Modeling Earth Systems and Environment (2024) 10:751–765
https://doi.org/10.1007/s40808-023-01797-0

ORIGINAL ARTICLE

ANN-SFLA based parameter estimation method for an unsaturated–saturated simulation model

Mamata Das[1] · Rajib Kumar Bhattacharjya[1] · Suresh A. Kartha[1]

Received: 15 January 2023 / Accepted: 26 May 2023 / Published online: 11 June 2023

Abstract
A numerical simulation of groundwater aquifers in saturated and unsaturated zones requires knowledge of the hydraulic parameters that govern the flow. However, these parameters may not be readily available and need to be estimated. The parameters can be estimated by using an inverse optimization model, where the model minimizes the error function between the observed and simulated hydraulic heads. Since parameter estimation is a non-convex problem, multiple solutions satisfy the imposed constraints and thus result in the non-uniqueness of solutions. On the other hand, due to the nonlinearity in the numerical flow models, high computational times are required for the simulations when coupled with the optimization model. This paper presents a novel technique to estimate the unsaturated and saturated flow parameters by employing the meta-heuristic Shuffled Frog Leaping Algorithm (SFLA). In addition, Artificial Neural Network (ANN) is combined uniquely in the simulations to reduce the computational time in predicting the hydraulic heads. The ANN-SFLA model successfully estimated the unsaturated and saturated parameters of a hypothetical three-dimensional groundwater aquifer simulation model. The efficacy of the proposed model is reflected by its high efficiency in computational time and performance prediction. In addition, a global sensitivity analysis is performed using variance decomposition technique to determine the relative importance of each flow parameter.

Keywords Non-convex · Hydraulic parameters · Constraints · Groundwater

Introduction

A significant portion of the precipitation that falls onto the earth's surface enters the subsurface through infiltration. The infiltrated water passes through the unsaturated zone before reaching the groundwater table. The movement of water through the unsaturated–saturated zone is highly complex since the moisture content of the soil changes within this zone. In order to study how water moves from the ground surface to the groundwater aquifers, it is necessary to develop a model that replicates the flow phenomena in unsaturated–saturated zones. Using numerical models to study groundwater flow, solute transport, and groundwater management has become essential over the past few decades. With the increased use of groundwater for irrigation and domestic purposes, the importance of such models has increased drastically. As such, it is necessary to incorporate the soil and hydraulic parameters to develop an accurate numerical simulation model along with natural boundary conditions at the field scale. The hydraulic parameters are those that define the relationship between hydraulic conductivity (K), volumetric water content (Θ), and pressure head (h). Such parameters are measured or estimated based on different experimental and empirical relations. It is, however, difficult to measure some of these parameters at the desired field or laboratory scale. In practice, if the hydraulic properties of the aquifer are unknown, these must be estimated using hydrogeologic data by the model calibration process.

The model calibration process has recently gained significant attention (McLaughlin and Townley 1996). However, hydraulic parameter identification or inverse problem involves using a mathematical or numerical model to identify hydraulic parameters from field or laboratory

✉ Mamata Das
mamatadas4@iitg.ac.in

Rajib Kumar Bhattacharjya
rkbc@iitg.ac.in

Suresh A. Kartha
kartha@iitg.ac.in

[1] Department of Civil Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

observations (Hyun and Lee 1998). In the subsequent step, the soil and hydraulic parameters are estimated by clubbing the numerical and optimization models. The parameters are estimated by satisfying the objective function, which minimizes the error function between the observed and predicted hydraulic heads (Dane and Hruska 1983; Kool et al. 1987; Yeh 1986). The observed hydraulic head was obtained from the field study, while the simulated head was obtained by running the numerical simulation model. The optimization model uses various algorithms to provide new solutions to attain the objective function. Estimating parameters in unsaturated flow studies have traditionally been carried out using gradient-based classical optimization methods (Eching and Hopmans1993; Kool and Parker 1988; Šimůnek and Van Genuchten 1996). However, due to the nonlinear behavior of the response function, they sometimes fail to find the optimal global solution to the problem. Woodbury and Ulrych (2000); Woodbury and Rubin (2000) applied a full-Bayesian approach using both Bayesian and maximum entropy to estimate transmissivity from the hydrostatic head and transmissivity measurements viewpoints. A simulation–optimization-based model was developed using a meshless local Petrov–Galerkin method and particle swarm algorithm to estimate saturated flow parameters (Swathi and Eldho 2018). This model predicted only one or two parameters at a time among hydraulic conductivity or transmissivity and specific storage. The model, however, could not provide conclusions about its suitability for different groundwater systems. Another model was developed to estimate the storage coefficient, transmissivity, and leakage factor by using pumping test data in one-dimensional confined and leaky confined aquifers (Ayvaz and Gurarslan 2019).In many groundwater studies, stochastic optimization techniques, such as Pattern Search, Genetic Algorithms, or Simulated Annealing, have been used to reach the optimal global solution. These models were developed to estimate parameters in groundwater aquifers (Huang et al. 2008; Şahin 2018; Samuel and Jha 2003).All such models independently estimated the soil and hydraulic parameters for the unsaturated zone or the saturated zone. Thus, an effective parameter estimation model is yet to develop to estimate the unsaturated and saturated flow parameter together in a single model.

This study proposes a methodology to estimate the unsaturated and saturated flow parameters together in a single inverse optimization model. As such, the numerical simulation model needs to be developed by considering both the unsaturated and saturated zone. Due to the presence of an unsaturated zone in the study domain, the groundwater flow model becomes highly nonlinear. Thus, it becomes computationally expensive to combine this simulation model with the optimization algorithm. This is because the simulation model will be called as many times as the number of population sizes, leading to time-consuming computations. In order to overcome this limitation, an alternate simulator should be used in conjunction with the optimization model to estimate the flow parameters. In the field of civil and environmental engineering, artificial neural networks (ANNs) have shown successful results in mapping complex nonlinear relations (Flood and Kartam 1994). The groundwater flow model developed by Balkhair (2002) could estimate transmission coefficients and storage coefficients using trained neural networks. Also, as a result of back propagation, training of multilayer perceptrons, complex relationships, such as rainfall-runoff processes, have been successfully modeled in hydrology and water resources (Smith and Eli 1995), and water quality parameters have also been forecasted (Maier and Dandy 1996).

There are many problems associated with parameter estimation models, including nonlinearity, non-uniqueness, and instability (Carrera and Neuman 1986). Non-identifiability of solutions occurs when a solution cannot be found with the proposed technique. Whereas multiple solutions that satisfy imposed constraints are indicative of the problem of non-uniqueness of solutions. Such types of problems can be solved using meta-heuristic algorithms, and those algorithms are effective for solving inverse optimization problems as well. One such efficient meta-heuristic algorithm is the Shuffled Frog Leaping Algorithm (SFLA). This algorithm solves highly nonlinear non-convex problems using a population-based metaheuristic and a memetic approach. It was designed the way that an army of frogs searched for food in a swamp. For a better search, they leap onto the nearest possible rock and communicate with each other. Consequently, they develop a strategy that allows them to gather the most food in the least amount of time. An optimization algorithm designed to replicate this process is called the Shuffled Frog Leaping Algorithm (SFLA). A combination of Particle Swarm Optimization (PSO) and Shuffled Complex Evolution (SCE) are the principles behind this algorithm. This algorithm is relatively very fast compared to the traditional meta-heuristic evolutionary genetic algorithm (Gandhi and Bhattacharjya 2020).

All the optimization models available in the literature estimated the flow parameters for unsaturated and saturated zone separately, whereas, in real field problems, there may be situations where both the unsaturated and saturated flow parameters have to be considered together in modeling. Thus, to overcome this limitation, this paper proposes an effective parameter estimation model to estimate both the unsaturated and saturated parameters together using Shuffled Frog Leaping Algorithm in conjunction with the simulation model. However, coupling the flow simulation model with the optimization algorithm for the entire computational domain requires more time. As such, an alternate simulator developed by using Artificial Neural Networks (ANN) that replicates the groundwater simulation model is used to reduce the computational time. In addition, it was found that the input values significantly affect the model's outputs. Therefore, Sobol's global sensitivity analysis based on variance decomposition is used to determine the most relevant flow parameters associated with the groundwater flow model.

Materials and methods

Estimation is performed by minimizing the error function between the observed and simulated hydraulic heads. The observed hydraulic head is obtained from the field study, and the simulated head is obtained from the groundwater simulation model. Initially, the numerical simulation model is developed to study the groundwater flow considering both the unsaturated and saturated zone. The governing equation that is used to develop the groundwater flow model is discussed below.

Flow equation

The three-dimensional unsaturated and saturated groundwater flow equation is the modified form of Richards' equation given by Dogan and Motz (2005).

$$\frac{\partial}{\partial x}\left(K_{xx}(h)\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}(h)\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}(h)\frac{\partial h}{\partial z}+K_{zz}(h)\right)+q_e = C(h)+S_wS_s\frac{\partial h}{\partial t} \tag{1}$$

where, θ is the water content; h is the pressure head [L]; K_{xx}, K_{yy}, and K_{zz} are the hydraulic conductivity along x, y, and z directions, considering the coordinate system as the principal directions of the hydraulic conductivity tensor [L T^{-1}]; q_e represents pumping or recharge rate [L^1 T^{-1}]; $C(h)$ is the specific moisture capacity (L^{-1}), S_w is the saturation ratio, S_s is the specific storage [L^{-1}]; and t represents the time.

Constitutive relationship:

From the above equations, it is observed that the specific moisture content $C(h)$, hydraulic conductivity $K(h)$, and $\Theta(h)$ are nonlinear, which makes the equation more complex. To overcome this nonlinearity, the model uses the constitutive relationship given by Van Genuchten and Nielsen (1985).

Constitutive relation for $K(h)$:

For $h<0$

$$K_r=\frac{K(h)}{K_s}=(1+\beta)^{-\frac{5}{2}(1-1/n)}\left[(1+\beta)^{(1-1/n)}-\beta^{(1-1/n)}\right]^2 \tag{2}$$

For $h \geq 0$

$$K_r=\frac{K(h)}{K_s}=1 \tag{3}$$

Constitutive relation for $C(h)$:

When $h \leq h_0$

$$C(h)=\frac{(n-1)\left(\theta_s-\theta_r\right)|h|^{n-1}}{\left|h_a\right|^n(1+\beta)^{2-1/n}} \tag{4}$$

When $h > h_0$

$$C(h)=0 \tag{5}$$

Constitutive relation for $\theta(h)$:

When $h \leq h_0$

$$\theta(h)=\theta_r+\left(\theta_s-\theta_r\right)(1+\beta)^{(1/n-1)} \tag{6}$$

$h > h_0$

$$\theta(h)=\theta_r+\left(\theta_s-\theta_r\right)\left(1+\beta_0\right)^{1/n-1}+S_S\left(h-h_0\right) \tag{7}$$

where, $\beta=\left|\frac{h}{h_a}\right|^n$, h_a is the air entry pressure [L], n is the fitting parameter in the moisture retention curve, or,$\beta_0=\left|\frac{h_0}{h_a}\right|^n$, h_0 is a parameter depending upon the Specific storage (S_S). When, $h \geq h_a$, the Eq. (1) solves for the saturated flow condition, i.e., $C(h)=0$, $K(h)=K_s$, $S_w=1$, and when h < ha, then the Eq. (1) solves for the unsaturated flow condition. Then $C(h) \neq 0$, $K(h)$ is the function of pressure head, $S_w<1$ and $Ss=0$.

This study uses the block-centered finite difference form to solve Eq. (1). In order to develop the model, the sum of inflows into and out of a unit volume of aquifer must be equal to the rate of change in the volume of storage within the cell. Since the modified form of Richard's equation is highly nonlinear, Picard iteration method is adopted at each time step to overcome the nonlinearity. Using the numerical scheme and applying the necessary boundary condition, a linear system of equations is developed at every modified Picard iteration level. This set of equations can be solved using the preconditioned conjugate gradient method (PCGM), which is more memory-efficient than other iterative methods and has a faster convergence rate (Celia et al. 1990; Clement et al. 1994).

Development of ANN model

The artificial neural network (ANN) model is a very effective and popular substitute for numerical aquifer simulations (Afzaal et al. 2020; Chang and Zhang 2019; Mohanty et al. 2013; Shen et al. 2018; Zhang et al. 2018, 2020). In this proposed methodology, the ANN model acts as the surrogate

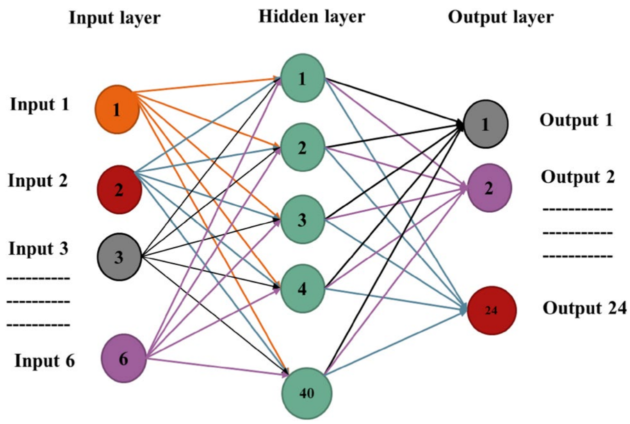


Fig. 1 ANN model network

model of the groundwater flow model. A three-dimensional unsaturated–saturated groundwater flow model developed using Eq. (1) is used to generate data for training the ANN model. To develop the ANN model, the input parameters are the flow parameters (**θs**, **θr**, **α**, $\mathbf{K_s}$, **n**, and $\mathbf{S_s}$) to be estimated, and the output of the ANN model are the hydraulic head at different observation well location for different time steps. The developed ANN model can further predict the hydraulic heads without evoking the numerical simulation model, thereby reducing the computational time considerably. In this study, six observation wells are considered, and as such, six ANN models are developed. A feed-forward neural network is used to generate the ANN pattern (Fig. 1), which features one hidden layer with 40 neurons and 1000 input–output patterns, which are generated using the groundwater simulation model. In total, 60% of the generated data is utilized for training ANN models, and 40% is used for testing and validating the ANN model. Training of the ANN model is carried out using the Levenberg–Marquardt (LM) algorithm. A unipolar sigmoidal transfer function and a purely linear transfer function are used for the hidden layer and the output layer of the network.

Parameter estimation model

In this study, the unsaturated hydraulic properties—water content (Θ), hydraulic conductivity (K), and pressure head (h) are related using Van Genuchten and Nielsen's (1985) constitutive relationship. The five hydraulic parameters ($\boldsymbol{\theta_s}$, $\boldsymbol{\theta_r}$, $\boldsymbol{\alpha}$, $\mathbf{K_s}$ and **n**) need to be estimated in order to get the constitutive relations. On the other hand, the specific storage ($\mathbf{S_s}$) is an important parameter in a saturated zone that also needs to be estimated. Henceforth, this optimization model considers six decision variables ($\boldsymbol{\theta_s}$, $\boldsymbol{\theta_r}$, $\boldsymbol{\alpha}$, $\mathbf{K_s}$, **n** and $\mathbf{S_s}$) that need to be estimated. In this inverse optimization technique, the numerical groundwater simulation model initially uses the candidate solutions generated by the optimization algorithm as input parameters. After that, using this simulation model, the spatial and temporal hydraulic head is generated and matched with the measured hydraulic head at the different observation well. In the optimization model, the objective function value is determined by the difference between the simulated and observed hydraulic heads. The candidate solution is modified based on the objective function value, and the process is repeated until the optimal solution is obtained. The objective function used to estimate all flow parameters is given by Eq. (8). However, this combination took around one day, 5 h, 45 min, and 10 s to estimate all the parameters, which is very time-consuming. To overcome this disadvantage, the numerical simulation model is replaced with an alternate simulator using an Artificial Neural Network (ANN). This ANN model is linked externally with the optimization model, and the whole methodology is represented in a flowchart, as given in Fig. 2.

$$\text{Minimize} \quad f_x = \sum_{j=1}^{M} \sum_{i=1}^{N} \left| OH_i^j - SH_i^j \right| \tag{8}$$

$$\text{Subject to} \quad H = f\left(\theta_s, \theta_r, \alpha, n, K_s, S_s\right)$$

$$\theta_s^{\min} \le \theta_s \le \theta_s^{\max}$$

$$\theta_r^{\min} \le \theta_r \le \theta_r^{\max}$$

$$\alpha^{\min} \le \alpha \le \alpha^{\max}$$

$$n^{\min} \le n \le n^{\max}$$

$$K_s^{\min} \le K_s \le K_s^{\max}$$

$$S_s^{\min} \le S_s \le S_s^{\max}$$

where $\boldsymbol{f_x}$ is the objective function for the present optimization model with $\boldsymbol{x}$ number of parameters; OH_i^j is the observed hydraulic head at $\boldsymbol{i}$th time step for $\boldsymbol{j}$th well location; SH_i^j is the simulated hydraulic head at $\boldsymbol{i}$th time step for $\boldsymbol{j}$th well location obtained from ANN model or numerical simulation model; $\boldsymbol{M}$ is the total number of observation wells and; N is the total number of time steps; $\theta_r^{\min}$ and $\theta_r^{\max}$ are the lower and upper bound of θ_r; $\theta_s^{\min}$ and $\theta_s^{\max}$ are the lower and upper bound of θ_s;$\alpha^{\min}$ and $\alpha^{\max}$ are the lower and upper bound of α; $K_S^{\min}$ and $K_S^{\max}$ are the lower and upper bound of K_s;$n^{\min}$ and $n^{\max}$ are the lower and upper bound of n and $S_s^{\min}$ and $S_s^{\max}$ are the lower and upper bound of S_s.

Shuffled frog leaping algorithm

In this study, Shuffled Frog Leaping Algorithm (SFLA) is used as the optimization algorithm for estimating the flow parameters. SFLA is a metaheuristic optimization algorithm that solves nonlinear non-convex optimization problems. As the flow parameter estimation is a non-convex problem having multiple local optima, the SFLA algorithm is suitably employed. The model developed for the ANN-SFLA study, as depicted in Fig. 2, is coded in MATLAB environment. The process of Shuffled Frog Leaping Algorithm (SFLA) is shown in Fig. 3. To begin

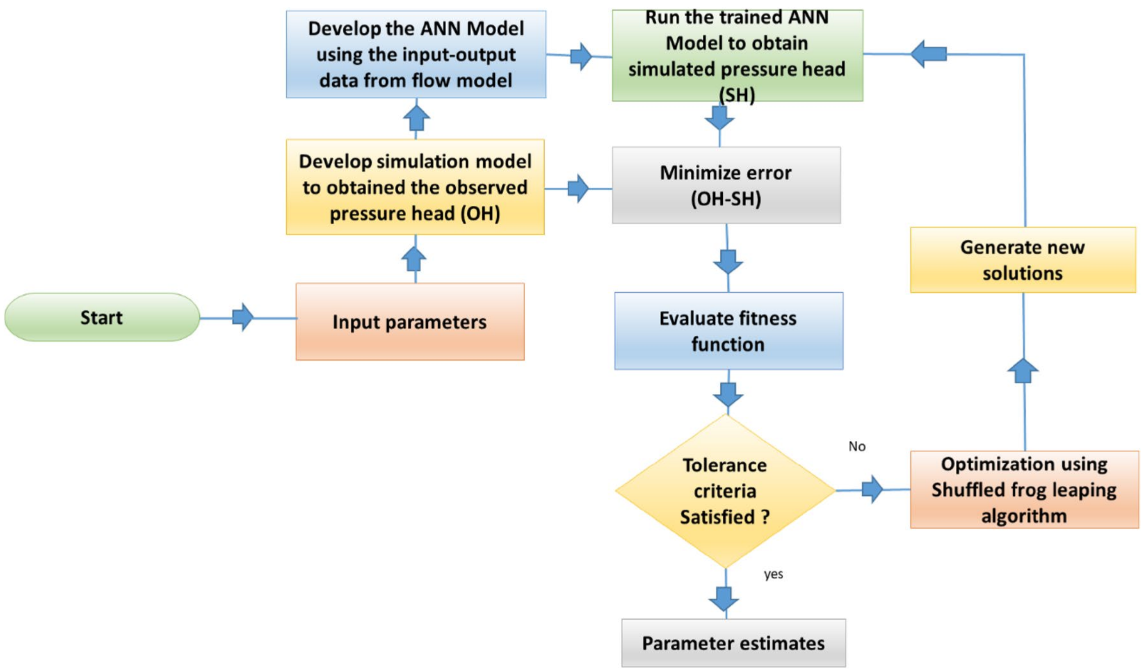


Fig. 2 Flow chart showing the ANN-SFLA based parameter estimation model

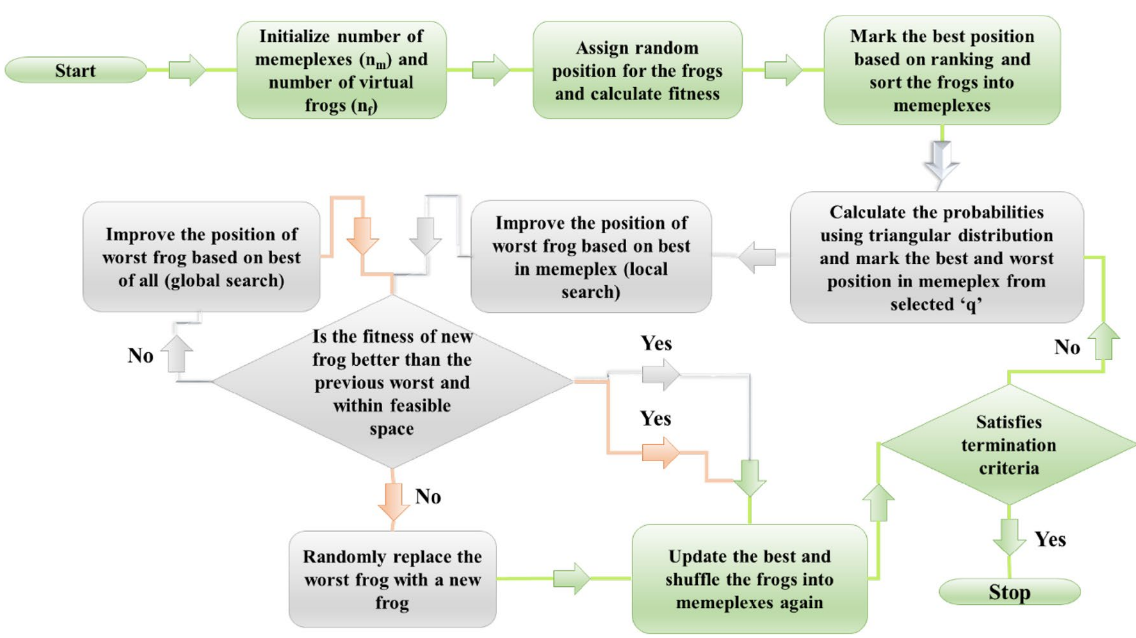


Fig. 3 Flowchart showing the process of shuffled frog leaping algorithm (SFLA)

the problem, the first step is to select the number of memeplexes (n_m) and virtual frogs within each memeplex (n_f). This gives the total number of frogs as $n_m \times n_f$. The algorithm continues by assigning a random position to all the frogs and calculating the corresponding fitness. The best frog is then marked as the global best, and then the frogs are sorted into memeplexes. This portion of the flowchart is presented in green colour. The next portion is local evolution in each memeplex, represented in grey colour. The frogs are distributed using the triangular distribution, as shown in Eq. (9), where p_i represents the probability for triangular distribution, and from them, q frogs are selected.

$$p_j = \frac{2(n_f+1-j)}{n_f(n_f+1)}, \quad \text{where } j = 1 \ldots n_f, \tag{9}$$

The best and the worst frog are then marked. The position of the worst frog is improved by choosing an appropriate step size based on the best of the memeplex. Then the condition of whether this frog is better than the previous worst and within the feasible space is checked. If the condition is satisfied, then the next step is to update the memeplex and shuffle the frogs. If the condition is not satisfied, then the worst position is improved based on the global best. This set of conditions is represented with blue connectors and arrows with the decision matrix in the flowchart. The improved position is again checked with the same condition as the previous one. If the condition is satisfied, then updating the memeplex and shuffling is continued. If the condition is not satisfied, the new position is improved randomly, and then the memeplex is updated. The flowchart represents this set of conditions with red connectors and arrows. The termination criteria are then checked after updating the global best, and reshuffling into the memeplexes is done. The global optimal solution is reached if the termination criteria are satisfied. If the termination criteria are not satisfied, then the algorithm resumes the step of evolution from each memeplex.

Results and discussion

Validation of the numerical flow model with one-dimensional infiltration problem

To validate the groundwater flow model considering both the unsaturated and saturated zone, a one-dimensional groundwater flow model is selected with transient infiltration towards the groundwater table (Paniconi et al. 1991). The model is simulated for 32 h with the same boundary conditions and input parameters. Figure 4 compares the solutions obtained from the numerical simulation model developed using the code written in MATLAB with the solution obtained by Paniconi et al. (1991). The scatter plots (Fig. 4b) correlate the pressure head (m) obtained from Paniconi et al. (1991) and the numerical simulation model. The regression coefficient is 0.9981, which ascertains that the numerical simulation groundwater flow model could provide an accurate solution as obtained by Paniconi et al. (1991).

Three-dimensional hypothetical groundwater flow model considering both unsaturated and saturated zone

In this study, a three-dimensional hypothetical numerical flow model is developed for a homogeneous medium considering both unsaturated–saturated zones. This hypothetical groundwater flow model is developed by solving Eq. (1) using MATLAB. The graphical representation of the groundwater flow model used in this study showing the

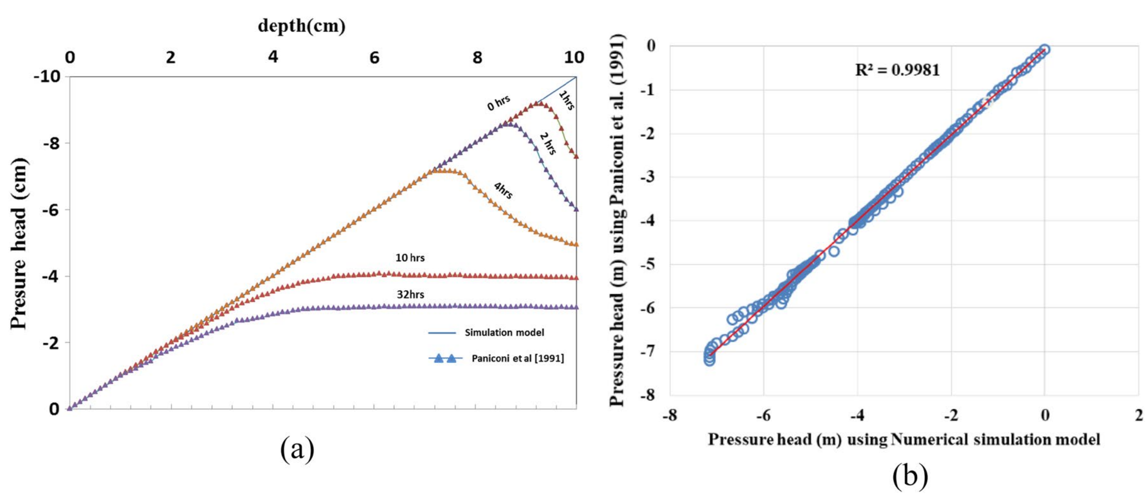


Fig. 4 Comparison of the pressure head distribution obtained from Paniconi et al. (1991) and numerical simulation model

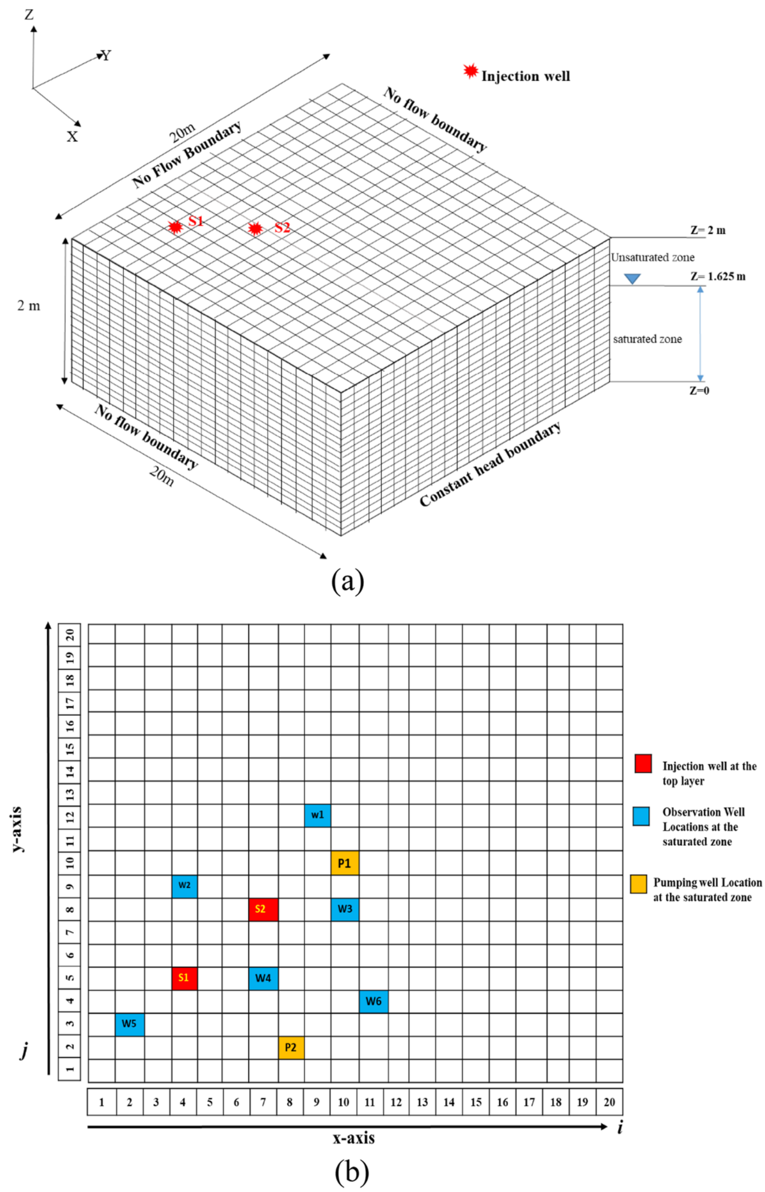


Fig. 5 Hypothetical study area showing **a** dimensions and boundary conditions, **b** injection wells, observation wells and pumping well location

location of the injection well, observation well, and pumping well is given in Fig. 5. The parameters used to develop this model are listed in Table 1. Further, it is assumed that two injection wells are applied at the ground surface, and two pumping wells are located to pump out water from the saturated zone.

The hydraulic head obtained at time steps 10 h, and 15 h is presented (Fig. 6) for both the unsaturated and saturated zones. For the unsaturated zone, the contour plots of the hydraulic head are shown at the top layer of the flow domain, whereas for the saturated zone, the hydraulic head is shown at a depth of 1.625 m from the datum (Initial position of the water table). The x-axis and y-axis are the dimensions of the flow domain in the XY plane. A total of $(20 \times 20 \times 40)$ grids are considered for the analysis. However, the solutions are refined for grid size (200×200) along the XY plane by interpolating and smoothening the results of the head obtained. Figure 6 shows an increase in hydraulic head in the location of injection wells, whereas a cone of depletion is seen in the pumping well location.

Performance of the ANN models

The numerical simulation model developed for the three dimensional groundwater aquifer is used to generate 1000 input–output data to develop the ANN model. The model uses six parameters as input data and the hydraulic head at different time steps as the output. Using the generated data, the ANN model is trained, tested, and validated. The

Table 1 Parameters used to develop the groundwater flow model as considered by Dogan and Motz (2005)

Sl. no.	Parameters	Value
1	Flow domain	$20 \times 20 \times 2$ (m^3)
2	Saturated Hydraulic conductivity (Ks)	0.35 m h^{-1}
3	Residual moisture content	0.01
4	Saturated moisture content	0.30
5	Air entry pressure ($h_a = 1/\alpha$)	1/3.3 m
6	Van Genuchten parameter(n)	4.1
7	Specific storage (Ss)	0.001 (m^{-1}) (assumed)
8	Injection wells (S1) active for 10 h	0.37 m h^{-1}
9	Injection wells (S2) active for 10 h	0.35 m h^{-1}
10	Pumping rate(P) active for 15 h	−0.25 m h^{-1}
11	Bottom, front, back and left boundary	Impervious—no flow condition
12	Top boundary	Prescribed flux at top layer as shown in Fig. 5
13	Right boundary	Constant head is maintained upto 1.625 m and the remaining is no flow
14	Initial pressure head	Hydrostatic equilibrium with horizontal water table at 1.625 m (i.e., $h + z = 1.625$ m)
14	Grid discretization	Grid size of $dx = 1$ m, $dy = 1$, $dz = 0.05$ m
15	Time increment	0.5 time step is maintained
16	Maximum simulation time	24 h

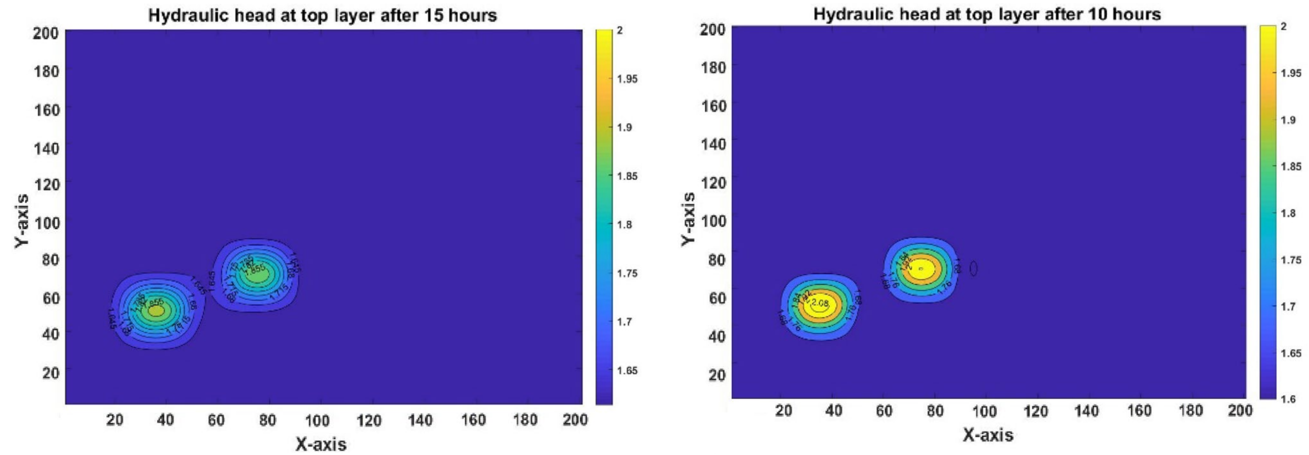
performance of the developed ANN simulator is shown in Figs. 7 and 8. For each observation well, scatter plots at twenty-four-time steps are plotted. The groundwater numerical flow model provides the observed hydraulic head (OH), while the ANN model provides the simulated hydraulic head (SH). Among all observation wells, the best and the worst coefficient of correlation (R^2) values are 0.9999 and 0.9375, respectively. It can be seen that R^2 is very close to 1, which implies a strong correlation between the actual hydraulic head and the predicted pressure head. Figure 8 shows the performance of all 6 ANN models for training, testing, and validating the data using Mean square error. This graph shows MSE for training, testing, and validating batches as it converges toward the best with each Epoch. The calculated error terms are found to be very negligible as it ranges from 2.3×10^{-2} to 4.07×10^{-6}. Therefore, we can conclude that the developed ANN model can serve as an approximate simulator for simulating the hydraulic head for the proposed study area.

Performance of the parameter estimation model

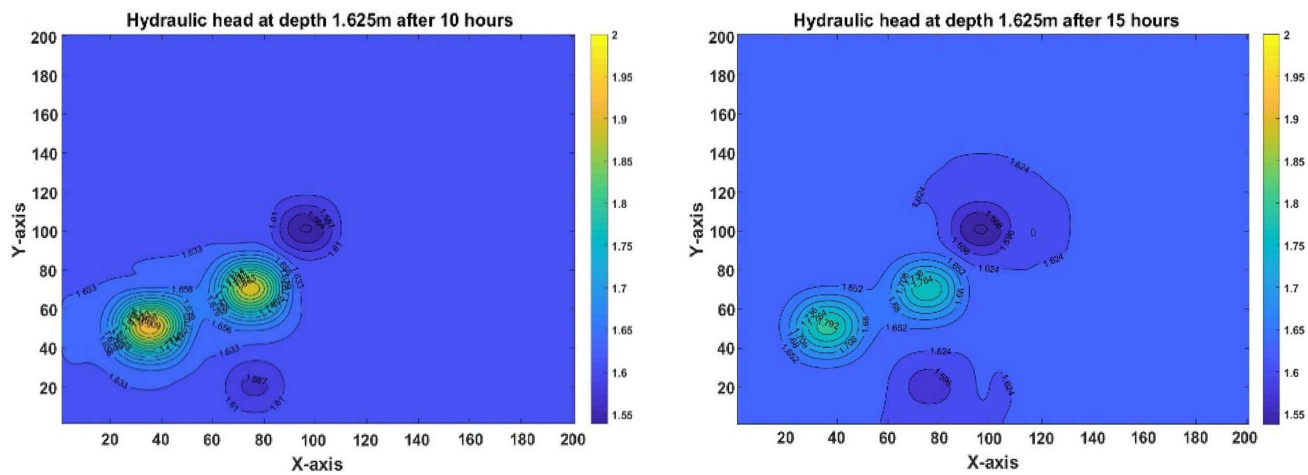
ANN-shuffled frog leaping algorithm

The ANN model was developed for six observation wells and is coupled with the optimization model (SFLA) to minimize the objective function with six input variables (i.e., $\boldsymbol{\theta_s}$, $\boldsymbol{\theta_r}$, $\boldsymbol{\alpha}$, $\mathbf{K_s}$, $\mathbf{n}$, and $\mathbf{S_s}$) as the decision variable. Since this study considers a hypothetical problem, the observed hydraulic heads (OH) for the optimization's objective function are the values taken from the numerical simulation model, and simulated heads (SH) for the same objective function are taken from ANN models. The lower and upper limits for these parameters are decided based on the previous experimental evidence, as listed in Table 2. The number of memeplexes is selected as 7, and the number of virtual frogs is chosen as the maximum number of variables plus one and is equal to 7. Therefore, 7 virtual frogs for each memeplex were selected, comprising a total of 49 frogs. The maximum step size was taken as 1 unit. The maximum number of evolutions in each memeplex is 6, the step length coefficient is 2, and the maximum number of iterations is restricted to 200. The predicted hydraulic parameters using the ANN-SFLA-based parameter estimation model are listed in Table 2.

Table 2 provides the predicted values obtained from the SFLA-ANN-based parameter estimation model. Here, the unsaturated and saturated parameters are estimated in a single model. The relative efficiency of the ANN-SFLA model in predicting the flow parameters is also checked by evaluating the relative error concerning the actual flow parameters. The model observations indicate that the model could predict all the parameters up to a fair degree of accuracy. Considering the relative error among the actual and predicted values, it ranges from 0.03 to 1.00%. But these values are subsequently low and considered within the acceptable accuracy range. The model converges toward the optimal solution when the objective function value of 6.3084E-05 is reached. It is further compared with another established ANN-Genetic Algorithm-based parameter estimation tool to illustrate the performance of our ANN-SFLA model.



(a) Hydraulic head (m) in unsaturated zone



(b) Hydraulic head (m) in saturated zone

Fig. 6 Contour plots showing the hydraulic head (m) at different time steps in **a** unsaturated zone and **b** saturated zone

ANN-genetic algorithm

The genetic algorithm (GA) available in the MATLAB toolbox compares the results with SFLA. Genetic algorithms search for optimal solutions through natural selection and genetic evolution (Abdel and El-Hadi 2009; Cavazzuti 2012; Holland 1992). Due to the non-gradient-based search method of GA, it typically produces nearly global optimal solutions instead of true solutions. Thus, the solution obtained by using ANN-GA is presented in Table 3. In this study, GA uses a population size of 50, a maximum generation of 100, a function tolerance of 1×10^{-5}, and a crossover probability of 0.8. Mutation functions are constraint-dependent, and the number of stall generations is 60.

When the relative error is calculated, it is observed that ANN-GA could correctly predict three parameters while the remaining three parameters showed errors. For a number of trials considering both the ANN-SFLA and ANN-Genetic algorithms, a number of solutions are generated to verify the accuracy of the proposed algorithms. Figure 9 shows a box plot representing the estimated parameters after 20 trials from both models. The plots show that the average value of α, Ks, and n for both models is very close to the optimal solution. In the ANN-GA model, the estimated value of θ_s, θ_r, and S_S varies with a wide range of values as compared to ANN-SFLA. The median value obtained for θ_s is 0.34, θ_r is 0.0135, and $S_S = 0.0018$ (m^{-1}) using the ANN-GA model, while for the ANN-SFLA model, the median values for θ_s are 0.315, θ_r is 0.011, and $S_S = 0.0012$ (m^{-1}). From the investigation, it is clear that the solution obtained after 20 trials shows better performance in the ANN-SFLA model than in the ANN-GA model.

The variation of the objective function with the iteration for the Genetic Algorithm (GA) and Shuffled Frog Leaping Algorithm (SFLA) is plotted in Fig. 10 to study the reason for this observation. A total number of 200 generations are

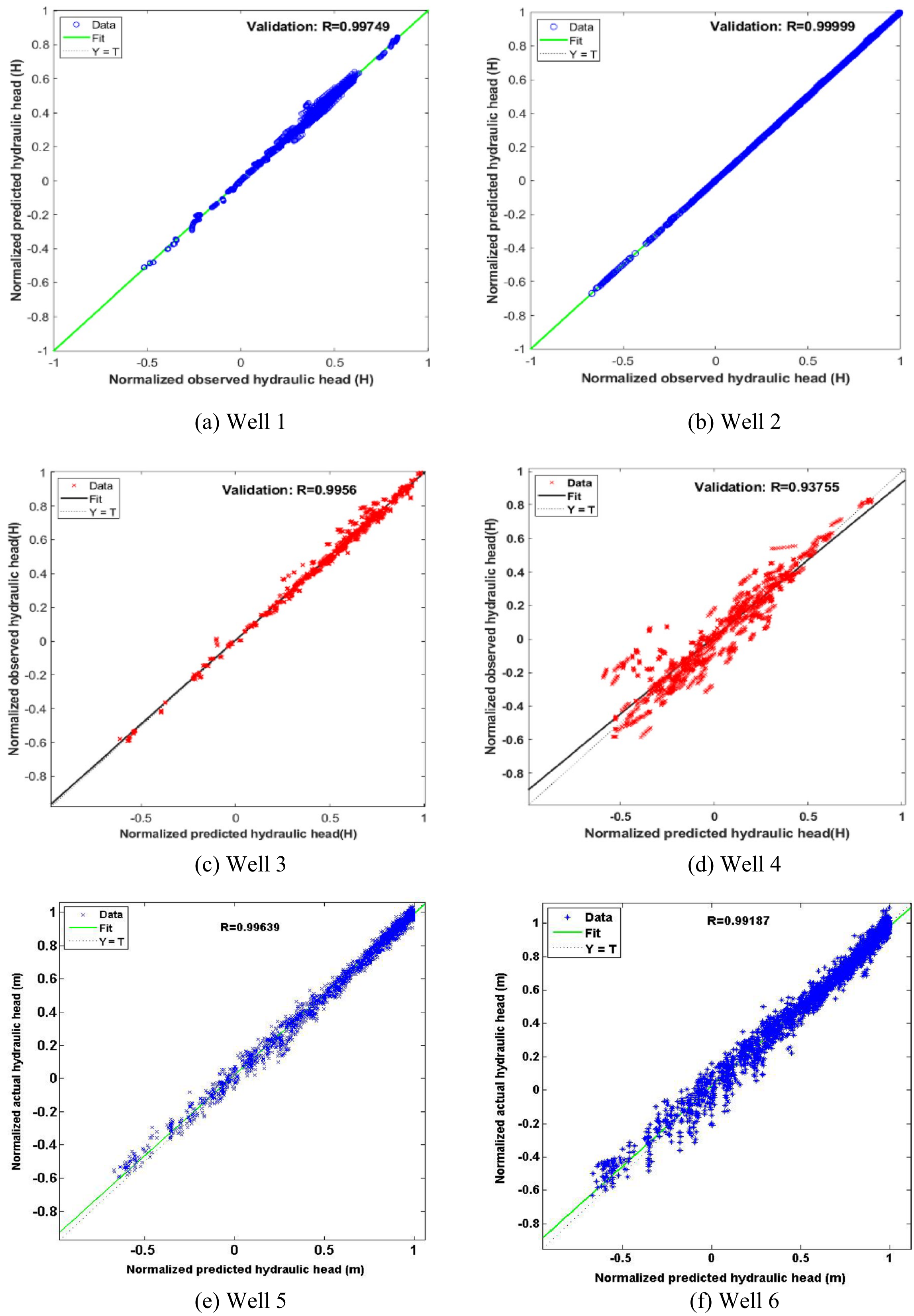


Fig. 7 Regression plot of the numerical simulation model with the six ANN models

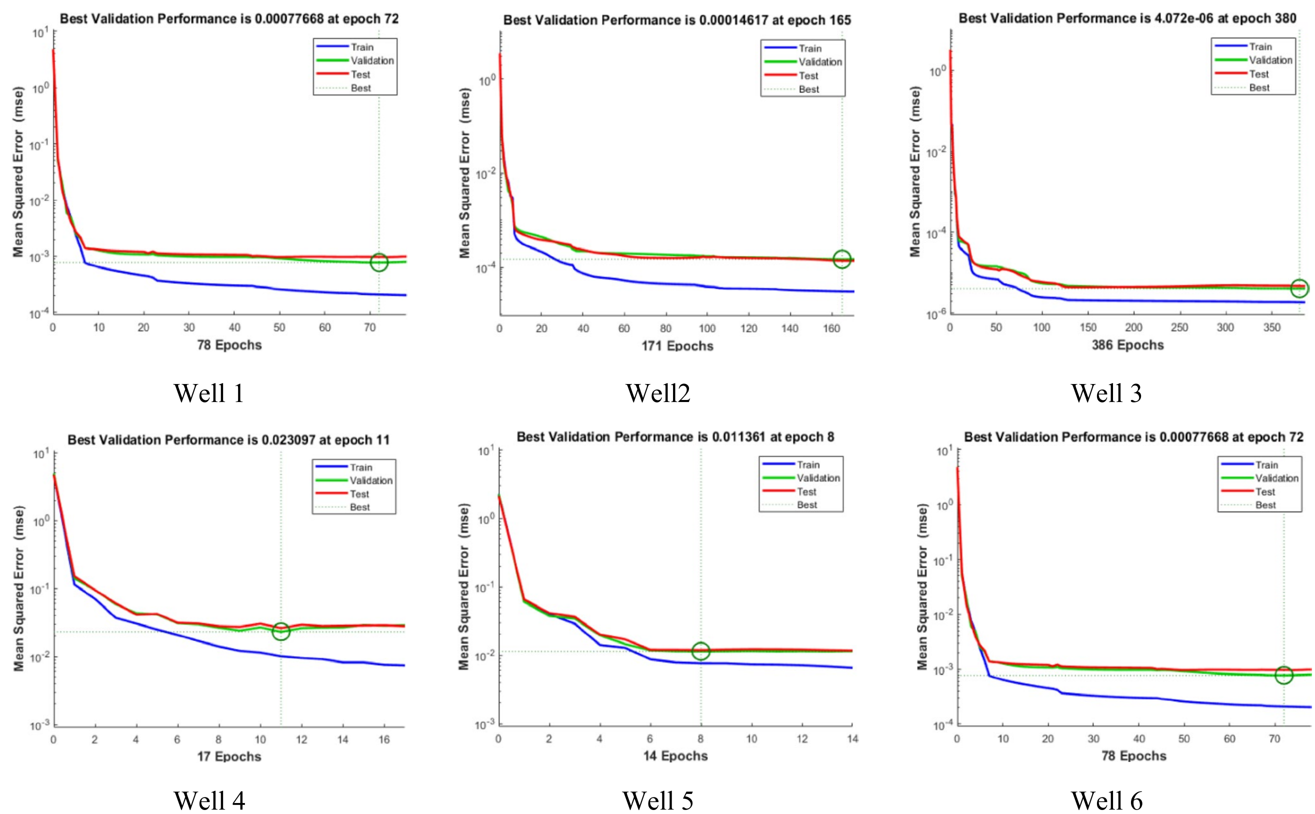


Fig. 8 Training, testing, and validation plots of the six ANN models

taken for both GA and SFLA. The population size is also relatively similar – 50 for GA and 49 for SFLA. Therefore, the total number of function evaluations is almost identical for both algorithms. As observed, SFLA gets convergence faster and yields better results (Fig. 10). This may be because the memetic evolution is faster and consists of different sets of evolution happening at the same time. Genetic evolution consists of the population (a set of solutions) and evolves altogether. On the other hand, memetic evolution follows a different approach where the population is divided into different memeplexes, and each memeplex evolves independently on a population basis. The population is mixed again to communicate so that the global best is updated, and reshuffling is done again to continue the evolution into the memeplexes. It may be noted that the problem considered in the study has multiple local optimal solutions. As such, a large population size must be taken to obtain the optimal global solution. The population size of 50 is considered in GA just to compare the result with SFLA. The GA may yield a better solution if we increase the population size by more than 50.

Sobol's sensitivity analysis

The Sobol's global sensitivity method is used to analyze the most influential flow parameters in the unsaturated–saturated flow model. With the use of variance decomposition, one can determine the effect of each parameter on the output, and the interactions between them, based on a large sample of input variables.

Using this method, we can deal with nonlinear and non-monotonic models due to its variance decomposition approach. Using functional representations, the models can be expressed as follows:

$$y = f(x) = f\left(x_1, x_2, \ldots, x_p\right) \tag{10}$$

where y is the goodness of fit metric for the model output, and x is the set of input parameters: (x_1, x_2,....,x_p). Sobol's method is a variance decomposition approach. D(y) represents the total variance of the function f. Depending on individual parameters and interactions, D(y) is subdivided into different components.

$$D(y) = \sum_i D_i + \sum_{i<j} D_{ij} + \sum_{i<j<k} D_{ijk} + \cdots + D_{12\ldots p} \tag{11}$$

Table 2 Optimization results 1: ANN-shuffled frog leaping algorithm

Parameters	Θ_s	Θ_r	α	n	K_s	S_s	Function value
Lower limit	0.1	0.009	1	1	0.1	0.0001	
Upper limit	0.5	0.1	10	10	0.5	0.01	
Actual	0.30	0.01	3.3	4.1	0.35	0.001	6.30843e-05
Predicted	0.2992	0.01003	3.3012	4.0812	0.347	0.00101	
Relative error (%)	0.2633	0.3	0.0392	0.4568	0.8571	1.00	

Table 3 Optimization results II: ANN-genetic algorithm

Parameters	Θ_s	Θ_r	α	n	K_s	Ss	Function value
Lower limit	0.1	0.009	1	1	0.1	0.0001	
Upper limit	0.5	0.1	10	10	0.5	0.01	
Actual	0.30	0.012	3.3	4.1	0.35	0.001	3.31e-03
Predicted	0.3430	0.0130	3.2574	4.0284	0.3436	0.0019	
Relative error (%)	14.3527	8.8666	1.2879	1.7441	1.8011	96.100	

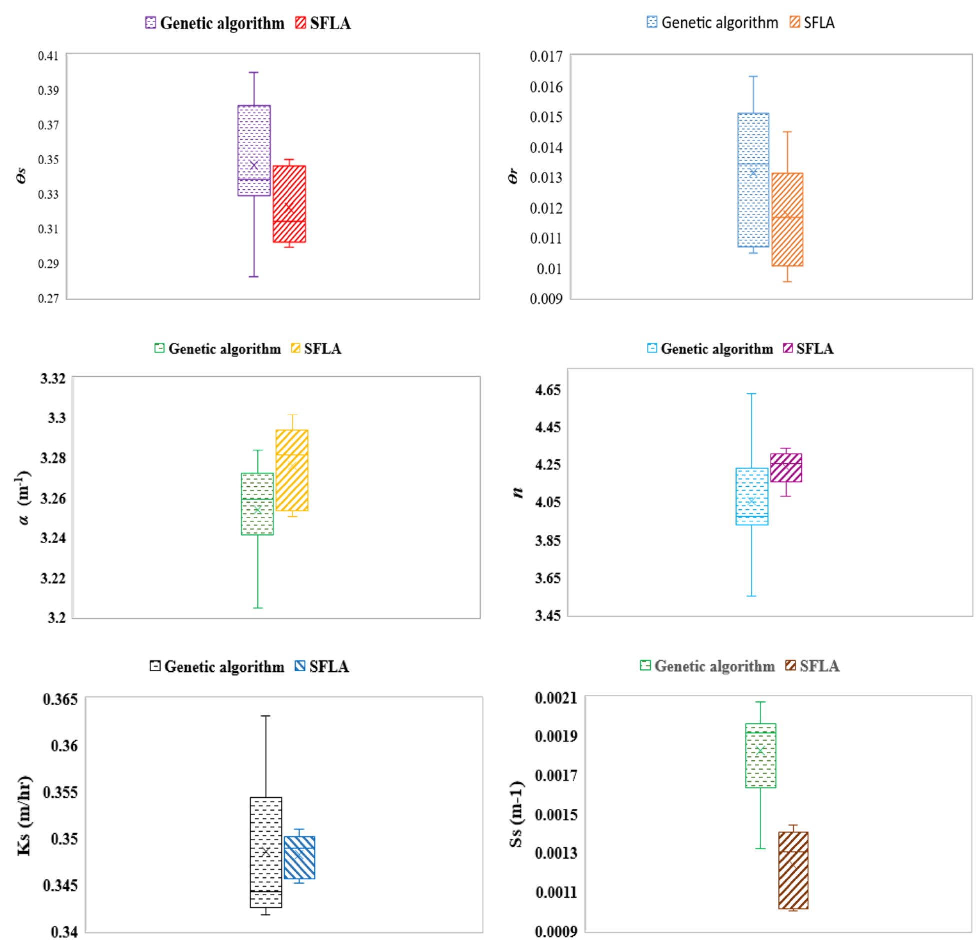


Fig. 9 Box plot representing the estimated value of the flow parameters using ANN-Genetic algorithm and ANN-SFLA

By considering the percentage contribution of the total variance D, Sobol's sensitivity indices are derived for different orders.

The first order sensitivity indices (S_i) on y is then defined as:

$$\text{First-order indices} \quad S_i = \frac{D_i}{D} \tag{12}$$

The second order indices (S_{ij}) on y due to the direct effect between the two parameters x_i and x_j is given by:

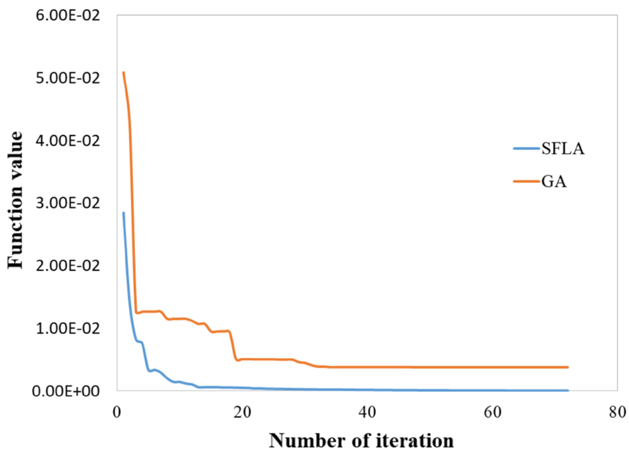


Fig. 10 Variation of the objective function with iteration for genetic algorithm (GA) and shuffled frog leaping algorithm (SFLA)

$$\text{Second-order indices} \quad S_{ij} = \frac{D_{ij}}{D} \tag{13}$$

The total-order indices (S_{Ti}) account for the direct effects between one parameter x_i with the other parameters and are given by Eq. (14):

$$\text{Total order indices} \quad S_{Ti} = 1 - \frac{D_{\sim i}}{D} \tag{14}$$

The variance due to the ith parameter is represented by D_i, while the variance between the two parameters is represented by D_{ij}. $D_{\sim i}$ represent the total variance relating to all parameters except the one for which total order indices are being calculated. In Eq. 19, the variance can be found by using Monte Carlo approximations based on these equations (Hall et al. 2005; Sobol 1993, 2001).

$$\hat{f}_0 = \frac{1}{n}\sum_{S=1}^{n} f(x_s) \tag{15}$$

$$\hat{D} = \frac{1}{n}\sum_{S=1}^{n} f^2(x_s) - \hat{f}_0^2 \tag{16}$$

$$\hat{D}D_i = \frac{1}{n}\sum_{S=1}^{n} f\left(x_s^{(a)}\right) f\left(x_{(\sim i)s}^{(b)} - x_{is}^{(a)}\right) - \hat{f}_0^2 \tag{17}$$

$$\hat{D}D_{ij}^{c} = \frac{1}{n}\sum_{S=1}^{n} f\left(x_s^{(a)}\right) f\left(x_{(\sim i,\sim j)s}^{(b)} - x_{(i,j)s}^{(a)}\right) - \hat{f}_0^2 \tag{18}$$

$$\hat{D}D_{ij} = \hat{D}D_{ij}^{c} - \hat{D}D_i - \hat{D}D_j \tag{19}$$

$$\hat{D}_{\sim i} = \frac{1}{n}\sum_{S=1}^{n} f\left(x_s^{(a)}\right) f\left(x_{(\sim i)s}^{(a)}, x_{is}^{(b)}\right) - \hat{f}_0^2 \tag{20}$$

where, Superscripts (a) and (b) represent different samples in the sampled unit hypercube, where n represents sample size, and x_s represents the sampled individual. Parameters that take their values from a sample (a) are represented by $x_s^{(a)}$. The variables $x_{is}^{(a)}$ and $x_{is}^{(b)}$ are variables that denote parameter x_{is} using sampled values from samples (a) and (b). The $x_{\sim is}^{(a)}$ and $x_{\sim is}^{(b)}$ symbols represent cases where all parameters, except x_{is}, are based on sampled values from samples (a) and (b). Parameters x_{is} and x_{js} are represented by $x_{(ij)s}^{(a)}$ in sample (a) with sampled values. Finally, $x_{(\sim i\sim j)s}^{(a)}$ illustrates the case when all parameters except x_{is} and x_{js} are based on sampled values from sample (b).

Selection of the sample size is one of the most significant steps while carrying out Sobol's sensitivity analysis. The sensitivity indices (total order effect and first-order effect) are calculated with the decision variable as input and hydraulic head as output with different sample sizes. The most suitable sample sizes are selected accordingly. As the sample size increases beyond 10,000, the values of Sobol's indices do not change. This means that at least 10,000 samples should be considered while performing the sensitivity analysis in this study. Using Eq. (12) to (14), the effect of all the parameters on the model output using Sobol's First Order indices (FOI), Second Order indices (SOI), and Total Order indices (TOI) is calculated. The FOI, SOI, and TOI values are shown in Fig. 15 for all the six parameters used in the groundwater flow model, considering a sample size of 15,000.

As per the SOBOL analysis, when the value of the FOI and TOI approaches to 1, this means that the parameter is highly sensitive. On the other hand, the value of FOI should always be less than TOI. In this study, it was observed that a high value of FOI and TOI (>0.9) is observed for Van Genuchten's fitting parameter (α), which means that the parameter (α) is a highly sensitive input parameter. The second-order Sobol indices are also determined to understand the influence of two parameters to the model output. In this flow domain, the highest value of SOI is obtained for (α–n), followed by (α–θ_s), (α–θ_r), (α–Ss), and (α–Ks). This result indicates that α is the most sensitive input parameter. When interacting with the other parameters, it shows the highest value. These findings indicated that the hydraulic head obtained from the model output had a synchronized effect when the parameter α interacted with the other flow parameters, which was impossible to observe during the FOI calculation.

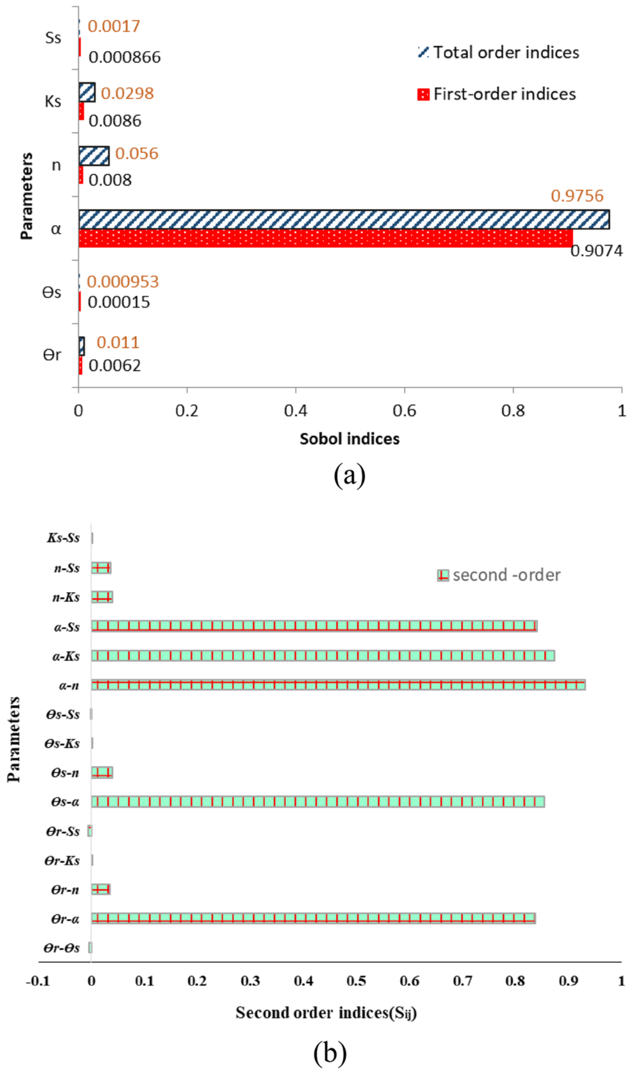


Fig. 11 Sensitivity analysis of **a** first-order indices and total indices and **b** second-order indices for groundwater flow model

Conclusion

This study proposes an effective methodology to estimate the unsaturated and saturated flow parameters together in a single inverse optimization model. As such three-dimensional hypothetical groundwater flow model is developed considering both the saturated and unsaturated zone. The parameters that need to be estimated are the hydraulic parameters given by Van Genuchten and Nielsen (1985) that are $\boldsymbol{\theta_s}$, $\boldsymbol{\theta_r}$, $\boldsymbol{\alpha}$, $\mathbf{K_s}$, **n**, and specific storage ($\mathbf{S_s}$), an essential parameter in the saturated zone. This parameter estimation model is developed using Artificial Neural Network (ANN) and Shuffled Frog Leaping Algorithm to achieve efficiency in computation time and predicting performance. The ANN model is trained using the data generated from the three-dimensional groundwater flow model considering both unsaturated and saturated zones. The result indicates that the ANN-SFLA-based parameter estimation model can predict all six flow parameters well with a minimum relative error and less computational time. Due to its faster convergence and better results, SFLA has shown competitive results when compared to Genetic Algorithm. This may be because memetic evolution in SFLA occurs more rapidly and consists of different evolution sets occurring simultaneously. On the other hand, in GA, genetic evolution is composed of a population (a set of solutions) and evolves as a whole. Therefore, we conclude that ANN-SFLA-based parameter estimation models are a better alternative to solve this parameter estimation problem. The sensitivity study shows that the fitting parameter (α) is a highly sensitive input parameter in the developed groundwater flow model. When analyzing the Sobol indices, it is observed that when α associates with other parameters, it provides high sensitivity values, as shown in Fig. 11. Thus, the Van Genuchten Parameter (α) is considered to be the most sensitive input parameter when developing a groundwater flow model considering both the unsaturated and saturated zones.

Author contributions All the authors have contributed in completing this work on flow parameter estimation in groundwater aquifer. The methodology, analysis were performed by MD, RKB and SAK. The first draft of the manuscript is written by MD, and has been read and approved by all the authors for the final submission of the manuscript.

Data availability statement The materials and data considered in this study are taken from the previous literature and are cited in the manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Abdel-Gawad HA, El-Hadi HA (2009) Parameter estimation of pumping test data using genetic algorithm. In: Thirteenth international water technology conference, IWTC 2009, vol 13

Afzaal H, Farooque AA, Abbas F, Acharya B, Esau T (2020) Groundwater estimation from major physical hydrology componentsusing artificial neural networks and deep learning. Water 12(1):5

Ayvaz MT, Gurarslan G (2019) A hybrid optimization approach for parameter estimation of confined and leaky confined aquifers. Water Supply 19(8):2359–2366

Balkhair KS (2002) Aquifer parameters determination for large diameter wells using neural network approach. J Hydrol 265(1–4):118–128

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. Water Resour Res 22(2):211–227

Cavazzuti M (2012) Optimization methods: from theory to design scientific and technological aspects in mechanics. Springer Science & Business Media, Berlin
Celia MA, Bouloutas ET, Zarba RL (1990) A general mass conservative numerical solution for the unsaturated flow equation. Water Resour Res 26(7):1483–1496
Chang H, Zhang D (2019) Machine learning subsurface flow equations from data. Comput Geosci 23:895–910
Clement TP, Wise WR, Molz FJ (1994) A physically based, two-dimensional, finite-difference algorithm for modeling variably saturated flow. J Hydrol 161(1–4):71–90
Dane JH, Hruska S (1983) In-situ determination of soil hydraulic properties during drainage. Soil Sci Soc Am J 47(4):619–624
Dogan A, Motz LH (2005) Saturated-unsaturated 3D groundwater model. II: Verification and application. J Hydrol Eng 10(6):505–515
Eching SO, Hopmans JW (1993) Optimization of hydraulic functions from transient outflow and soil water pressure data. Soil Sci Soc Am J 57(5):1167–1175
Flood I, Kartam N (1994) Neural networks in civil engineering. I: Principles and understanding. Journal of computing in civil engineering 8(2):131–148.
Gandhi BR, Bhattacharjya RK (2020) Introduction to shuffled frog leaping algorithm and its sensitivity to the parameters of the algorithm. Nature-inspired methods for metaheuristics optimization: algorithms and applications in science and engineering. Springer, Berlin, pp 105–117
Hall JW, Tarantola S, Bates PD, Horritt MS (2005) Distributed sensitivity analysis of flood inundation model calibration. J Hydraul Eng 131(2):117–126
Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT Press, Cambridge
Huang YC, Yeh HD, Lin YC (2008) A computer method based on simulated annealing to identify aquifer parameters using pumping-test data. Int J Numer Anal Meth Geomech 32(3):235–249
Hyun Y, Lee KK (1998) Model identification criteria for inverse estimation of hydraulic parameters. Groundwater 36(2):230–239
Kool JB, Parker JC (1988) Analysis of the inverse problem for transient unsaturated flow. Water Resour Res 24(6):817–830
Kool JB, Parker JC, Van Genuchten MT (1987) Parameter estimation for unsaturated flow and transport models—a review. J Hydrol 91(3–4):255–293
Maier HR, Dandy GC (1996) The use of artificial neural networks for the prediction of water quality parameters. Water Resour Res 32(4):1013–1022
McLaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. Water Resour Res 32(5):1131–1161
Mohanty S, Jha MK, Kumar A, Panda DK (2013) Comparative evaluation of numerical model and artifcial neural network for simulating groundwater fow in Kathajodi-Surua inter-basin of Odisha, India. J Hydrol 495:38–51
Paniconi C, Aldama AA, Wood EF (1991) Numerical evaluation of iterative and noniterative methods for the solution of the nonlinear Richards equation. Water Resour Res 27(6):1147–1163
Şahin AU (2018) A particle swarm optimization assessment for the determination of non-Darcian flow parameters in a confined aquifer. Water Resour Manag 32:751–767
Samuel MP, Jha MK (2003) Estimation of aquifer parameters from pumping test data by genetic algorithm optimization technique. J Irrig Drain Eng 129(5):348–359
Shen C, Laloy E, Elshorbagy A, Albert A, Bales J, Chang FJ, Ganguly S, Hsu KL, Kifer D, Fang Z, Fang K, Li D, Li X, Tsai WP (2018) HESS opinions: incubating deep-learning-powered hydrologic science advances as a community. Hydrol Earth Syst Sci 22:5639–5656
Šimůnek J, Van Genuchten MT (1996) Estimating unsaturated soil hydraulic properties from tension disc infiltrometer data by numerical inversion. Water Resour Res 32(9):2683–2696
Smith J, Eli RN (1995) Neural-network models of rainfall-runoff process. J Water Resour Plan Manag 121(6):499–508
Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1(4):407–414
Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55(1–3):271–280
Swathi B, Eldho TI (2018) Aquifer parameter and zonation structure estimation using meshless local Petrov-Galerkin method and particle swarm optimization. J Hydroinf 20(2):457–467
Van Genuchten MT, Nielsen DR (1985) On describing and predicting the hydraulic properties. Ann Geophys 3(5):615–628
Woodbury AD, Rubin Y (2000) A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site. Water Resour Res 36(1):159–171
Woodbury AD, Ulrych TJ (2000) A full-Bayesian approach to the groundwater inverse problem for steady state flow. Water Resour Res 36(8):2081–2093
Yeh WW (1986) Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour Res 22(2):95–108
Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol 561:918–929
Zhang A, Winterle J, Yang C (2020) Performance comparison of physical process-based and data-driven models: a case study on the Edwards Aquifer, USA. Hydrogeol J 28:2025–2037