ij} h_j + a_i, 1 \right) \quad (10)$$

where v_i follows the Gaussian distribution whose mean is $\sum_{j=1}^m w_{ij} h_j + a_i$ and variance is 1. To simplify the calculation, the variance is set to 1 [25,26].

In this paper, the partially supervised pre-training method was first used to train the first hidden layer, and the unsupervised training was carried out on each layer of DCN subsequently. When training the first layer, $\mathbf{x} = [x_1, x_2, \dots, x_n]$ would be used as the import of RBM1 and $\{\mathbf{w}_1, \mathbf{a}_1, \mathbf{b}_1\}$ would be gained. Then, the activation probability function of the hidden unit in RBM1 was considered an import of RBM2 and $\{\mathbf{w}_2, \mathbf{a}_2, \mathbf{b}_2\}$ gained based on Equation (8). The activation probability function of the hidden unit in RBM2 was

considered as the import of RBM3, and the unsupervised learning was continued based on Equation (8). Afterward, the layers were followed until the top level, and the original values of the weight and bias of DCN could be gained.

(2) Supervised fine adjustment phase based on the LM algorithm.

After pre-training, all parameters in DCN needed an intense adjustment to gain a global optimal solution for the network parameters. Given the training dataset $\{\mathbf{x}_{(t)}, y_{(t)}\}$ ($t = 1, 2, \dots, N$), the relation of import–export is described as:

$$\hat{y}_{(t)} = f(\mathbf{x}_{(t)}, \boldsymbol{\theta}) \quad (11)$$

where f represents a nonlinear function, $\mathbf{x}_{(t)} = [x_1, x_2, \dots, x_n]$ is the import variable from the t th training dataset, $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]$ is an unknown vector composed of all parameters of weights and bias, and p is the number of parameters.

The expression of the error loss function (sum of the squares of error between the export value of the neural network and the target export value) is:

$$E(\boldsymbol{\theta}) = \sum_{t=1}^N e_t^2 = \sum_{t=1}^N (y_{(t)} - f(\mathbf{x}_{(t)}, \boldsymbol{\theta}))^2 \quad (12)$$

where N represents the number of training samples and $y_{(t)}$ represents the target export of the t th training sample.

In this paper, LM algorithm was first used after the pre-training phase to adjust the network parameters of DCN globally. A set of parameters of $\boldsymbol{\theta}$ was acquired to minimize the error loss function on account of the parameters gained in the pre-training phase. If the outcome of the k th iteration was $\boldsymbol{\theta}_k$, the updated calculation equation of the parameter vector was as follows:

$$\Delta \boldsymbol{\theta}_k = -[\mathbf{J}^T(\boldsymbol{\theta}_k)\mathbf{J}(\boldsymbol{\theta}_k) + \mu \mathbf{I}]^{-1} \mathbf{J}^T(\boldsymbol{\theta}_k)E(\boldsymbol{\theta}_k) \quad (13)$$

where the damping factor is a constant and $\mu > 0$. \mathbf{I} is the Jacobian matrix, and:

$$\mathbf{J}(\boldsymbol{\theta}_k) = \begin{bmatrix} \frac{\partial e_1(\boldsymbol{\theta}_k)}{\partial \theta_1} & \frac{\partial e_1(\boldsymbol{\theta}_k)}{\partial \theta_2} & \dots & \frac{\partial e_1(\boldsymbol{\theta}_k)}{\partial \theta_p} \\ \frac{\partial e_2(\boldsymbol{\theta}_k)}{\partial \theta_1} & \frac{\partial e_2(\boldsymbol{\theta}_k)}{\partial \theta_2} & \dots & \frac{\partial e_2(\boldsymbol{\theta}_k)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(\boldsymbol{\theta}_k)}{\partial \theta_1} & \frac{\partial e_N(\boldsymbol{\theta}_k)}{\partial \theta_2} & \dots & \frac{\partial e_N(\boldsymbol{\theta}_k)}{\partial \theta_p} \end{bmatrix} \quad (14)$$

The LM algorithm found the appropriate damping factor by constantly searching for iterations. When $\mu = 0$, the algorithm was transformed into the Gauss–Newton method of least squares solution. When the value μ was large, the algorithm was transformed into the Fastest Gradient Descent method. Therefore, in the early phase of ISIC, the LM algorithm had the characteristics of large initial descent and rapid convergence of the Fastest Gradient Descent method. In the later phase, the LM algorithm had the characteristics of local convergence of the Gauss–Newton method so that the network parameters could quickly converge to the global optimal solution.

2.3. State Evaluation Function

The state evaluation function can be written as (Lu et al., 2020) [27]:

$$\eta(X) = p(X|S)p(Y|X, S) \quad (15)$$

where X represents the unknown variable, Y represents contaminant concentration monitoring data, S represents designed concentration monitoring, $p(X|S)$ denotes prior probability, $p(Y|X, S)$ and denotes likelihood.

2.4. Variable Radius Free Search Method

The free search (FS) is an algorithm of swarm intelligence algorithm proposed by Penev and Littlefair (2005) [28]. The FS algorithm is not a simple simulation of the living habits of a social animal, but a comprehensive simulation of the biological characteristics and living habits of a variety of organisms. In the process of searching, individuals consider the accumulated experience/knowledge acquired in the past but are not restricted solely to this information and can search freely in any region within the prescribed scope. Flexibility is an important feature of FS algorithms. Individuals can conduct local or global searches and determine the size of the search step by themselves.

In this paper, a linear decreasing variable radius strategy was first introduced to improve the FS algorithm, and the variable radius FS (VRFS) algorithm was first constructed. To enhance the search efficiency and search precision in ISIC, the probe walk strategy in the VRFS algorithm was first used for the attempt point selection in this paper. The size of the search radius constantly changes in ISIC. In the early phase of ISIC, a larger radius was used to reinforce global search with high search ergodicity, and a smaller radius was used to reinforce local search and enhance search accuracy in the later phase.

The operational process of the VRFS algorithm is as follows:

Step 1: Set the range of the search radius $[R_{\min}, R_{\max}]$. The present number is k ; the total iteration number is N , and the sum of unknown variables is n .

Step 2: Set the initial value range of each unknown variable and the values of m start points throughout the entire search iteration process. $X_1^k, X_2^k, \dots, X_m^k$ is determined based on professional experience. In the first iteration, $k = 1$. Among the m start points, the j th start point is $X_j^k = (x_{j1}^k, x_{j2}^k, \dots, x_{jn}^k)$, $j = 1, 2, \dots, m$, and the i th unknown variable is x_{ji}^k .

Step 3: Determine m attempt points corresponding to m start points separately, and the j th attempt point is $X_j^{k'} = (x_{j1}^{k'}, x_{j2}^{k'}, \dots, x_{jn}^{k'})$.

The calculation formula for the i th unknown is as follows:

$$x_{ji}^{k'} = x_{ji}^k - \Delta x_{ji}^k + 2\Delta x_{ji}^k \text{random}_{ji}^k(0, 1) \quad (16)$$

$$\Delta x_{ji}^k = R_{ji}^k (x_{i\max}^k - \Delta x_{i\min}^k) \text{random}_{ji}^k(0, 1) \quad (17)$$

$$R_{ji}^k = R_{\max} - k(R_{\max} - R_{\min})/N \quad (18)$$

For each set of start and attempt points, the start point for the next iteration is determined based on a certain state transition criterion, and a total of m start points is obtained.

Step 4: Let $k = k + 1$, go back to step 3 until $k > N$, stop the iteration, and export all the above values of unknown variables.

2.5. The Tsallis Formula Based on State Evaluation Function

The concept of the *Tsallis* formula proposed in this paper was originally derived from the simulated annealing (SA) algorithm. The SA algorithm is widely used to solve inverse problems because of its simple principle and ease of operation. To enhance the computing efficiency of the SA algorithm, many improved methods have been proposed, such as the fast simulated annealing (FSA) and very fast simulated annealing (VFSA) algorithms.

In the FSA algorithm, the new acceptance probability expression, i.e., the mathematical expression of the *Tsallis* formula, is:

$$P = [1 - (1 - h)\Delta E/T]^{1/(1-h)} \quad (19)$$

where P denotes the acceptance probability; ΔE represents the energy difference; T is the temperature; and h is a real number.

The annealing method in the VFSA algorithm is as follows (Robini, 2013) [29]:

$$T(k) = T_0 \exp(-Ck^{1/N}) \quad (20)$$

where N is the number of unknown variables, T_0 is the initial temperature, k is the number of iterations, and C is the attenuation factor, which is a given constant.

To enhance the search efficiency and precision in ISIC, the *Tsallis* formula was first constructed to judge the state transition in this paper. We embedded the state evaluation function into the *Tsallis* formula and used the calculation result of the *Tsallis* formula to judge the state transition. The approaching degree of the search path to the target state was improved by utilizing the guidance and correction function of the useful information contained in the state evaluation function during ISIC. Moreover, by learning from the annealing method in the VFSA algorithm, the controlled factor in the *Tsallis* formula was adjusted. By adjusting the size of the controlled factor, the ISIC can take search ergodicity and efficiency into account. The useful information includes field monitoring data and professional experience. The *Tsallis* formula was computed in each iteration for judging the state transition.

If the condition of state transition was met, the state transition occurred, and the attempted point would be taken as the starting point for the next iteration. Otherwise, the state transition did not occur, and the start point of the current iteration would be taken as the start point for the next iteration.

The judging process of state transition is as follows.

Step 1: Given the initial value T_1 of the controlled factor T , the current value T_l and the attenuation factor C ($0.7 \leq C < 1$) were determined. The sum of unknown variables is n , $T_l = T_1 \exp(-Cl^{1/n})$. The number of adjustments of the controlled factor is L , and under the conditions of the same controlled factor T , the number of iterations is N , and the label of the iteration number is k .

Step 2: An initial estimate for each unknown variable is given according to experience, and the start point X^k of ISIC is determined, $X^k = (x_1^k, x_2^k, \dots, x_n^k)$. In the first iteration, $k = 1$, $l = 1$, $T = T_l$.

Step 3: The attempt point $X^{k'}$ in this iteration is determined using one method, and the value of the state evaluation function $\eta(X)$ of the start point and the attempt point is computed, $\Delta\eta = \eta(X^k) - \eta(X^{k'})$. If $\Delta\eta < 0$, then let $X^{k+1} = X^{k'}$, otherwise calculate $\beta = (1 - (1 - h) \frac{\Delta\eta}{T_l})^{\frac{1}{1-h}}$ ($h > 1$). In this study, $h = 6$. Compare β with the random number $u = \text{rand}(0, 1)$. If $\beta > u$, then make $X^{k'} = X^{k+1}$; otherwise, let $X^k = X^{k+1}$.

Step 4: Let $k = k + 1$, go back to step 3, until $k > N$, perform step 5.

Step 5: Let $l = l + 1$, $T = T_l$, return to step 3 until $l > L$, cease the iteration and export all the iteration values.

In this study, the original values of the controlled factor T and the attenuation factor were $T_1 = 10$, $C = 0.98$.

3. Case Study

3.1. Site Overview

This work chose a two-dimensional case of groundwater inorganic pollution. The specific condition of the research region could be seen in Figure 4. The BC border and DA border of the research region were borders of no flow (relative confining borders). The AB border and CD border were borders of varying specified heads, and the variation was nonlinear (the river completely cuts the aquifer). The groundwater flow was unsteady, and the groundwater pollution in this region was conservative. No pollutants were in the research region before the groundwater pollution was released. The simulation time (a total of 24 months) for the groundwater pollution numerical simulator was evenly divided into eight periods, and the research region was divided into six zones in terms of hydraulic conductivity. During the first three periods, the pollution source continuously released pollutants into the aquifer and then stopped. There were three monitoring wells in the research region, and the concentrations of groundwater pollution at each monitoring well were collected. The release intensity in the first three periods, the longitudinal and transverse dispersity, and the hydraulic conductivity in six zones were unknown variables in this research region. Table 1 shows the truth values of unknown variables. Other

parameters describing the condition of the groundwater pollution source and aquifer were considered known (Table 2). By importing the truth values of unknown variables into the numerical simulator, the simulator has been operated to obtain the pollution concentration exports at three monitoring wells in eight periods, considered the monitoring data of groundwater pollution (Figure 5).

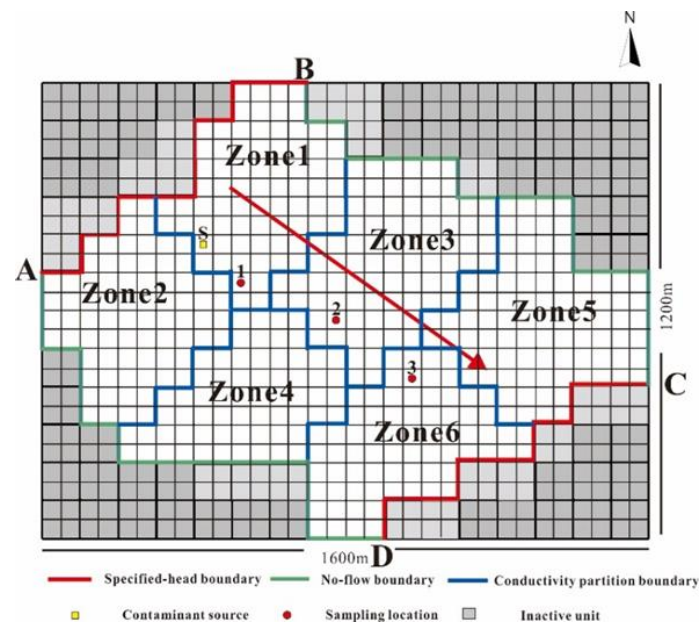


Figure 4. Hypothetical aquifer model of the study case.

Table 1. Truth values of release intensity of the pollution source, longitudinal and transverse dispersivity and hydraulic conductivity. P1, P2, and P3 indicate the release intensity of the pollution source in the first three periods, respectively. K1, K2, K3, K4, K5, and K6 represent the hydraulic conductivity in zones 1, 2, 3, 4, 5, and 6. a_L and a_T represent the longitudinal and transverse dispersivity.

Unknown Variables	Truth Value
P1 (g/s)	29.3
P2 (g/s)	21.3
P3 (g/s)	12.8
K1 (m/d)	143.5
K2 (m/d)	36.2
K3 (m/d)	13.4
K4 (m/d)	212.4
K5 (m/d)	59.8
K6 (m/d)	7.2
a_L (m)	20
a_T (m)	4

Table 2. Parameters describing the condition of the aquifer for the case study.

Parameter	Value
Effective porosity	0.3
Confined aquifer thickness (m)	30
Grid spacing in x -direction (m)	50
Grid spacing in y -direction (m)	50

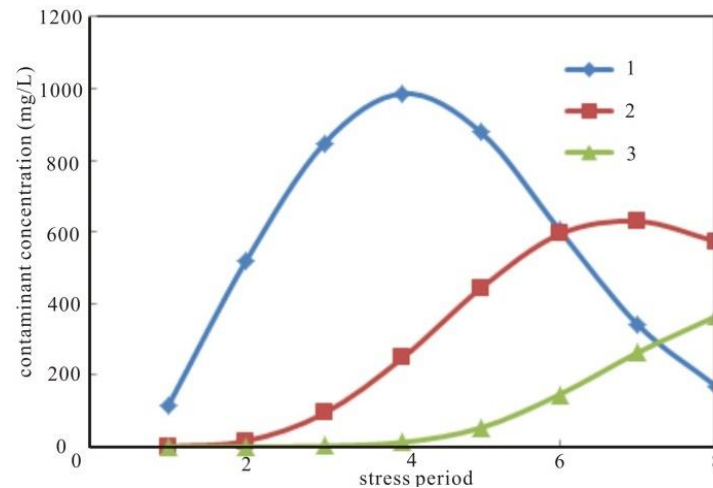


Figure 5. Accurate pollution concentrations at three monitoring wells.

3.2. Application of the Improved DCN Substitute

The above eleven unknown variables constituted an eleven-dimensional vector as the import vector of the improved DCN substitute for the monitoring well. The groundwater pollution concentrations in eight periods were considered as the exports of the improved DCN substitute. Initially, the import variables in 400 sets of training samples and 100 sets of testing samples were obtained by a sampling approach in their feasible regions. After that, the numerical simulator was operated to acquire the corresponding pollution concentrations (exports), so 500 sets of import–export samples were obtained. Then, 400 import–export training samples were utilized to establish a substitute for three monitoring wells, and the 100 import–export test samples were utilized to test the approaching precision of the substitute to the simulator. The establishment process for the improved DCN substitute in this work was as follows:

(1) Dataset normalization processing.

The import dataset of the substitute was normalized.

(2) Partially supervised pre-training.

We used the training samples to partially supervise pre-train the DCN substitute and ascertain the original network parameters of the DCN substitute. The DCN substitute in this work was composed of a three-layer RBM and a top-level BP neural network. We used the training dataset to train the DCN substitute and ascertained the number of units in the hidden layer. First, we ascertained the optimal quantity of hidden units in the first hidden layer and fixed it. Then, we added a hidden layer and ascertained the optimal quantity of hidden units in the second hidden layer. And so on, until the precision of the network is no longer enhanced. The neuron quantities in the three layers of RBM were finally set to 128, 64 and 64, respectively.

(3) Supervised fine adjustment.

We used the LM algorithm to fine-adjust the weight and bias of the network and obtained the optimal solution for the DCN substitute. The import dataset was imported into the fine-adjusted DCN substitute to gain the export dataset.

(4) Inverse dataset normalization.

We reversely normalized the export dataset in step (3).

3.3. Application of State Evaluation Function

3.3.1. Prior Information

Table 3 shows the prior information on the unknown variables in this work. The start point of ISIC was ascertained according to the prior information.

Table 3. Prior information for eleven unknown variables.

Unknown Variable	Preliminary Estimating Value	Preliminary Value Range	Prior Distribution	Preliminary Mean	Preliminary Variance
K1 (m/d)	100	(50, 200)	Normal distribution	100	500
K2 (m/d)	50	(20, 100)	Normal distribution	50	200
K3 (m/d)	40	(20, 100)	Normal distribution	40	150
K4 (m/d)	150	(100, 300)	Normal distribution	150	700
K5 (m/d)	30	(10, 70)	Normal distribution	30	100
K6 (m/d)	20	(5, 60)	Normal distribution	20	90
P1 (g/s)	5	(0, 50)	Normal distribution	5	40
P2 (g/s)	10	(0, 50)	Normal distribution	10	80
P3 (g/s)	35	(0, 50)	Normal distribution	35	150
a_L (m)	30	(5, 40)	Normal distribution	25	80
a_T (m)	6	(1, 8)	Normal distribution	4	30

3.3.2. Likelihood Function

To decrease the huge calculation load in ISIC and maintain the computational precision, rather than the numerical simulator, the DCN substitute was used to obtain the export value of the groundwater pollution concentration in likelihood. The expression of the likelihood can be referred to our previous research [30]. The monitoring well number was 3, and the period number was 8 in this work.

3.4. Informed Search Iterative Course

We first designed an informed search iterative course (ISIC) and made the best of the guiding function of the groundwater pollution monitoring data, considering search ergodicity and efficiency. The selection of attempt points and judgment of state transition constituted the content of each iteration. For the first iteration, the start points were ascertained by field investigation and specialized knowledge. For the following iterations, the start points were ascertained by the judging criterion of state transition. In each iteration, the attempt points were ascertained by the VRFS method, and the *Tsallis* formula was calculated to judge whether the state transition occurred. If the condition for state transition was met, the state transition occurred, and the attempted point would be taken as the starting point for the next iteration. Otherwise, the state transition did not occur, and the start point of the current iteration would be taken as the start point for the next iteration. The ISIC is as follows: When the state of termination was met, the overall process ceased. Meanwhile, the recognition results for groundwater pollution source information and aquifer parameters were acquired.

4. Results and Discussion

4.1. Performance of the Improved DCN Substitute

The 100 import–export test samples in the above research were utilized to test the approaching precision of the substitute to the numerical simulator. The improved DCN substitute was calculated based on the 100 imports, and 100 exports of the substitute were obtained. A total of 100 exports involved 800 groundwater pollution concentration values,

and the exports from substitute and numerical simulator were compared. The scatter diagram of the numerical simulator exports and substitute exports of three monitoring wells is shown in Figure 6. The scatters were distributed mainly around the line $y = x$, which proved that the export of the numerical simulator was very close to that of the improved DCN substitute.

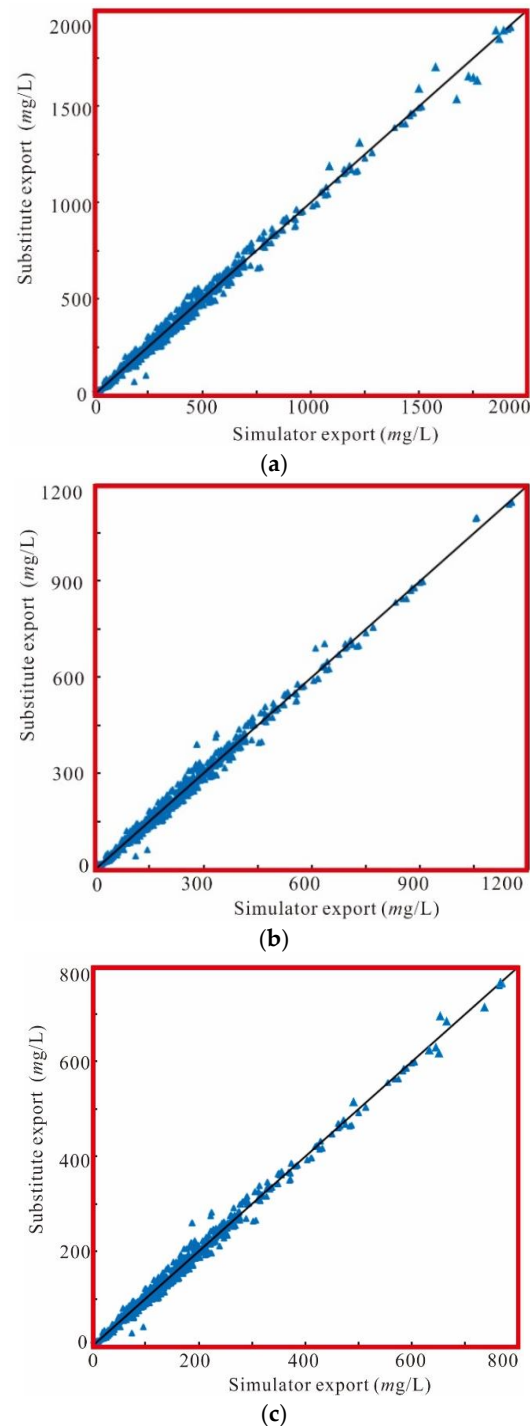


Figure 6. Fitting graph of the improved DCN substitute exports and the simulator exports of monitoring wells 1 (a), 2 (b) and 3 (c).

The precision evaluation index of the correlation coefficient of the improved DCN substitute was calculated (Table 4). The closer the correlation coefficient value was to 1, the higher the approaching precision of the substitute was to the numerical simulator. The

computed outcome of the precision index demonstrated that the improved DCN substitute possessed a high approaching precision to the numerical simulator for three monitoring wells. Therefore, the improved DCN substitute could replace the numerical simulator in the following synchronous recognition of groundwater pollution source information and aquifer parameters.

Table 4. Precision evaluation index of the improved DCN substitute of three monitoring wells.

Monitoring Well	Correlation Coefficient
1	0.9912
2	0.9927
3	0.9935

4.2. Recognition Results

The iteration number in ISIC was 20,000, and the number of start points or attempt points in each iteration was set to five. As a result, the simulator would be operated 100,000 times. Each operation of the simulator took 20 s on a computer, while the improved DCN substitute took only 0.8 s, greatly reducing the computation load. In this work, the scale reduction score was considered as the convergence criterion to judge the occurrence of convergence [31]. Figure 7 shows the variation of the scale reduction score of eleven unknown variables, which are <1.2 at around 15,000 iterations. Hence, ISIC converged at around 15,000 iterations, and the first 15,000 iterations were considered as “burn-in”. In this paper, the last 500 iterations after convergence were selected, the corresponding 2500 values of unknown variables were considered sampled from the distribution of unknown variables, and the statistical analysis of each unknown variable was carried out.

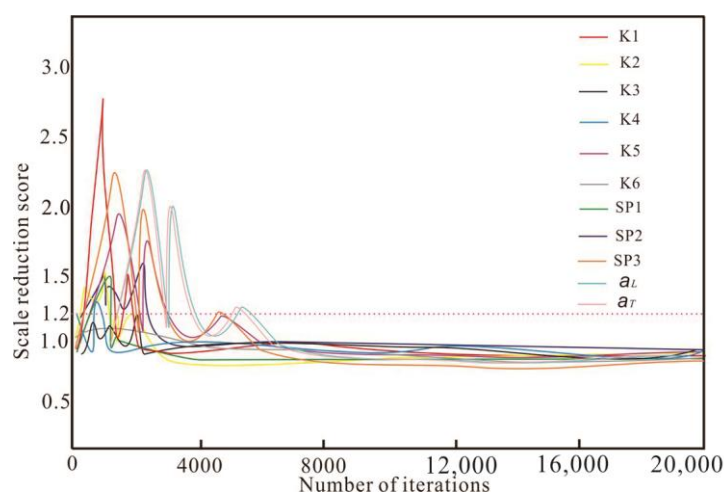


Figure 7. Curves of scale reduction score for eleven unknown variables.

The statistics analysis included the point and interval estimations. For the interval estimation, the posterior probability density function curves were plotted based on 2500 values of all unknown variables (Figure 8), which clearly showed the interval estimations for eleven unknown variables. For the point estimation, the value with the maximum posterior probability density function (PPDF) was considered as the point estimation for each unknown variable. The outcome showed that the point estimations were close to the truth values (Table 5). For the eleven unknown variables, the PPDFs are all concentrated near the truth value. As for the hydraulic conductivity in six zones, there are two peak values in the PPDF. The two peak values are relatively close for K1 and K6 but dispersed for K2~K5. As for the release intensity in the first three periods, the peak values are relatively dispersed for P1 and P2 but more concentrated for P3. Meanwhile, for longitudinal and transverse dispersivity, the two peak values are relatively concentrated around the truth value.

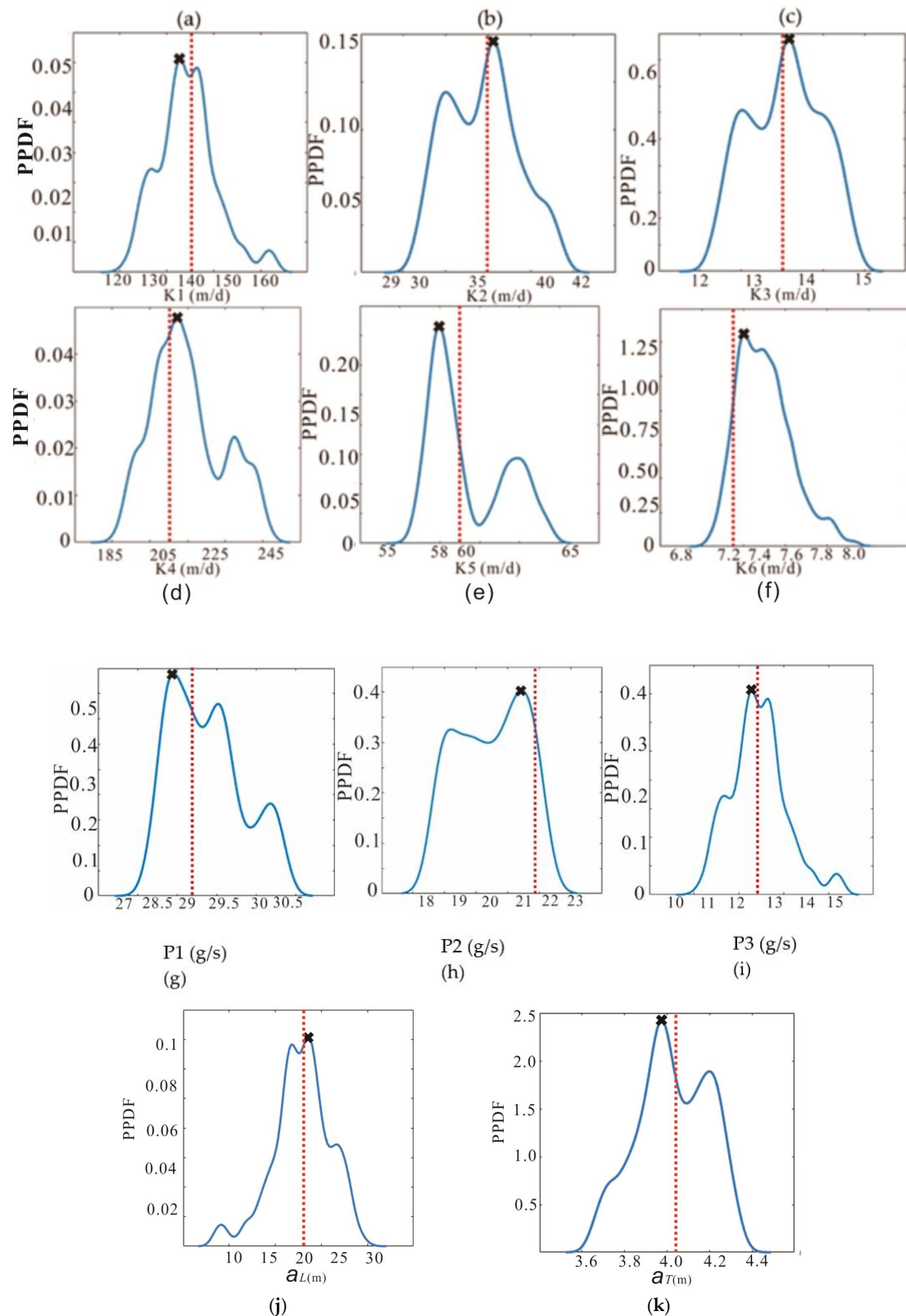


Figure 8. PPDF curves (blue lines) with true values (vertical dotted red lines) of eleven unknown variables. The black crosses are the point estimations of eleven unknown variables. (a) PPDF curves of K_1 ; (b) PPDF curves of K_2 ; (c) PPDF curves of K_3 ; (d) PPDF curves of K_4 ; (e) PPDF curves of K_5 ; (f) PPDF curves of K_6 ; (g) PPDF curves of P_1 ; (h) PPDF curves of P_2 ; (i) PPDF curves of P_3 ; (j) PPDF curves of a_L ; (k) PPDF curves of a_T .

Table 5. Relative errors between point estimations and true values of eleven unknown variables.

Unknown Variable	True Value	Point Estimation	Relative Error
K1 (m/d)	143.5	139.6	2.72%
K2 (m/d)	36.2	37.2	2.76%
K3 (m/d)	13.4	13.7	2.24%
K4 (m/d)	212.4	216.8	2.07%
K5 (m/d)	59.8	58.4	2.34%
K6 (m/d)	7.2	7.4	2.78%
P1 (g/s)	29.3	28.7	2.05%
P2 (g/s)	21.3	20.8	2.35%
P3 (g/s)	12.8	12.5	2.34%
a_L (m)	20.0	20.7	3.50%
a_T (m)	4.0	3.87	3.25%

5. Conclusions

The ISIC was designed for GPSR in this paper, and each iteration included the determination of attempt point and state transition. Two improvements were adopted in this paper in ISIC to improve search efficiency and precision. The first improvement was that the VRFS method was exploited for the attempt point selection, and the size of the search radius was regulated constantly, taking ergodicity and efficiency into account. The second improvement was that the *Tsallis* formula was used for state transition judgment, and the controlled factor in the *Tsallis* formula was regulated constantly so that the search could consider both ergodicity and efficiency. Furthermore, to reduce the huge calculation load caused by repeatedly solving the numerical simulator in ISIC, the deep confidence network (DCN) method was introduced to build a substitute for the simulator. An improved DCN substitute was established to improve the approximation precision of the substitute to the simulator further. As ISIC went on, the synchronous recognition results for groundwater pollution sources and aquifer parameters eventually gained when ISIC ceased. The above-mentioned methods were verified through a case of groundwater pollution, and the outcome indicated that the ISIC with an improved DCN substitute could assist in synchronously recognizing groundwater pollution source information and aquifer parameters with high precision and efficiency.

Author Contributions: All authors contributed to the study conception and design. Conceptualization was performed by G.L.; methodology was performed by H.W.; software was performed by J.G.; Validation was performed by J.Z. and W.L. The first draft of the manuscript was written by G.L. and all authors commented on previous versions. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 41972252) (funder: Wenxi Lu).

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Asher, M.J.; Croke, B.F.W.; Jakeman, A.J.; Peeters, L.J.M. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* **2015**, *51*, 5957–5973. [[CrossRef](#)]
2. Ayvaz, M.T. A linked simulation-optimization model for solving the unknown groundwater pollution source identification problems. *J. Contam. Hydrol.* **2010**, *117*, 46–59. [[CrossRef](#)] [[PubMed](#)]
3. Yao, Y.; Yang, F.; Suuberg, E.M.; Provoost, J.; Liu, W. Estimation of contaminant subslab concentration in petroleum vapor intrusion. *J. Hazard. Mater.* **2014**, *279*, 336–347. [[CrossRef](#)] [[PubMed](#)]
4. Zanini, A.; D'Oria, M.; Tanda, M.G.; Woodbury, A.D. Coupling empirical Bayes and Akaike's Bayesian information criterion to estimate aquifer transmissivity fields. *Math. Geosci.* **2020**, *52*, 425–441. [[CrossRef](#)]

5. Guo, J.; Lu, W.; Yang, Q.; Miao, T. The application of 0–1 mixed integer nonlinear programming optimization model based on a surrogate model to identify the groundwater pollution source. *J. Contam. Hydrol.* **2019**, *220*, 18–25. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Mirghani, B.Y.; Zechman, E.M.; Ranjithan, R.S.; Mahinthakumar, G. Enhanced Simulation-Optimization Approach Using Surrogate Modeling for Solving Inverse Problems. *Environ. Forensics* **2012**, *13*, 348–363. [\[CrossRef\]](#)
7. Sreekanth, J.; Datta, B. Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple-realization optimization. *Water Resour. Res.* **2011**, *47*, W04516. [\[CrossRef\]](#)
8. Zhao, Y.; Qu, R.; Xing, Z.; Lu, W. Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms. *Adv. Water Resour.* **2020**, *138*, 103540. [\[CrossRef\]](#)
9. Mirarabi, A.; Nassery, H.R.; Nakhaei, M.; Adamowski, J.; Akbarzadeh, A.H.; Alijani, F. Evaluation of data-driven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems. *Environ. Geol.* **2019**, *78*, 489.1–489.15. [\[CrossRef\]](#)
10. Prakash, O.; Datta, B. Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations. *Environ. Monit. Assess.* **2012**, *185*, 5611–5626. [\[CrossRef\]](#)
11. Zhang, J.; Vrugt, J.A.; Shi, X.; Lin, G.; Wu, L.; Zeng, L. Improving simulation efficiency of MCMC for inverse modeling of hydrologic systems with a Kalman-inspired proposal distribution. *Water Resour. Res.* **2020**, *56*, e2019WR025474. [\[CrossRef\]](#)
12. Datta, B.; Chakrabarty, D.; Dhar, A. Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *J. Hydro-Environ. Res.* **2011**, *5*, 25–36. [\[CrossRef\]](#)
13. Lapworth, D.J.; Baran, N.; Stuart, M.E.; Ward, R.S. Emerging organic contaminants in groundwater: A review of sources, fate and occurrence. *Environ. Pollut.* **2012**, *163*, 287–303. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Zanini, A.; Tanda, M.G.; Woodbury, A.D. Identification of transmissivity fields using a Bayesian strategy and perturbative approach. *Adv. Water Resour.* **2017**, *108*, 69–82. [\[CrossRef\]](#)
15. Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–107. [\[CrossRef\]](#)
16. Sadegh, M.; Vrugt, J. Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC). *Water Resour. Res.* **2015**, *50*, 6767–6787. [\[CrossRef\]](#)
17. Han, Z.; Lu, W.; Fan, Y.; Lin, J.; Yuan, Q. A surrogate-based simulation-optimization approach for coastal aquifer management. *Water Supply* **2020**, *20*, 3404–3418. [\[CrossRef\]](#)
18. Matott, L.S.; Rabideau, A.J. Calibration of complex subsurface reaction models using a surrogate-model approach. *Adv. Water Resour.* **2008**, *31*, 1697–1707. [\[CrossRef\]](#)
19. Zhao, Y.; Lu, W.; Xiao, C. A Kriging surrogate model coupled in simulation-optimization approach for identifying release history of groundwater sources. *J. Contam. Hydrol.* **2016**, *185–186*, 51–60. [\[CrossRef\]](#)
20. Xing, Z.; Qu, R.; Zhao, Y.; Fu, Q.; Ji, Y.; Lu, W. Identifying the Release History of a Groundwater Contaminant Source Based on an Ensemble Surrogate Model. *J. Hydrol.* **2019**, *572*, 501–516. [\[CrossRef\]](#)
21. Zanini, A.; Woodbury, A.D. Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion. *J. Contam. Hydrol.* **2016**, *185–186*, 74–86. [\[CrossRef\]](#)
22. Zhao, Y.; Lu, W.; An, Y. Surrogate model-based simulation-optimization approach for groundwater source identification problems. *Environ. Forensics* **2015**, *16*, 296–303. [\[CrossRef\]](#)
23. Hinton, G.E.; Salakhutdinov, R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2014**, *18*, 1527–1554. [\[CrossRef\]](#)
25. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [\[CrossRef\]](#)
26. Kong, X.; Zheng, F.; E, Z.; Cao, J.; Wang, X. Short-term load forecasting based on deep belief network. *Autom. Electr. Power Syst.* **2018**, *42*, 133–139.
27. Lu, W.; Wang, H.; Li, J. Parallel heuristic search strategy based on a Bayesian approach for simultaneous recognition of contaminant sources and aquifer parameters at DNAPL-contaminated sites. *Environ. Sci. Pollut. Res.* **2020**, *27*, 37134–37148. [\[CrossRef\]](#)
28. Penev, K.; Littlefair, G. Free Search—A comparative analysis. *Inf. Sci.* **2005**, *172*, 173–193. [\[CrossRef\]](#)
29. Robini, M. Theoretically grounded acceleration techniques for simulated annealing. *Intell. Syst. Ref. Libr.* **2013**, *38*, 311–336.
30. Wang, H.; Lu, W.; Chang, Z. An iterative updating heuristic search strategy for groundwater contamination source identification based on an ACP SO-ELM surrogate system. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2153–2172. [\[CrossRef\]](#)
31. Wang, H.; Lu, W. Recognizing groundwater DNAPL contaminant source and aquifer parameters using parallel heuristic search strategy based on Bayesian approach. *Stoch. Environ. Res. Risk Assess.* **2020**, *35*, 813–830. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.