Journal of Groundwater Science and Engineering
ISSN 2305-7068
Indexed by ESCI CABI CSA Scopus GeoRef AJ CNKI

Aquifer hydraulic conductivity prediction via coupling model of MCMC-ANN

GUI Chun-lei, WANG Zhen-xing, MA Rong, ZUO Xue-feng

DOI: 10.19637/j.cnki.2305–7068.2021.01.001

Aquifer hydraulic conductivity prediction via coupling model of MCMC-ANN

GUI Chun-lei, WANG Zhen-xing*, MA Rong, ZUO Xue-feng

The Institute of Hydrogeology and Environmental Geology, Key Laboratory of Groundwater Sciences and Engineering, MNR, Shijiazhuang 050061, China.

Abstract: Grain-size distribution data, as a substitute for measuring hydraulic conductivity (K), has often been used to get K value indirectly. With grain-size distribution data of 150 sets of samples being input data, this study combined the Artificial Neural Network technology (ANN) and Markov Chain Monte Carlo method (MCMC), which replaced the Monte Carlo method (MC) of Generalized Likelihood Uncertainty Estimation (GLUE), to establish the GLUE-ANN model for hydraulic conductivity prediction and uncertainty analysis. By means of applying the GLUE-ANN model to a typical piedmont region and central region of North China Plain, and being compared with actually measured values of hydraulic conductivity, the relative error ranges are between 1.55% and 23.53% and between 14.08% and 27.22% respectively, the accuracy of which can meet the requirements of groundwater resources assessment. The global best parameter gained through posterior distribution test indicates that the GLUE-ANN model, which has satisfying sampling efficiency and optimization capability, is able to reasonably reflect the uncertainty of hydrogeological parameters. Furthermore, the influence of stochastic observation error (SOE) in grain-size analysis upon prediction of hydraulic conductivity was discussed, and it is believed that the influence can not be neglected.

Keywords: Grain-size distribution; Hydraulic conductivity; ANN; GLUE; MCMC; Stochastic Observation Error (SOE)

Received: 15 Jul 2020/ Accepted: 12 Sep 2020

GUI Chun-lei, WANG Zhen-xing, MA Rong, *et al.* 2021. Aquifer hydraulic conductivity prediction via coupling model of MCMC-ANN. Journal of Groundwater Science and Engineering, 9(1): 1-11.

Introduction

The southwest part of the North China Plain (NCP) has been widely covered by Cenozoic unconsolidated sediments. For each local area of the plain, Quantization of hydraulic conductivity (K) at different scales can provide scientific support for groundwater exploitation and pollutant transport model-based groundwater resources assessment & pollution remediation. At present, in addition to laboratory tests, there are more numerical simulation inversions to be used to obtain hydraulic conductivity (Smiles and Youngs, 1963; Namunu *et al.* 1989; JI Rui-li *et al*. 2016). Despite some merits, those methods have such demerits as high costs and limited applications due to regional scales (Koekkoek and Booltink, 1999; DONG Pei, 2010; Alfaro Soto *et al*. 2017).

Since the grain-size distribution is the most basic property of sediments and is easily available, researchers both at home and abroad attach much importance to the relationship between the grain-size distribution and hydraulic conductivity to predict the numerical value of hydraulic conductivity (Mahmoud *et al*. 1993; Salarashayeri and Siosemarde, 2012; FAN Gui-sheng *et al*. 2012). As a matter of fact, much attention has been paid to developing empirical or semi-empirical formulas on the basis of grain-size distribution data to provide reliable predication of the K value since

*Corresponding author. E-mail: wangzhxing9987@126.com

the end of 19th century. The Hazen formula (David and Asce, 2003) is the earliest model to predict K value that connects the grain size distribution with hydraulic conductivity through empirical coefficient. From then on, researchers have developed more empirical prediction formulas by selecting a certain effective grain size as their parameters, which can be applied to different kinds of samples (Russell, 1989; Justine, 2007). Compared with other methods of determining the K value, grain-size data has often been used to indirectly obtain the K value for it has been one of the most economical approaches to get K value, and moreover, it does not have any dependence on hydrogeological conditions of study areas during the computing process (Awad and Bassam, 2001). However, there are certain risks to use these empirical formulas to predict K value. For instance, for the well-konwn KC equation, the constant $C_{K\text{-}C}$ has been proved not to be a constant but a function between porosity and fractal dimension, which changes positively with porosity increase (XU Peng *et al*. 2011), thus enlarging the uncertainty of K value prediction of the equation. Furthermore, the most important limitation of these empirical and semi-empirical formulas is that all of them use only one or several parameters of grain-size distribution data, which may omit or neglect information that is included in complete grain-size distribution data system and correlated with hydraulic conductivity prediction.

In recent years, researchers have taken high interest in developing models that can be contrasted with grain-size distribution relationship to calibrate the K value. Artificial neural net (ANN), which serves as a widely used computational tool in a wide range of research areas, has played a satisfying role in the inversion calculation of the K value. Some researchers have employed ANN technology to establish hydraulic property models of the rock mass or soil, including the K value as a function of grain-size distribution. LI Shou-ju *et al*. (2002) constructed a numerical method of K value recognition based on ANN technology and water-head observation data of rock seepage field as well as a priori information of pumping tests, which has empirically been proved to be able to enhance the accuracy of water-head prediction. Nakhaei (2005) used 8 cumulative grain-size fractions for predicting log-transformed hydraulic conductivity for loamy sand with ANN technology, indicating that individual modeling of different soil types is superior to joint modeling. Hasan *et al*. (2006) used grain-size distribution, bulk density, and three different kinds of porosity as input parameters to build ANN model and multi-linear regression model for calculating K value of vadose zone soil, indicating that compared with multiple-linear regression model, ANN model produced more accurate results. TANG Xiao-song *et al*. (2007) used coarse-grained soil of Three Gorges Reservoir area as samples and got K values of different graduation soil through seepage experiments. Then, they employed the powerful nonlinear and dynamic processing capacity of ANN to predict K values. Trough a comparison between the predicted K values and the ones from the seepage experiments, the result indicated that it was feasible to use ANN to predict K of coarse-grained soil. Isik *et al*. (2012) used an ANN models which had three input parameters and one output parameter to predict the KL value of coarse-grained soils with different calculation methods of ANN, suggesting that those different calculation methods of ANN had almost the same prediction capability.

Nowadays most of applications of ANN for predicting K values are confined within the soil research area; there are few ANN applications, which take contents of clay, silt and sand as input variables of grain-size distribution, used for regional water resources assessment. Furthermore, in the framework of stochastic modeling and risk assessment, the quantification of uncertainty variables related to these predictions is of equal importance, which seems to be rarely mentioned and noticed. Therefore, this study combines the general likelihood uncertainty estimation (GLUE) with ANN technology to establish a holistic model for the aquifer hydraulic conductivity inversion and related uncertainty analysis, which takes complete grain-size distribution data of samples as input parameters of ANN model to predict hydraulic conductivity. As a long-term used non-linear modeling approach, it is not uncommon to apply the ANN model to deduce hydrogeological parameters; however, it is not so common to couple ANN with GLUE as an overall model for the K value prediction and its uncertainty analysis. This paper tries to make an empirical study in an area of North-China Plain by coupling ANN model with

GLUE. In addition, besides common uncertainty analysis of parameter, in recent years researchers has paid close attention to the influence of stochastic observation error on simulation results (WANG Dong *et al*. 2009; LU Le and WU Ji-chun, 2010), the study also adequately discussed the influence.

1 Model construction

The model is established through the coupling between ANN and GLUE. The details are illustrated as follows.

1.1 ANN technology

As a deeply-applied computing tool in a wide range of research areas, the ANN can be regarded as a form of nonlinear regression, and multi-layer feedforward network has been proved to possess the property of universal approximation (Haykin, 2004). Some particular ANN models (one type of their structures is shown in Fig. 1) are capable of better predicting hydraulic conductivity (Erzin *et al*. 2009; Park, 2011; Das *et al*. 2012).

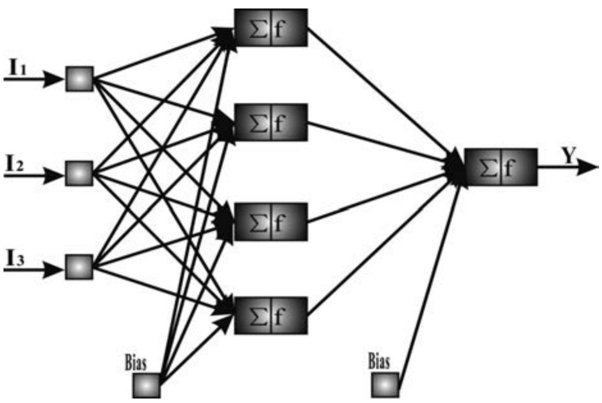

Fig. 1 Neural network architecture

1.2 Bayesian inference

The phenomenon of equifinality for different parameters (Keith, 2006) leads to very large uncertainty for optimal selection of parameters of hydrogeological models. When the degree of uncertainty needs to be expressed, probability and probability distribution is the best language (MAO Shi-song, 1999). As a method of probability analysis, Bayesian method currently is less applied in hydrogeological parameter identification process (LU Le *et al*. 2008). The Bayesian method performs its statistic inference based on the general information, sample information and a priori information, and its density function is shown in Equation (1).

$$\pi(\theta \mid x) = \frac{p(x \mid \theta)\pi(\theta)}{\int_{\theta} p(x \mid \theta)\pi(\theta)d\theta} \tag{1}$$

$\pi(\theta)$ is a prior probability that is the K value probability distribution obtained from rock and soil sample test with definite particle size component;

$\pi(\theta|x)$ is a posterior probability that is the proba-bility that a rock and soil sample with an uncertain particle size component has a certain K value.

1.3 MCMC-ANN coupling model

Markov chain Monte Carlo method (MCMC) was introduced into parameter uncertainty research to estimate the Bayesian distribution sampling of parameters in the 1990s (Smith and Robert, 1993). Compared with Monte Carlo method, MCMC's sampling efficiency is enhanced dramatically and the calculation work is remarkably reduced (GONG Guang-lu and QIAN Min-ping, 2003). The single component adaptive metropolis algorithm (SCAM) (Haario *et al*. 2005) is adopted here to replace Monte Carlo method of traditional GLUE in the paper. SCAM counts the parameter set as a multidimensional vector in which each component represents a parameter. The predicting model of hydraulic conductivity constructed in the study is a coupling model between three-layered ANN and improved GLUE. An ANN model sample set is generated by considering the following random variations of the model parameters: (i) the variation range of each grain-size composition in the given study area; (ii) the quantity variation of the model hidden nods; (iii) initial values of the network weight and bias. The specific steps of model construction is as follows: (1) according to borehole data of the study area and related literature review, the range of parameter values of the model is determined; (2) a likelihood function needs to be selected, and the model efficiency coefficient R^2 is adopted here:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(K_{oi} - K_{pi})^2}{\sum_{i=1}^{N}(K_{oi} - \overline{K})^2} \tag{2}$$

K_{0i} is the actually measured hydraulic conductivity; K_{pi} is predicted hydraulic conductivity; $\overline{K}$ is the mean value of actually measured hydraulic conductivity series; N represents the length of the measured series; (3) according to a priori distribution of parameters, the initial sample X_0 of ANN model is randomly generated; (4) for the i^{th} component of the t^{th} sample, new sample component z^i is generated by using one-dimensional normal distribution $N(X^i_{t-1}, V^i_t)$, and the meaning and calculation method of parameters can be seen in the above-mentioned 24th literature review; (5) new candidate sample component z^i is accepted by a given probability α:

$$\alpha = \min\left[\frac{\dfrac{p(X_t^1,\mathrm{K}\,,X_t^{i-1},z^i,X_{t-1}^{i+1},\mathrm{K}\,,X_{t-1}^{d})\,p(y_t^1,\mathrm{K}\,,X_t^{i-1},z^i,X_{t-1}^{i+1},\mathrm{K}\,,X_{t-1}^{d})}{\int p(X)p(y\,|\,X)dX}}{\dfrac{p(X_t^1,\mathrm{K}\,,X_t^{i-1},X_{t-1}^{i},\mathrm{K}\,,X_{t-1}^{d})\,p(y\big|X_t^1,\mathrm{K}\,,X_t^{i-1},X_{t-1}^{i},\mathrm{K}\,,X_{t-1}^{d})}{\int p(X)p(y\,|\,X)dX}}\right] \quad (3)$$

(6) steps 4 and 5 are repeated until all components of the t^{th} sample is regenerated; (7) steps 4 to 6 are repeated until sufficient samples are obtained, and those samples need to be adjusted to reach the prediction accuracy set in advance; (8) selected likelihood function values are taken as measuring standard, and the samples that are not up to the standard are abandoned, and then the contrastive scatter plot between prediction values and measured values needs to be drawn; (9) it is required to define upper and lower bound of predicted K value and renew likelihood function value. Based on the defined valve value and the ranking order of likelihood function values, the ANN model uncertainty of given confidence level should be estimated.

2 Application example

The constructed GLUE-ANN coupling model has been applied to predict the K value of a study area in the NCP. Details are illustrated as follows:

2.1 Introduction to the study area

The study area is located in the southeast of Shijiazhuang city, southwest of Hebei province. It is situated in the northern latitude of 37°45′-37°51′ and the eastern longitude 114°34′-114°40′ with an area of 126.78 km^2, which is part of the inclined piedmont plain region where east Taihang Mount meets the NCP. The area belongs to the alluvial-proluvial-fan groundwater system of Hutuo River and Huaisha River of Ziya River basin, in which the groundwater mainly lies in the pore of quaternary unconsolidated rock layers. The aquifers below the area transit from single-layer, bi-layer to multi-layer structures. In the horizontal direction, single-layer thickness of the aquifers gradually thickens from west to east, its grain size becomes coarser, and the number of layers increases with the water yield property getting stronger. In the vertical direction, the grain size in the upper and lower layers is finer with relatively thinner aquifers, compared with coarser grain size and thicker aquifers in the middle layers.

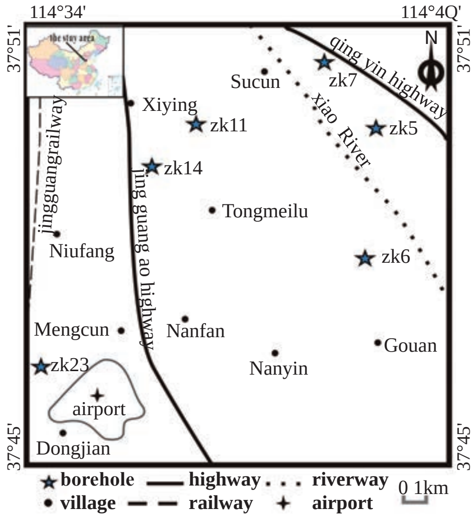

Fig. 2 Study area and borehole distribution

2.2 Data measurement

Six boreholes (zk05, zk06, zk07, zk11, zk14 and zk23) were selected in this study, and their distribution is shown in Fig. 2. High-recovery and low-disturbance boosting core sampler was used to take 0.4 m undisturbed soil samples every 2 m, and then DZS70 constant head permeameter and TST55 variable head permeameter were applied to test the K value of sandy soil samples and clay soil ones. During tests, deionized water was injected to the bottom of the samples under a constant pressure and K value was determined by flow measurements in accordance with the Darcy’s law. For high clay contents, the measurement could last for 2 to 3 weeks until steady flow was achieved. In the case of more silty or sandy samples, a 1 m high water column was employed with the testing time not less than one day and the accuracy could reach about 10%. Meanwhile, samples used for grain size analysis were taken from each undisturbed soil sample. There were altogether 351 soil samples for grain size analysis, including 170 sets used for obtain contrastive data between grain size composition and actually measured values of hydraulic conductivity.

2.3 Generation of ANN model samples

By taking the model efficiency coefficient R^2 as the likelihood function, the SCAM method was adopted here to replace the Monte Carlo to take samples. 12 000 ANN model samples were taken which contained actually measured values of grain size data. The a priori distribution of each parameter was tested by the Bayesian assumption, assuming that parameters within their value range subject to uniform distribution, as shown in Table 1.

Table 1 Parameters in the ANN model

Parameter	Physical meaning	Value range
D_{2000+}	Gravel content, diameter more than 2 mm	0~0.02
D_{500}	Coarse sand content, diameter between 0.5~2 mm	0.01~0.46
D_{250}	Medium sand content, diameter between 0.25~0.5 mm	0.01~0.82
D_{75}	Fine sand content, diameter between 0.075~0.25 mm	0.02~0.51
D_5	Silt content, diameter between 0.005~0.075 mm	0.01~0.68
D_{5-}	Clay content, diameter less than 0.005 mm	0.01~0.65
N_2	The number of hidden layer nodes	7~21
$W^p_{i,j}$	Network weight initial value	-1~1
b^p_i	Bias	-1~1

2.4 Model training and uncertainty analysis

150 sets of data sets from borehole zk05, zk06, zk07, zk11, zk14 and zk23 were applied to train the models. For each K value of the 150 samples, it was assumed that the ensemble of model samples consisted of 1 000 ANN models. The number of models was set to 1 000 so that the estimated distribution can keep balance to achieve reasonable convergence and balanced calculation load. All of the analyses and calculations were processed by using relevant functions of neural network BP toolkit in the Matlab environment. Whole grain size fractions, initial weights, bias data, and the number of node in hidden layers were used as input parameters of the network and output data was taken as the samples' hydraulic conductivity. Standardization of input is capable of speeding up neural network training and reducing the probability of being blocked in the local optimization process, so the input data of network was normalized actually-measured values of grain size fractions. In order to ensure that output data could range from 0 to 1, all the transfer functions used in hidden layers and output layers were the logsig functions.

Levenberg-Marquart rule was adopted to train the network and the maximal training step was set to 2 500. Error values were obtained according to Equation (4):

$$E = \frac{1}{2}\sum_{p_l=1}^{l}\sum_{K=1}^{m}(t_{k,}p_l - y_{k,}p_l)^2 \tag{4}$$

Where: E value is 10^{-14} m/s. In the formula: E is the total error between actually measured and output values; pl are the grain size components; t_k are actually measured values; y_k are output values. Based on related literature review and the testing results, the minimum of hydraulic conductivity can be $1.12*10^{-9}$ m/s, and it was used as minimum target value, and after squaring it, the error accuracy reached 10^{-18} m/s. Therefore it is reasonable to value the total error E as 10~14 m/s. For minimizing the error, initial weight and bias needed to be reasonably selected. Even-distributed decimals between -1 and 1 are generally chosen. The number of node in hidden layers has larger random, and its optimal number depends on complexity of problem, which is usually determined by trial and error. Generally speaking, the number does not exceed twice as much as the number of input nodes.

After training the accuracy of the model reached the requirement. The value of likelihood function R^2 was 0.7. After those ANN model samples below the value were abandoned, 5 520

ANN model samples were finally obtained. When likelihood function value was 0.82, the scatter diagram of output values vs. measured values is shown in the Fig. 3.

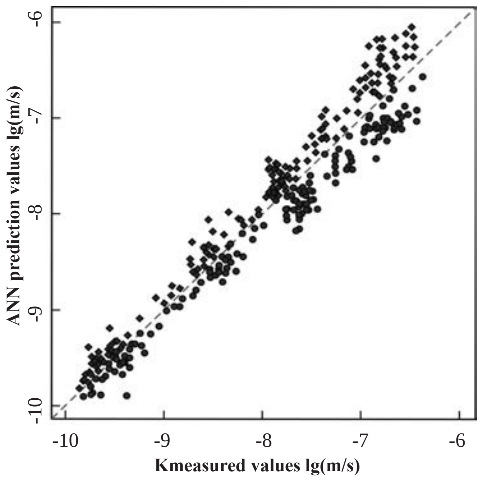

Fig. 3 The output versus measured value scatter plots (R^2=0.82)

The critical value of model efficiency coefficient took 0.91 when uncertainty of model parameter was analyzed, and zero was assigned to the likelihood values of parameter sets below the critical value. The likelihood values of parameter sets above the critical value were normalized, and sorted by size of the likelihood values, the uncertainty range of models that are below the 90% confidence level needs to be calculated.

2.5 Application

It's required to use the before-mentioned model to predict the hydraulic conductivities of 20 samples from the borehole zk11 of the same study area and the comparison results of predicted and measured values are shown in Table 2.

Table 2 Comparison between the ANN model's output and measured values

Sample number	Gravel	Coarse sand	Medium sand	Fine sand	Silt	Clay	Predicted K 10^{-6} (cm/s)	Measured K 10^{-6} (cm/s)	Relative error %
	>2	2~0.5	0.5~0.25	0.25~0.075	0.075~0.005	<0.005			
	content %								
zk11-01	1.15	4.05	12.33	5.10	64.06	13.31	18.01	18.33	1.75
zk11-02	0.00	4.10	14.59	8.18	57.12	16.01	13.34	13.59	1.84
zk11-03	0.00	3.35	8.09	10.17	33.12	45.27	2.42	2.98	18.79
zk11-04	0.53	7.26	20.20	4.11	49.07	18.83	9.29	9.10	2.08
zk11-05	0.00	10.32	13.52	24.97	42.15	9.04	28.91	28.47	1.55
zk11-06	0.00	15.16	46.94	28.05	8.14	1.71	602.56	665.30	9.43
zk11-07	0.00	8.20	1.08	3.34	62.30	25.08	2.65	2.51	5.58
zk11-08	1.00	1.13	1.19	5.17	65.40	26.11	3.80	3.92	3.06
zk11-09	0.00	31.22	39.14	1.35	26.54	1.75	536.11	589.78	9.10
zk11-10	0.00	1.26	5.79	1.62	54.45	36.88	1.87	1.97	5.08
zk11-11	0.00	32.49	27.55	6.88	30.87	2.21	410.41	456.98	10.19
zk11-12	0.00	28.01	38.40	10.02	21.91	1.66	445.09	415.70	7.07
zk11-13	0.00	20.34	42.15	4.37	31.51	1.63	404.02	448.31	9.88
zk11-14	0.00	2.30	2.27	13.25	35.90	46.28	2.21	2.89	23.53
zk11-15	0.00	3.36	1.33	1.03	54.73	39.55	1.23	1.36	9.56
zk11-16	0.00	39.22	20.68	7.21	30.70	2.19	394.80	439.18	10.11
zk11-17	0.00	24.98	55.17	11.15	6.97	1.73	589.85	644.71	8.51
zk11-18	0.00	2.24	1.12	2.33	56.66	37.65	1.18	1.05	12.38
zk11-19	0.00	1.27	2.97	2.10	60.15	33.51	0.88	0.92	4.35
zk11-20	2.06	5.10	4.55	8.51	53.28	26.50	5.03	5.24	4.01

In Table 2, K values were measured by DZS70 constant head permeameter; grain-size composition contents were calculated based on the grain-size testing report made by the monitoring center of groundwater & mineral water & environment, Ministry of Land and Resources, China.

The test data in the Table 2 was obtained from 20 sedimentary samples of aquifer and aquitard at various depths in the borehole zk11, which includes various lithologies from coarse sand, fine sand to silt clay. The predicted results showed that for silt and fine sand samples (in zk11-01, zk11-02, zk11-05, zk11-07 and zk11-19, *etc*.), ANN model has higher calculation accuracy with relative error from 1.55% to 5.58%. Most of the order of magnitude of predicted hydraulic conductivity values of these samples are 10^{-9} m/s. With the content of clay in samples rising, the predicted orders of magnitude drop a little, which indicates the model has higher sensitivity to clay content. For coarse sand and medium coarse sand samples (in zk11-06, zk11-09, zk11-12 and zk11-13, *etc.*), the relative errors of predicted values are between 7.07% and 10.19%, and the prediction accuracy is a little lower than that of silt samples. Among all the samples, the predicted and measured values of zk11-06 and zk11-17 are the highest, which shows that the BP neural network can better reflect the rule that the higher sorting capability of the grain is, the larger their hydraulic conductivity is. For silt clay and some silt samples that have higher clay content (zk11-03, zk11-10, zk11-14 and zk11-15, *etc*.), the relative errors of predicted values of ANN model are between 9.56% and 23.53%. The relative errors and variation range are relatively larger, which indicates that the model needs to be further improved.

In order to validate the model's applicability to variant sedimentary environment, the established model above was used to predict K values of 7 sedimentary samples in a scientific borehole of central region of NCP, and the comparison results of predicted and measured values are shown in Table 3.

Table 3 The comparison between the ANN model's output and measured values of samples from middle region of NCP

Sample number	Gravel	Coarse sand	Medium sand	Fine sand	Silt	Clay	Predicted K 10^{-6} (cm/s)	Measured K 10^{-6} (cm/s)	Relative error %
	>2	2~0.5	0.5~0.25	0.25~0.075	0.075~0.005	<0.005			
	content %								
S18-01	0.00	23.52	20.58	33.54	12.25	10.11	17.33	20.17	14.08
S18-02	0.00	25.08	17.41	45.88	9.38	2.25	398.75	475.32	16.11
S18-03	0.00	25.49	22.72	28.10	10.89	12.80	42.36	36.74	15.30
S18-04	0.00	29.37	25.26	21.65	14.17	9.55	31.91	40.47	21.15
S18-05	0.00	28.90	19.21	41.20	8.92	1.77	332.97	412.55	19.29
S18-06	0.00	28.83	19.88	37.21	11.12	2.96	274.10	376.61	27.22
S18-07	0.00	41.28	22.43	17.57	4.22	14.50	13.73	18.39	25.34

In Table 3, K values were measured by DZS70 constant head permeameter; grain-size composition contents were calculated based on the grain-size testing report made by the monitoring center of groundwater & mineral water & environment, Ministry of Land and Resources, China.

In Table 3, seven sedimentary samples of borehole S18 that were obtained from aquifer and aquitard in different depth are contained, these samples consist of coarse sand, medium sand and fine sand *et al*. The data in the table shows that the relative errors of prediction results of GLUE-ANN model are between 14.08% and 27.22%, the prediction accuracy also conform to basic requirements of groundwater resources evaluation.

Posterior distributions of some models' parameters that were obtained by using improved GLUE are shown in Fig. 4, which reflect value range probability of parameters in the whole domain of definition. As the figure shows, the high probability areas of parameters' posterior distribution are discontinuous, and the areas of global optimum are easily determined from the diagram. For instance, the posterior distribution of hidden node number appears to reach relatively high probability value when the number is around

14, which corresponds to the number of hidden node of hidden layers used in this study when the model efficiency coefficient is at higher levels. For most of the models, posterior distributions of their parameters have higher searching performance because of the application of SCAM method. The posterior distribution of D5, however, appears to be flat, indicating that the model still needs to be improved to enhance its sensitivity to some parameters.

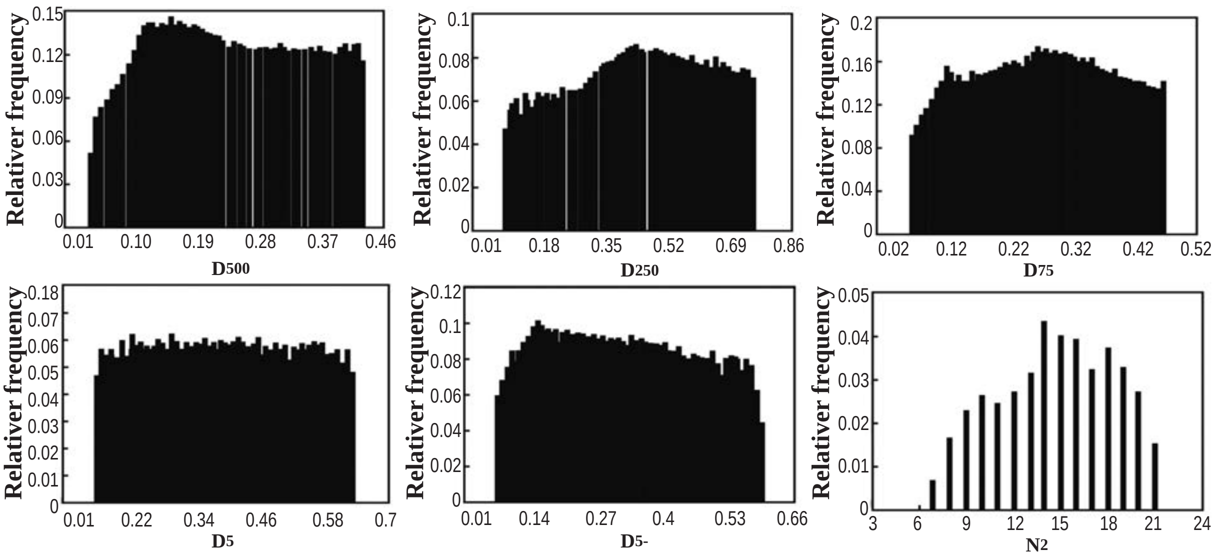

Fig. 4 Posterior probability distribution of GLUE-ANN model's parameters

When R^2 is values as 0.91, uncertainty range of ANN model coefficient at the 90% confidence level is shown in Fig. 5, which includes the actually-measured values of hydraulic conductivity and GLUE-ANN model's prediction values. As shown in the figure, all of the prediction values are within two orders of magnitude of corresponding actually-measured values, and furthermore, most of the differences are within one order of magnitude. When K values are between 10^{-7} m/s and 10^{-8} m/s, the model prediction values have less deviation, demonstrating that the prediction and uncertainty analysis made by the constructed model are satisfying.

3 Stochastic observation error and uncertainty

In grain-size test of sedimentary sample, instrument's measuring error and inexpert command of measuring skill can all cause observation error. Because the observation error is stochastic, using the data with stochastic observation error (SOE) to simulate aquifer parameter can lead to uncertainty of simulation results.

In order to be compared with uncertainty of simulation results with SOE, the uncertainty of a sedimentary sample for reference is analyzed firstly. The contents of medium sand and fine sand of the assumed reference sample are respectively 22.5% and 35.10%. After using MCMC to sample the reference sample, GLUE-ANN model is applied to predict and evaluate parameter uncertainty. When the number of samples was beyond 45 000, the series of reference sample converge to posterior distribution. As shown in Fig. 6.

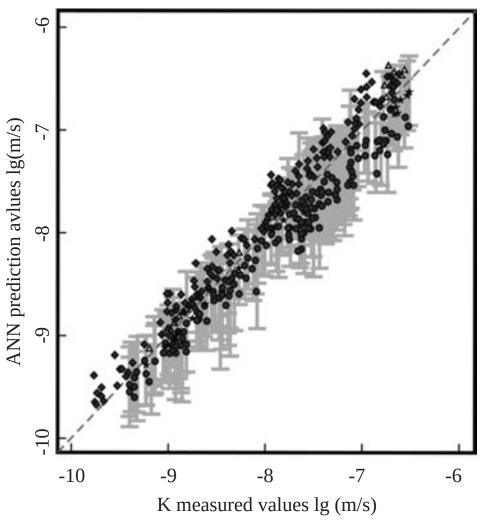

Fig. 5 Actually-measured K values versus the GLUE-ANN model prediction and uncertainty estimation

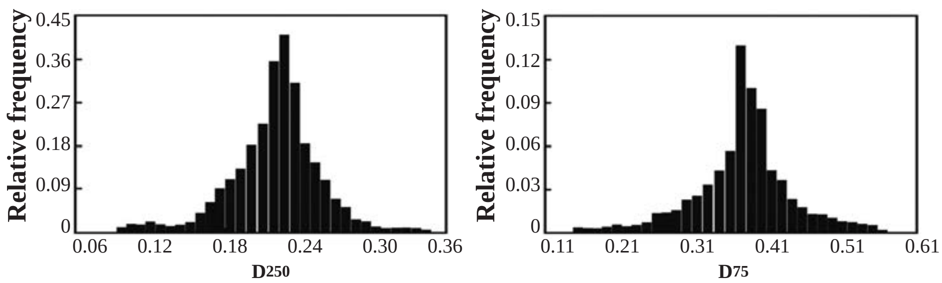

Fig. 6 The posterior distribution of refence sample

For inspecting the influence of SOE in grain-size analysis upon uncertainty of model output, the relative error that mean is 0 and variance is 0.001 is artificially added to the observation data of medium sand and fine sand of reference samples. Simulation and uncertainty analysis is performed using the observation data with SOE thereby to explore the effect of SOE on uncertainty of simulation results. When the number of samples exceeds 30 000, the samples converge to parameter's posterior distribution as shown in Fig. 7.

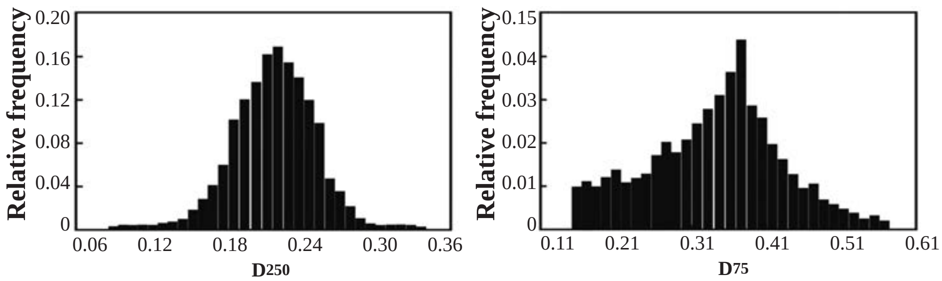

Fig. 7 The posterior distribution of reference samples with SOE

As shown in Fig. 7, compared to reference sample's posterior distribution (Fig. 6), the posterior distributions of the two parameters after adding SOE to them appear more scattered. It is evident that SOE has significantly increased the uncertainty of posterior parameters and consequently increased uncertainty of simulation results.

4 Conclusions

An improved GLUE-ANN model has been established in this study by coupling ANN technology and MCMC method, which replaces the Monte Carlo algorithm in traditional GLUE method. Compared with the Monte Carlo method, SCAM possesses superior searching performance for posterior distribution of model parameters, ensuring that the MCMC method can both search the global scale of posterior distribution and locate the extent of optimum values of parameters. Results of this study show that the improved GLUE-ANN model is capable of effectively predicting the K value and analyzing uncertainty of the parameters.

The advantages of the GLUE-ANN model created in the paper lies in taking fully into account the influence of different grain sizes and their content on permeability and the availability and universality of grain-size data in hydrogeological borehole data. The input parameters of the model adopted a complete grain size component consisting of clay, silt and gravel that can reflect the real soil structure in the study area. Therefore, the K values predicted by the model are closer to the value in natural state. If larger quantities of typical samples are available, predictions from the established model may become more rational. The study has not taken organic matter, carbonate content, soil porosity and its density as input parameters, and if these factors were considered, the prediction accuracy could be further enhanced. This reflects that there are many influencing factors on the hydraulic conductivity calculation leading to the uncertainty of predictions, which needs to be further studied.

Overall, the prediction results made by the GLUE-ANN model are relatively closer to actually-measured values with satisfying uncertainty estimation, which can serve as a precise way of predicting the hydraulic conductivity. It has been proved that it is capable of accurately predicting the hydraulic conductivity in the study area and other similar alluvial-proluvial plain regions and providing fundamental data for the improvement of solute transport model.

By way of adding SOE to observation data, the influence of SOE on simulation results is markedly reflected, which indicates that the uncertainty of simulation results caused by SOE can not be neglected.

Acknowledgements

This study was supported by Key Laboratory of Groundwater Sciences and Engineering, Ministry of Natural Resources (MNR) and the China Geological Survey project (No. DD20190252).

References

Alfaro Soto MA, Chang HK, Th M, *et al*. 2017. Fractal-based models for the unsaturated soil hydraulic functions. Geoderma, 306: 144-151.

Awad HS, Bassam AM. 2001. A computer program to calculate hydraulic conductivity from grain size data in Saudi Arabia. International Journal of Water Resources Development, 17(2): 237-246.

Das SK, Samui P, Sabat AK. 2012. Prediction of field hydraulic conductivity of clay liners using an artificial neural network and support vector machine. International Journal of Geomechanics, 12(5): 606-611.

David CW, Asce F. 2003. Goodbye, Hazen; Hello, Kozeny-Carman. Journal of Geotechnical and Geoenvironmental Engineering, 129(11): 1054-1056.

DONG Pei. 2010. Laboratory studies of sand column on the dynamic evaporation on a shallow water table. M.S. thesis, Beijing: China University of Geosciences: 10-22. (in Chinese)

Erzin Y, Gumaste SD, Gupta AK, *et al*. 2009. Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soil. Canadian Geotechnical Journal, 46: 955-968

FAN Gui-sheng, XING Ri-xian, ZHANG Ming-bin. 2012. Experimental study on permeability of the sandy gravel media with different gradation. Journal of Taiyuan University of Technology, 43(3): 373-378. (in Chinese)

GONG Guang-lu, QIAN Min-ping. 2003. Applied stochastic coursebook and its application to algorithm and intelligent computing. Beijing: Tsinghua University Press: 35-58. (in Chinese)

Haario H, Saksman E, Tamminen J. 2005. Componentwise adaptation for high dimensional MCMC. Computational Statistics, 20(2): 265-273.

Hasan M, Ozer C, Ramazan M, *et al*. 2006. Comparison of artificial neural and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil and Tillage Research, 90(2): 108-116.

Haykin S. 2004. Neural Networks: A comprehensive foundation, second edition. Beijing: China Machine Press: 46-55. (in Chinese)

Isik Y, Marian M, Martin B. 2012. Neural computing models for prediction of permeability coefficient of coarse-grained soils. Neural Computing and Applications, 21(5): 957-968.

JI Rui-li, ZHANG Ming, SU Rui, 2016. Research of in-situ hydraulic test method by using double packer equipment. Journal of Groundwater Science and Engineering, 4(1): 41-51.

Justine O. 2007. Evaluation of empirical formulae for determination of hydraulic conductivity based on Grain-size analysis. Journal of American Science, 3(3): 54-60.

Keith B. 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2): 18-36.

Koekkoek JW, Booltink H. 1999. Neural network models to predict soil water retention. European Journal of Soil Science, 50(3): 489-495.

LI Shou-ju, LIU Ying-xi, WANG Deng-gang, *et al*. 2002. Inversion algorithm of permeability coefficients of rockmass and its application based on artificial neural network. Chinese Journal of Rock Mechanics and Engineering, 21(4): 479-483. (in Chinese)

LU Le, WU Ji-chun. 2010. Bayesian analysis of uncertainties in groundwater numerical simulation. Journal of Hydraulic Engineering, 41(3): 264-271.

LU Le, WU Ji-chun, CHEN Jing-Ya. 2008. Identification of hydrogeological parameters based on the bayesian method. Hydrogeology and Engineering Geology, 58(5): 58-63. (in Chinese)

Mahmoud S, Alyaman I, Zekai S. 1993. Determination of hydraulic conductivity from complete Grain-Size Distribution Curves. Ground Water, 31(4): 551-555.

MAO Shi-song. 1999. Bayesian statistics. Beijing: China Statistics Press. (in Chinese)

Nakhaei M. 2005. Estimating the saturated hydraulic conductivity of granular material using artificial neural network based on grain size distribution curve. Islam Repub Iran, 16(1): 55-62.

Namunu JM, Ian PK, Arulanandan K. 1989. An expression for the permeability of anisotropic granular media. International Journal for Numerical and Analytical Methods in Geomechanics, 13(6): 575-598.

Park HI. 2011. Development of neural network model to estimate the permeability coefficient of soils. Marine Georesources & Geotechnology, 29(4): 267-278.

Russel GS. 1989. Correlations of permeability and grain size. Ground Water, 27(5): 633-636.

Salarashayeri AF, Siosemarde M. 2012. Prediction of soil hydraulic conductivity from particle-size distribution. World Academy of Science, Engineering and Technology, 61: 454-457.

Smiles DE, Youngs EG. 1963. A Multiple-well method for determining the hydraulic conductivity of a saturated soil in Situ. Journal of Hydrology, 1(4): 279-287.

Smith AFM, Roberts GO. 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo Methods. Journal of Royal Statistical Society, Series B, 55: 3-24.

TANG Xiao-song, ZHENG Ying-ren, DONG Cheng. 2007. The prediction of seepage coefficient of coarse-grained soil by neurotic network. Rock and Soil Mechanics, 28: 133-136,143. (in Chinese)

WANG Dong, Vijay PS, ZHU Yuan-shen, *et al*. 2009. Stochastic observation error and uncertainty in water quality evaluation. Advances in Water Resources, 32: 1526-1534.

XU Peng, QIU Shu-xia, JIANG Zhou-ting, *et al*. 2011. Fractal analysis of Kozeny-Carman constant in the homogenous porous media. Journal of Chongqing University, 34(4): 78-82. (in Chinese)