Journal of Groundwater Science and Engineering 13(2025) 434–448

Journal homepage: http://gwse.iheg.org.cn

Research Article

ANN-based prediction model for single-hole water inflow from piedmont to inland plain areas of Hebei Province, North China Plain

Hong-wei Song[1,2], Fan Xia[3,4*], Wei-qiang Wang[1,2], Ming-sen Shang[1,2], Jian-ye Gui[1,2]

[1] *Institute of Hydrogeology and Environmental Geology*, *Chinese Academy of Geological Sciences*, *Shijiazhuang* 050061, *China*.
[2] *Key Laboratory of Groundwater Sciences and Engineering*, *Ministry of Natural Resources*, *Shijiazhuang* 050061, *China*.
[3] *Hebei Provincial of Academy of Ecological and Environmental Sciences*, *Shijiazhuang* 050031, *China*.
[4] *Hebei Water Environmental Science Laboratory*, *Shijiazhuang* 050031, *China*.

Abstract: This study, based on Artificial Neural Network (ANN) technology, develops a quantitative prediction model for the unit water flow rate of the Quaternary strata in the Shijiazhuang-Hebei Plain area. The study area extends from the piedmont region of Shijiazhuang, at the eastern foothills of the Taihang Mountains, to the hinterland area of Hengshui within the plain in Hebei Province section of the North China Plain. The hydrological and exploration boreholes selected for modeling are primarily located in the southeastern part of Shijiazhuang urban area — the southern region of Xinji County — north of Hengshui City near the Shenzhou County area. By employing the Induced Polarization method (IP) and Vertical Electrical Sounding (VES), apparent resistivity (ρS), apparent polarization rate (ηS), half-decay time (Th), and decay degree (D) were obtained as initial input parameters. These were combined with the measured water flow rates from borehole pumping tests to build the training sample set. To address the prevalent issue of high-salinity interference in the study area, multiple regression analysis revealed that when the inverted resistivity (ρ) is less than 5 Ω·m and the inverted polarization rate (η) is greater than 8%, the contribution of groundwater salinity to the resistivity parameter reaches 42%±6%. Based on this, a comprehensive parameter T"=ρ*H/ρ' was established, where ρ is the aquifer resistivity, ρ' is the aquitard resistivity, and H is the aquifer thickness. The resistivity ratio effectively eliminates the coupling effect between electrical parameters and salinity. The input neurons of the improved model were adjusted to a four-parameter system consisting of decay time (Th), decay degree (D), deviation degree (σ), and the comprehensive parameter (T"). Experiments showed that the prediction error of the model on the validation set was reduced from the original model's 5%-10% to 0.9%-5%. The introduction of the T" parameter reduced the prediction error in high salinity areas (Cl->500 mg/L) to within 7%. The study confirms that the composite parameter T" based on geophysical inversion parameters can effectively characterize the coupling features of aquifer thickness and water quality. Even with a small sample size, through algorithm optimization, data augmentation, and model structural improvements, it is entirely possible to effectively enhance prediction accuracy and generalization ability, providing a new parameterization method for the quantitative evaluation of Quaternary pore water in plain areas.

Keywords: Artificial Neural Network; Single-Hole; Aquifer Thickness; Resistivity; Induced Polarization

Received: 13 Jan 2025/ Accepted: 12 Aug 2025/ Published: 10 Oct 2025

*Corresponding author: Fan Xia, *E-mail address*: fanlevip@163.com

DOI: 10.26599/JGSE.2025.9280064

Song HW, Xia F, Wang WQ, et al. 2025. ANN-based prediction model for single-hole water inflow from piedmont to inland plain areas of Hebei Province, North China Plain. Journal of Groundwater Science and Engineering, 13(4): 434-448.

Introduction

The Induced Polarization methods (IP), Vertical Electrical Soundings (VES) and Electrical Resistivity Tomography (ERT) are widely used geophysics techniques that have become some of the most practical and cost-effective tools for hydrogeological exploration. These methods are endorsed by many hydrogeologists due to their simplicity,

speed, high efficiency and low cost (Gyeltshen et al. 2023). They are minimally affected by topographic relief and the inhomogeneity of surrounding rock resistivity, and they can fully leverage detection in the time or frequency domain to indirectly locate groundwater (Liu et al. 2007; Zhong et al. 2021). The electrical parameters obtained, such as apparent resistivity, apparent polarization, half-decay time and attenuation degree, have high resolution and accuracy for classifying aquifer structures and delineating water-conducting zones (Lin et al. 2013; Ling et al. 2007; Persico et al. 2023; Scanlon et al. 2000; Song et al. 2020).

In recent years, the rapid development of machine learning modeling technology and their application across various fields have brought new opportunities and significantly accelerated research progress (Saravanan et al. 2024; Yilmaz et al. 2022). In terms of model structure and algorithm optimization: Early studies constructed groundwater dynamic prediction models using the standard BP algorithm. For example, input parameters were determined through autoregressive analysis and groundwater depth prediction models were established by using trial-and-error methods to determine the number of hidden layer nodes. Later, to address issues such as the slow convergence speed of the BP algorithm and its tendency to fall into local minima, researchers introduced various optimization algorithms. For instance, the LM (Levenberg-Marquardt) algorithm, with its second-order convergence characteristics, was applied to groundwater dynamic prediction and significantly improved training efficiency (Devi et al. 2022). The Bayesian Regularization (BR) algorithm introduced a regularization term to inhibit overfitting, demonstrating better prediction performance than the GDX and LM algorithms. Furthermore, the development of hybrid algorithms, such as GN-BFGS, further enhanced the accuracy of parameter inversion for complex aquifers.

In terms of regional adaptability and interdisciplinary integration, current research covers diverse hydrogeological settings, including arid regions, karst spring areas, and alluvial plains. For instance, accelerated genetic algorithms have been used to optimize neural networks and have been successfully applied to the inversion of heterogeneous aquifer parameters (Li et al. 2024; Sivakrishna et al. 2022; Xia et al. 2019). Additionally, genetic algorithms combined with BP networks have been used to predict groundwater quality in oil-contaminated areas, significantly improving prediction accuracy. The integration of ANN models with traditional numerical models (such as MODFLOW) has also become increasingly popular. By combining ANN as a supplement to numerical models, predictions can be made through system-level analysis in data-scarce areas. Some studies further integrate ANN with grey system theory and fuzzy theory to better address uncertainty. It is particularly well integrated with geophysical prospecting in the fields of exploration and monitoring. For example, ANN methods have been applied to the investigation of environmental pollution, groundwater resource exploration, mineral reserves estimation, archaeology, underground air-raid shelter mapping and the delineation of brackish-freshwater interfaces in areas affected by seawater intrusion (Liu et al. 2011; Maiorov et al. 2021; Mnasri et al. 2023; Patra et al. 2016).

The introduction of Artificial Neural Network (ANN) technology overcomes the limitations of traditional regression-based prediction models by establishing a quantitative relationship between the abnormal parameters derived from electrical methods and the degree of groundwater enrichment. This approach shifts the interpretation from purely qualitative delineation to the quantitative estimation, enabling IP data to not only identify the spatial extent of groundwater occurrence using anomalous values but also to provide a quantitative explanation of formation water content (Behzad et al. 2010; Maiti et al. 2012).

In conventional ANN-based water content prediction models, input neurons typically include parameters such as resistivity, apparent polarizability, half decay time, and attenuation. However, the acquisition of these parameters is often susceptible to external disturbance, which can lead to abnormal values and cause deviation or instablility in predictions, making it challenging to accurately evaluate unknown hydrogeological units and limiting broader application. To mitigate the influence of external factors such as salinity, additional parameters such as aquifer thickness and deviation degree have been introduced. These additions enhance the model's robustness and stability, significantly improving the prediction accuracy of ANN models (Davydycheva et al. 2006; Dong et al. 1996; Duragasi et al. 2023).

In this study, a large number of boreholes with known hydrogeological data from the Shijiazhuang and Hengshui areas of the North China Plain were used as training samples. Electrical anomaly parameters obtained through borehole electrical sounding were used as input neurons, and the unit water inflow of a single hole served as the output neurons to establish an ANN-based

water content prediction model, yielding satisfactory results (Ghosh et al. 2023; Hayashi et al. 1993).

1 Materials and methods

1.1 Study area

Because the model is primarily designed for application in sedimentary strata within the plain area, the North China Plain serves as an ideal comprehensive test site due to its extensive distribution and diverse geological conditions. Accordingly, the study area was selected as an inland region extending from the southeastern part of Shijiazhuang City, located at the eastern foothills of the Taihang Mountains, passing through the southern part of Xinji County, and reaching the northern part of Hengshui City near Shenzhou County in Hebei Province. This area is characterized by abundant drilling activities (Fig. 1).

In this study area, nine exploration holes and seven monitoring holes (K1-K9 and J1-J7, respectively, Fig. 2) were selected as modeling samples, all of which have known pumping test data. Induced polarization and resistivity logging were conducted near these wells. The resulting electrical logging curves were processed and interpreted through inversion to calculate and statistically analyze the resistivity and induced polarization parameters of the aquifers.

The obtained data were then used for model training and validation through a cyclic cross-validation approch. Finally, the known pumping wells located in the Shijiazhuang and Hengshui study areas were used as test sets to evaluate the model's generalization ability. The groundwater level contour map (Fig. 2) illustrates the regional groundwater flow direction, while the hydrogeological profiles (Fig. 3) depict the lithological variation trends from the foothill zone to the center of the plain.

1.1.1 Shijiazhuang

Shijiazhuang is situated in the core zone of the Piedmont Plain of the North China Plain, bordering the majestic Taihang Mountains to its west. Geotectonically, this region lies within the Taihang Mountain meridional structural belt and is influ-

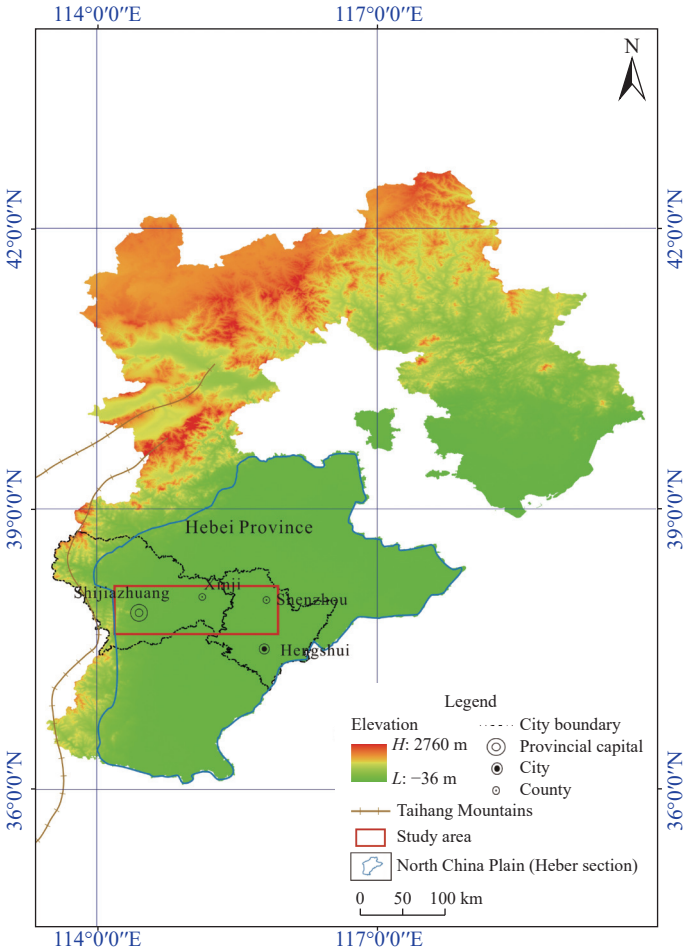

Fig. 1 Study area

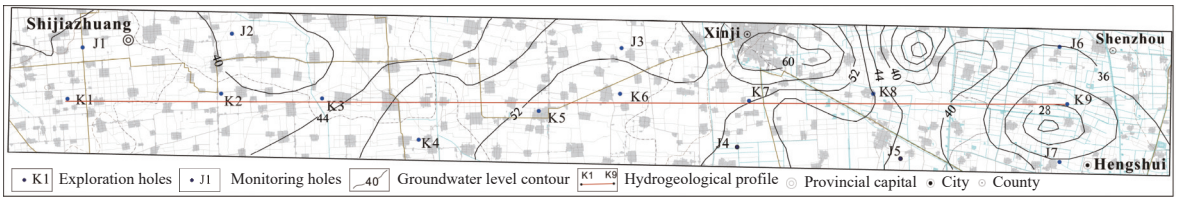

Fig. 2 Groundwater level contour

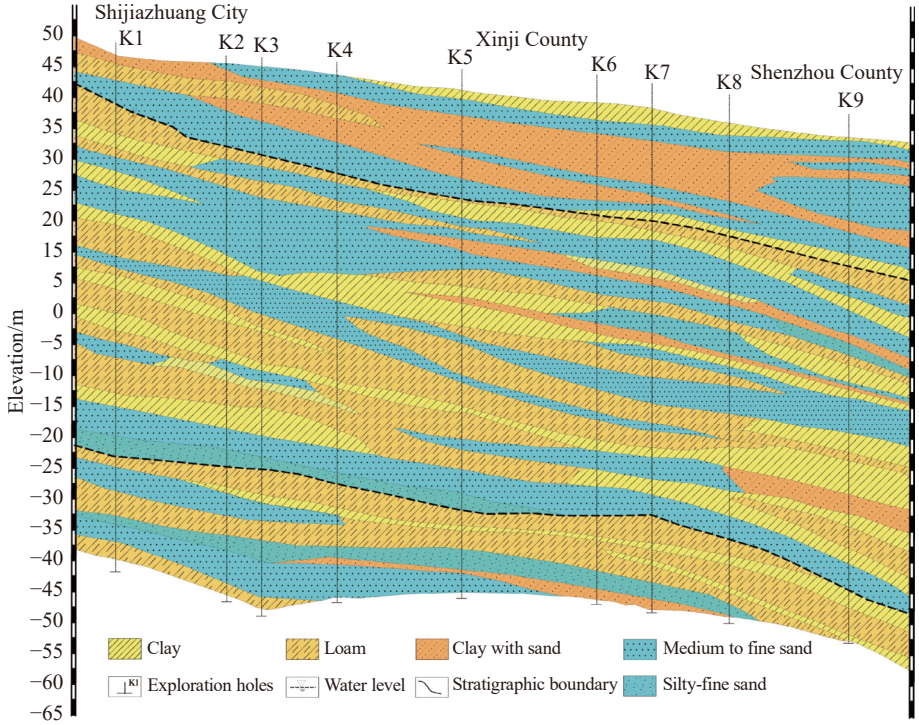

Fig. 3 Hydrogeological profiles

enced by the Neocathaysian tectonic regime. These structural forces have shaped an inland rift basin, which is overlain by Quaternary unconsolidated deposits and thick diluvial strata.

The upper strata of this basin mainly consist of alluvial and aeolian deposits, with dominant lithologies of fine sand, silt, and complex sand-clay interbeds. The sand layers exhibit clear stratification, with individual thicknesses ranging from 20 meters to 80 meters. In the middle section, the strata are characterized by alternating coarse- and fine-grained layers, with lithologies dominated by loam, sandy soil, and interbedded clay.

1.1.2 Hengshui

The stratigraphic sequence in the Hengshui region, from top to bottom, consists of Quaternary, Tertiary, Paleozoic, and Proterozoic layers. The study area lies within the alluvial plain of Hebei Province and features a gently sloping terrain extending from southwest to northeast, with elevations ranging from 12 m to 30 m. The surface gradient slopes eastward at approximately 1/8000 to 1/10,000 and westward at about 1/4000. The area is traversed by numerous rivers, which have shaped the alternating distributions of sediments through flooding events and river diversions. This has resulted in the formation of gentle hills, mildly inclined flats, and low-lying areas.

The lower sections comprise exceptionally thick Quaternary lacustrine deposits, which are predominantly composed of muddy sandy clay, interspersed with localized landforms. The geological features of the region are complex and diverse, reflecting a rich and dynamic geological history. The alternating sedimentary layers, each with distinct characteristics and occasional fossil records, testify to the area's evolving depositional environments.

The predominance of muddy sandy clay in the lower Quaternary lacustrine sediments indicates prolonged wet and humid conditions that favored the accumulation of fine-grained materials. Such conditions likely persisted for extended periods, enabling the formation of thick, continuous sedimentary sequences.

1.2 BP artificial neural network

ANN is sophisticated network system, comprising numerous interconnected simple neurons, designed to mimic the neural processing mechanisms of the human brain. This architecture enables efficient information processing and nonlinear transformations. Among these, the Back Propagation (BP) neural network is a widely used type of unidirectional, multi-layered forward ANN. It typically consists three or more layers: An input layer, one or more hidden layers (also known as the middle layers), and an output layer. The BP neural network learns by propagating errors backward through the network and simultaneously adjusting its parameters to minimize these errors. This iterative correction process allows the network to approximate the desired input-output mapping effectively. Consequently, BP neural networks are well-suited for developing predictive models in groundwater geophysics based on electrical parameters (Lohani et al. 2015). Furthermore, the adaptive learning capability of BP networks enables them to model complex nonlinear relationships inherent in hydrogeological systems. By analyzing patterns in electrical parameters associated with groundwater systems, BP networks can uncover insights that are often difficult to detect using traditional analytical approaches.

The development a BP neural network model for groundwater geophysics involves several key steps. Firstly, a comprehensive dataset must be compiled, encompassing a wide range of electrical parameters and corresponding groundwater characteristics. This dataset forms the foundation for training the network, enabling it to learn the intricate relationships among the variables. Subsequently, the network architecture is designed by specifying the number of layers, the number of neurons within each layer, and the choice of activation functions. These design decisions are pivotal and must be tailored carefully to the specific nature of the problem, as they significantly influence the model's performance. Once the architecture is established, the training phase commences. During training, the network is fed with input data from the dataset, and its outputs are compared against known target values. The errors computed between predictions and actual outputs are propagated backward through the network, allowing the iterative adjustment of network parameters to enhance prediction accuracy progressively. Upon completion of sufficient training, the BP neural network can predict groundwater characteristics from new, unseen input data. These predictions provide valuable insights into groundwater system behaviors and support informed decision-making in groundwater management and exploration activities.

In summary, the BP neural network is a powerful tool for modeling complex, nonlinear relationships in groundwater geophysics. By leveraging its capacity to learn and adapt from data, researchers and practitioners can gain deeper understanding of the intricate processes governing groundwater systems, enabling more informed decisions in their management and exploration.

1.3 Model structure

The BP (Back Propagation) neural network has the advantage of realizing nonlinear mapping from input space to output space without establishing an explicit mathematical model. In this study, resistivity (ρ) of the underground aquifer is determined through multiple linear or nonlinear regression analysis based on experimental and statistical data. The input neurons of the prediction model include polarizability (η), attenuation degree (D) and half decay time (Th), while the predicted output is the unit water inflow (q) of a single hole, which serves as an indicator of groundwater abundance in the survey area (Lin et al. 2010; Liu et al. 2008).

The network architecture employs the widely used three-layer Backpropagation (BP) neural network structure, common in artificial neural network applications. The setup is comprised of three distinct layers: An input layer, a hidden layer, and an output layer. Key components of the network include the basis function, the transfer function, and the training function, all of which play essential roles in its operation. The input layer acts as the gateway for data entry, receiving raw information from external sources. It may preprocess the data as needed before passing it on to the subsequent layers. The hidden layer, located between the input and output layers, performs the majority of the computational work. It consists of interconnected neurons, each with its own weights and biases, which process the incoming signals and extract meaningful features or patterns from the data.

The transfer function, also known as the activation function, plays a pivotal role in the hidden layer. It introduces non-linearity into the model, enabling the network to learn and represent complex relationships within the data. The choice of transfer function can significantly impact the model's learning capabilities and overall perfor-

mance.

The output layer generates the model's predictions based on the processed information from the hidden layer. Depending on the task, the output layer may consist of a single neuron, typically used for regression problems, or multiple neurons for classification tasks. The basis function in this context refers to the mathematical function that defines how the signals from the hidden layer are combined to produce the final output.

The training function manages the entire learning process by iteratively adjusting the weights and biases of the neurons to minimize a predefined loss function. This optimization is generally performed using gradient descent or its variants, updating model parameters in the direction that reduces the error between predicted and actual outputs.

The number of neurons in the hidden layer is related to the number of input and output layers, commonly estimated by the empirical formula:

$$y = \sqrt{mn} + a \tag{1}$$

Where: y represents the number of hidden layer neurons, m and n denote the respective numbers of input and output layer neurons, and a is an arbitrary integer between 1 and 10. To ensure both stability and flexibility of the model, this formula serves only as a guideline; the optimal number of hidden neurons is determined through iterative testing and validation.

The fundamental functionality of the model is linear in nature, wherein the output t represents the weighted summation of the input P and the threshold θ. The transfer function dynamically selects suitable functions based on the operational requirements of the model, introducing the necessary nonlinearity for effective learning. The overall model architecture is illustrated in Fig. 4. In this configuration, the input elements P_i (where i = 1, 2, ⋯) correspond to the distinct sets of resistivity, polarizability, attenuation, and half-attenuation, respectively, while the output element t encompasses the set of predicted unit water inflow quantities Q.

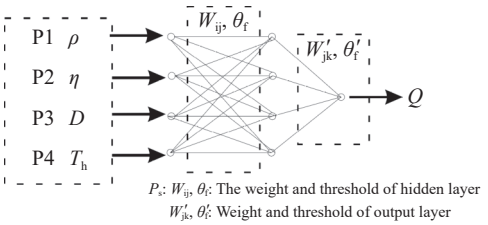

Fig. 4 ANN model

The input-output relationship of the model can be expressed in matrix form as follows:

$$Input: P = \begin{bmatrix} P_{11} & & P_{1n} \\ P_{21} & \cdots & P_{2n} \\ P_{31} & & P_{3n} \\ P_{41} & & P_{4n} \end{bmatrix},$$
$$Output: T = [Q_1 \ Q_2 \cdots Q_n], (n = 1, 2 \cdots) \tag{2}$$

Each column of the Input Neuron Matrix P corresponds directly to the associated water inflow values Q_i in the output Neuron Matrix T.

1.4 Normalization

Due to the differences in the units and scales of the input neuron data, the numerical values across various datasets may differ significantly, sometimes spanning multiple orders of magnitude. Consequently, it is essential to perform normalization of the input layer data within the range [0,1]. This study applies a linear function transformation, expressed as follows:

$$y = (x - \mathrm{Min}(x))/(\mathrm{Max}(x) - \mathrm{Min}(x)) \quad \text{(Song et al. 2012)} \tag{3}$$

Where: x represents the value prior to conversion, and y represents the normalized value. Here, Max(x) and Min(x) signify the maximum and minimum values within the given sample set, respectively.

The parameters, resistivity, half-decay time, decay degree, and polarizability, were obtained through geophysical exploration near the boreholes. For this study, a total of 16 boreholes within the study area were selected, including exploration boreholes (designated with "K") and monitoring boreholes (designated with "J"). The geophysical parameters were derived from measurements conducted adjacent to each borehole, and the corresponding water inflow values were collected from drilling data. After applying normalization to the input parameters, the normalized data are presented in Table 1. Since the output is a single value (unit water inflow) with only minor differences in magnitude, normalization of the output neuron was not required (Janič et al. 2019; Sanchez-Vila et al. 2006; Song et al. 2018).

1.5 Prediction result

After meticulously debugging, it was determined that the model's transfer function adopts S-type tangent function (tansig) and the linear function (purelin). The training function adopts the trainlm algorithm, which leverages the Levenberg-Marquardt method, a full-scale optimization algorithm known for its fast convergence and high predic-

tion accuracy. Six boreholes from the datasets were selected as test samples, while the remaining boreholes were use as training samples to generate the prediction results shown in Table 2. Upon examination, it is evident that the performance of this model does not yet meet expectations.

Despite the initial promise of the model structure and the advanced algorithms employed, the discrepancies between the predicted results and the actual pumping test data suggest that further refinement is required. These deviations may be attributed to several factors, including the complexity of the geological setting, fluctuations in groundwater levels, or potential uncertainties in the input parameters.

1.6 Deviation parameter

In the application of the model and during subsequent research, it has been discovered that acquiring reliable electrical parameters under complex geological conditions requires careful consideration of various external factors, which cannot be overlooked.

For example, groundwater salinity can significantly influence apparent polarizability and apparent resistivity values, while poor grounding conditions may distort electrical parameters. To address these challenges and improve the robustness and stability of the prediction model, the concepts of deviation degree and comprehensive parameters have been introduced through systematic research and analysis. The detailed analysis is as follows (Almuhaylan et al. 2020; Cheng et al. 2022).

Numerous practical investigations and research efforts have demonstrated that, in the application of the Direct Current Induced Polarization (DCIP) method, there inevitably exists a measurable deviation between the empirically obtained discharge curve and the theoretical "ideal" straight line. This deviation refers specifically to the discrepancy between the actual discharge curve and the

Table 1 Normalized data of input neuron

Hole number	ρ/Ω·m	Th/s	D/%	η/%	Q/m³/h
K1	1.3	1.2	0.730	1.1	57
K2	0.879	0.689	0.760	0.635	67
K3	0.687	0.569	0.820	0.605	36
K4	0.333	0.433	0.670	0.675	38
K5	0.881	0.373	0.627	0.883	58
K6	0.665	0.661	0.650	0.582	45
K7	0.480	0.320	0.355	0.240	79
K8	0.468	0.450	0.589	0.579	60
K9	0.312	0.355	0.755	0.079	63
J1	0.303	0.23	0.534	0.020	34
J2	0.223	0.231	0.509	0.587	41
J3	0.456	0.352	0.405	0.668	53
J4	0.423	0.155	0.680	0.502	51
J5	0.934	0.786	0.620	0.872	83
J6	0.218	0.556	0.601	0.618	55
J7	0.33	0.283	0.630	0.212	42

Table 2 Comparison of training error (all input data has been normalized)

Hole number	ρ/Ω·m	Th/s	D/%	η/%	Actual Q/m³/h	Predicted Q/m³/h
K1	1.3	1.2	0.730	1.1	57	60.8165
K2	0.879	0.689	0.760	0.635	67	62.5421
K3	0.687	0.569	0.820	0.605	36	39.3233
J5	0.934	0.786	0.620	0.872	83	78.3232
J6	0.218	0.556	0.601	0.618	55	51.6692
J7	0.33	0.283	0.630	0.212	42	44.2

expected straight-line behavior and is quantitatively described by the mean square relative error, denoted as σ. The deviationσis calculated using the following formula:

$$\sigma = \frac{1}{\overline{\eta_i}} \sqrt{\frac{\sum_{i=1}^{n} \left(\eta_i + K\log t_i - B\right)^2}{n}} \times 100\% \quad (4)$$

In the formula, n represents the number of sampling points, and $\overline{\eta_i} = \frac{1}{n}\sum_{i=1}^{n}\eta_i$ is the average deviation of the polarizability for each sampling point in the observed time period, representing the degree of deviation from the theoretical "Ideal" line. It is evident that a smaller σ corresponds to greater linearity exhibited by the measured discharge curve.

Extensive research and statistical analysis have shown that the deviation (σ) value is physically correlated with several influencing factors, such as the duration of power supply, the humidity during formation, the concentration and composition of pore fluid, and the clay content of the formation. Notably, statistical comparisons (Table 3) between polarization rate and the degree of deviation illustrate that the degree of deviation is less susceptible to external factors than the polarization rate alone. Furthermore, it exhibits a stronger correlation with the unit water inflow, thereby enhancing the stability and robustness of the predictive model (Hansen et al. 1997; Goldman et al. 1994; Hálek et al. 2011; Hiskiawan et al. 2023).

Table 3 Correlation analysis of electrical parameters specific capacity

Analytic quantity	Neurons	
	σ	*η*
Correlation (r)	0.969	0.886
Coefficient of significance (p)	0.982	0.819

1.7 Composite parameter

Given the occurrence of the equivalence phenomenon uring the inversion interpretation of electrical sounding, coupled with the generally high resistivity observed in Quaternary aquifers, there is an increased risk of generating interpretation errors related to this T equivalence effect. Additionally, an aquifer's resistivity is inherently linked to its specific composition and structure, whereas the unit water inflow per hole is not only influenced by these factors but also by the aquifer's thickness.

To minimize interpretation errors and enhance the correlation between input neurons and the prediction quantity, thereby improving the model's generalization capabilities, the aquifer resistivity obtained from electrical sounding and inversion is multiplied by its respective layer thickness and employed as an input neuron. However, due to the variable salinity levels of groundwater across different regions and the distinct conductivity of pore water, the resistivity of similar water-rich strata can still differ significantly.

To mitigate the impact of pore water conductivity on prediction precision and to highlight the dominant porosity characteristics of the aquifer, a composite parameter is introduced, based on Archie's formula. The derived expression is given as below:

$$T'' = \frac{T'}{\rho'} = \frac{\rho}{\rho'}\mathrm{H} \quad (5)$$

Where: ρ is aquifer resistivity at measuring point; ρ' is aquiclude resistivity at measuring point; H is aquifer thickness.

Thus, the original primary input neuron and H have been integrated into a single composite input neuron. This new parameter effectively strengthens the correlation with the unit water inflow per borehole, as illustrated in Table 4 (Batu, 1998; Bear, 2012; Krenker, 2011).

Comprehensive parameter T'' = ρ×h/ρ'. It is constructed by combining the resistivity (ρ and ρ') with the thickness of the aquifer (h). This combination approach captures the characteristics of the aquifer from multiple perspectives:

Resistivity ρ and ρ': They represent the resistivity of the aquifer at the measuring point and the reference resistivity, respectively. By comparing their ratio, the numerical difference in absolute resistivity caused by variations in groundwater salinity can be effectively eliminated.

By integrating these factors, the composite parameter T" reduces equivalence errors commonly encountered in the inversion of electrical sounding data and significantly enhances its correlation with the unit water inflow (r=0.992, Table 4).

Additionally, the deviation degree (σ), calculated from polarization decay curves, quantifies the stability of the measurement and further improves the robustness of the prediction model (p<0.05).

Table 4 Correlation and significance analysis

Analytic quantity	Neurons		
	ρ	**H**	T''
Correlation (r)	0.893	0.866	0.992
Coefficient of significance (p)	0.998	0.977	0.989

The aquifer thickness (H) can be obtained by inverting and interpreting the resistivity and polarization data required through resistivity sounding or induced polarization sounding, as shown in Fig. 5.

1.8 Model architecture

1.8.1 Input variables

The improved model uses four parameters as input features: Half-life (Th), Decay rate (D), Deviation (σ), Comprehensive parameter (T"=ρ h /ρ').

Output variable: Single-hole unit flow rate (Q).

1.8.2 Network type

Multilayer Perceptron (MLP) regression model is adopted, consisting of an input layer, two hidden layers, and an output layer.

Input layer: 4 neurons (corresponding to 4 input parameters)

Hidden layer 1: 8 neurons (ReLU activation function)

Hidden layer 2: 4 neurons (ReLU activation function)

Output layer: 1 neuron (linear activation function)

1.8.3 Model calibration cycle

Maximum iterations: 500 epochs

Early Stopping Mechanism: Training stops if the validation loss does not decrease for 20 consecutive epochs.

Dynamic Learning Rate: If the loss plateaus for more than 10 epochs, the learning rate decays to 10% of its initial value.

1.8.4 Calibration parameters

Number of neurons in hidden layers: Optimized through grid search:

L2 Regularization Coefficient: $\lambda \in [0.001,0.1]$

Dropout Rate: Set to 0.2 during validation phase to prevent overfitting.

Batch Size: Comparative testing with 8, 16 and 32.

1.8.5 Model validation cycle

Cross-Validation: Five-fold cross-validation is used to calculate the average R^2 and MAE. The network weights are reset (He initialization) for each validation.

Sensitivity Analysis: The influence of each input parameter on Q is quantified using SHAP values. The physical correlation between T" parameter and Q must be verified (require $|r|\geq 0.6$).

Field Validation: Three additional boreholes are selected for pumping tests in the field. The prediction error must satisfy: $|(Q_{\_pred}-Q_{\_obs})/Q_{\_obs}| \leq 25\%$

1.8.6 Model accuracy evaluation criteria

Quantitative Indicators:

Root Mean Square Error (RMSE) ≤0.5 $m^3/(d\ m)$.

Mean Absolute Percentage Error (MAPE) ≤ 15%.

Determination Coefficient (R^2) ≥0.85 (Test Set)

Residual Analysis:

Residuals must pass the normality test (K-S test $p>0.05$).

Residual distribution: 95% of prediction errors should fall within the ±1.96 σ range.

Engineering Applicability:

Classification accuracy for discharge predictions (categorized as <5, 5–10, >10 $m^3/(d\ m)$) ≥ 90%.

Parameter sensitivity ranking must align with hydrogeological principles (the weight of T" should be higher than that of σ)

1.9 Effect analysis of the improved model

In the refined model, the input neurons consist of

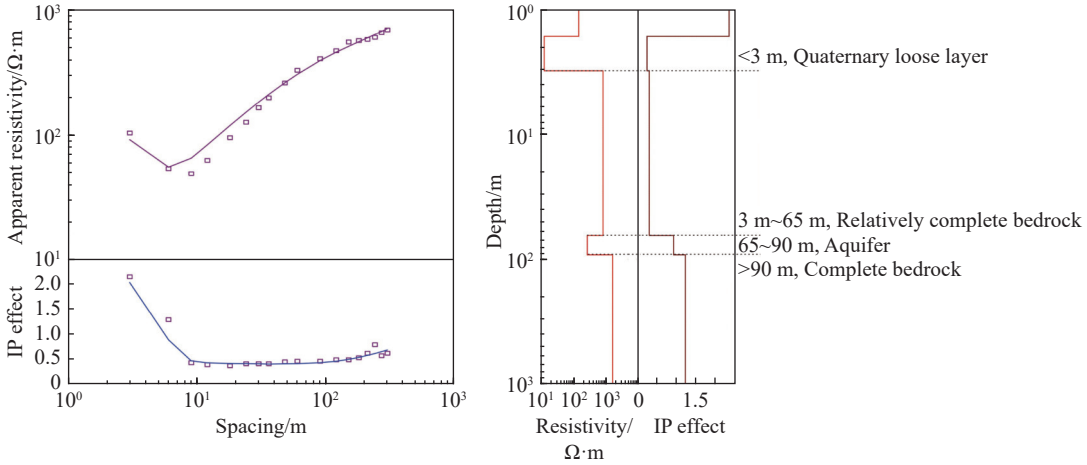

Fig. 5 IP sounding inversion curve

decay time (Th), attenuation degree (D), deviation degree (σ), and the composite parameter (T"), while the output neurons represent the unit water inflow of a single borehole. Validation tests were conducted using three established pumping wells in Shijiazhuang and three additional wells in Hengshui. The predicted convergence curve, as depicted in Fig. 6, illustrates that the model reached the target error threshold at step 276, indicating stable convergence. The prediction results, presented in Table 5, show a notable improvement in prediction accuracy compared to the initial model, confirming that the enhancements are effective and satisfactory.

Notably, the inclusion of comprehensive parameters as input neurons has significantly strengthened the model's performance by better capturing the combined effects of a broader range of influencing factors. This comprehensive approach enables the model to produce more accurate and reliable estimates of unit water inflow under varying geological and hydrogeological conditions. As a result, the improved model provides a robust tool for practical applications, offering enhanced precision and generalization for groundwater resource evaluation and management.

The prediction outcomes are presented in Table 5, and a noteworthy improvement in prediction accuracy is evident, underscoring the effectiveness of the advancements achieved in the updated model. Moreover, the enhancements to the model's architecture have led to a more robust performance across various test cases, demonstrating its ability to handle complex and variable data with greater precision. This improvement in accuracy is particularly critical for real-world applications, where even minor prediction errors can have significant hydrogeological and engineering implications.

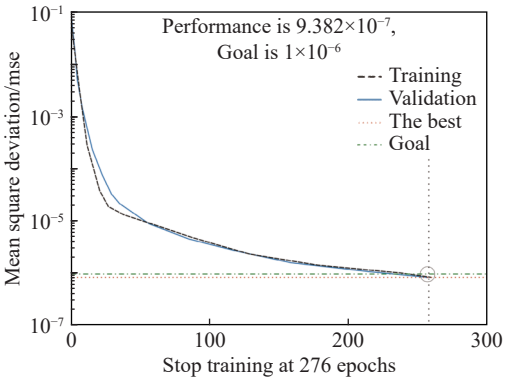

Fig. 6 Errors convergence curve

Table 5 Forecast results of new model

Borehole names	Actual Q/ m^3/h	Predicted Q/ m^3/h	Error/ %
K1	57	55.232	3.1018
K2	67	66.363	0.9507
K3	36	34.582	3.9389
J5	83	84.367	1.647
J6	55	52.563	4.4309
J7	42	41.368	1.5048

2 Results and discussion

2.1 Results

To further validate the effectiveness of the improved model, a detailed comparative analysis was conducted against the previous version. The data presented in Table 6 clearly demonstrate that the predictive accuracy of the new model, enhanced by the inclusion of additional input neurons and optimized parameters, has significantly improved compared to the origional model. Specifically, the previous prediction error rates, which ranged from 5% to 10%, have been substantially reduced to 0.9% to 5%. This remarkable improvement highlights the critical role of the refined model architecture and the introduction of the composite and deviation parameters in boosting prediction prediction.

Additionally, it is imperative to highlight the enhanced stability of the improved model. This is evidenced by the consistent convergence behavior illustrated in Fig. 6, where the model reliably reaches the predefined error target within a reasonable number of iterations. Such convergence performance is indicative of the model's robustness and reliability. This level of stability is of paramount importance for practical applications, as it ensures that the model can deliver timely and accurate predictions, thereby supporting effective decision-making in real-world hydrogeological

Table 6 Comparison of the prediction performance between the original and the improved models

Hole number	Prediction error of original model /%	Prediction error of new model /%
K1	6.70	3.10
K2	6.56	0.95
K3	9.23	3.94
J5	5.63	1.65
J6	6.06	4.43
J7	5.24	1.50

assessments.

The model prediction outcomes clearly indicate that the comprehensive parameter (T"), apparent resistivity (ρS) and deviation degree (σ) are pivotal factors in forecasting single hole water inflow (Fig. 7). These factors, when considered collectively, provide a robust understanding of the subsurface conditions that govern unit water inflow. Notably, the T" parameter plays a critical role in enhancing prediction precision by capturing multiple factors that influence fluid flow dynamics within the aquifer. The interplay between T" and ρS underscores the necessity for a nuanced analytical approach in their analysis. Resistivity, which measures the opposition to electric current flow, is a fundamental property that can significantly influence the permeability and, consequently, the water inflow rate. In contrast, the comprehensive parameter T", which integrates multiple factors affecting fluid flow dynamics, adds an additional layer of complexity to the prediction model.

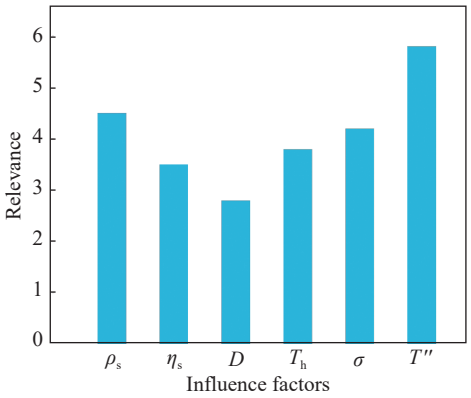

Fig. 7 Importance analysis of various influence factors

The deviation degree (σ), though less directly related to physical hydrogeological properties, serves as an essential indicator of data quality and model uncertainty. By quantifying the degree to which the measured discharge curve deviates from the theoretical ideal, σ provides insights into the stability and reliability of the geophysical measurements. By incorporating σ as an input neuron, the model can better capture and quantify potential prediction errors, thereby improving the overall accuracy and reliability of the forecasting system.

In addition to these key input factors, the improved model's architecture and normalization strategy also play crucial roles in delivering the accurate forecasts. The well-structured network design ensures that the neural network can learn complex, nonlinear relationships effectively, while input data normalization enhances convergence speed and training stability. Together, these improvements contribute to the overall performance and reliability of the prediction model.

2.2 Discussion

Understanding the specific contribution of each factor and how they interact is crucial for developing accurate and reliable predictive models. Therefore, further research into the underlying mechanisms governing these parameters and their relationships could lead to significant improvements in the prediction accuracy. It is important to delve deeper into the science basis of these factors to better understand how they influence each other and, consequently, the overall water inflow.

Additionally, it is essential to consider the spatial and temporal variability of these factors. Variations in the geological formations, water table levels, and other environmental conditions can significantly affect the values of T" and resistivity, thereby impacting the water inflow predictions. Incorporating such dynamic factors into the model would help ensure that predictions remain robust and applicable across a wider range of scenarios. This adaptability allows the model to respond to changes over time and space, providing more accurate predictions in diverse environments.

Moreover, the inclusion of historical data on water inflow and related parameters could enhance the model's capability to capture long-term trends and seasonal variations. This would improve the anticipation of future water inflow conditions, particularly in regions prone to extreme weather events or cyclical changes. Furthermore, integrating real-time monitoring systems with the predictive model could provide timely updates and continuous adjustments, ensuring the model remains current and reflective of prevailing conditions. Such advancements would not only improve the precision of water inflow predictions but also support more informed decision-making in groundwater management and resource allocation.

Inflow underscores the importance of a multifaceted approach in groundwater modeling. This analysis demonstrates that relying solely on traditional methods may not adequately capture the full complexity of groundwater system. By integrating advanced parameters such as T" and considering the dynamic nature of resistivity, we can move towards more sophisticated and accurate predictive tools. This understanding paves the way for future research to delve deeper into the interactions among these parameters and to explore inno-

vative ways of incorporating them into groundwater modeling frameworks. Ultimately, the goal is to develop models that not only predict water inflow with high precision but also offer valuable insights into the underlying geological and hydrological processes, thereby enhancing our overall understanding of groundwater dynamics.

In conclusion, recognizing T" and resistivity as two of the most critical factors in predicting single hole water inflow represents a significant advancement in the field. When these factors are properly understood and effectively utilized, they can greatly enhance the precision and reliability of predictive models. Continued research and development in this area are therefore essential to further improve the accuracy and applicability of water inflow predictions.

It should also be noted that even with algorithmic optimization, small-sample learning remains inherently sensitive to data quality. In groundwater systems, for example, regions characterized by strong heterogeneity, such as karst aquifers, may not be adequately captured by a limited number of samples due to the dynamic and complex nature of their flow patterns. Furthermore, many existing methods largely rely heavily on purely data-driven approaches and often lack the explicit embedding of hydrogeological mechanisms, which can introduce biases when making extrapolative predictions beyond the training domain.

The next step could involve exploring transfer learning techniques, which allow models to draw on parameters from data-rich areas and fine-tuning them to adapt to the target areas with limited data. For example, a groundwater model trained in a well-monitored, water-abundant region could be transferred to other hydrogeological units and fine-tuned using a small amount of local data, thereby improving its generalization capability. Additionally, combining different modelling approaches, such as Gaussian processes, SVMs, and random forests, could enhance prediction stability under small-sample conditions by leveraging the complementary advantages of these heterogeneous models.

For few-shot learning problems, the limitations of existing methods in data-scarce scenarios are becoming increasingly evident. Dynamic changes in environmental conditions, such as geological structures and groundwater levels, can significantly alter the values of key parameters like T and resistivity. A small number of samples often struggle to capture the full range of dynamic variation patterns, making purely data-driven methods prone to extrapolation prediction biases and potentially affecting the accuracy of water yield predictions. However, it is important to note that the study area selected for this research, a stable and relatively homogeneous Quaternary sedimentary plain, has demonstrated consistently high prediction accuracy through drilling validation, and the aforementioned issues have not yet been encountered.

Therefore, there is considerable scope for expanding the validation and application of the model. For instance, future work could compare and analyze prediction results against a broader range of actual drilling data, particularly across diverse geological conditions and hydrogeological units, to more comprehensively assess the model's precision and feasibility. Additionally, integrating real-time monitoring systems to develop time-series-based prediction models could allow timely updates and adjustments of model parameters, better reflecting current groundwater conditions and further improving prediction accuracy. This would provide stronger support for informed water resource management and allocation decisions.

This study also highlights the value of incorporating dynamic factors into the model. Compared to traditional methods that rely solely on static factors, this approach better accommodates the complexity and variability of real-world groundwater environments, thereby enhancing the model's robustness and broader applicability. In highly heterogeneous regions, transfer learning presents a promising solution. Several studies, particularly in fields such as natural language processing, have demonstrated that transferring pre-trained model parameters to target tasks can significantly improve few-shot learning performance. The transfer learning strategy proposed in this study, which leverages model parameters from data-rich regions and fine-tunes them for the target area, holds strong potential for improving prediction stability in data-scarce conditions. Furthermore, integrating heterogeneous models such as Gaussian processes, support vector machines, and random forests could capitalize on their respective strengths, further enhancing prediction stability and accuracy.

In summary, this study has made meaningful progress in key factor identification, model construction, and validation. However, limitations remain. By comparing results with existing research and analyzing drilling results, the precision and practical feasibility of this study have been further demonstrated. Future research should continue to deepen the exploration of these approaches, refine the models and methods, and improve the accuracy and reliability of water yield predictions, ultimately providing a stronger scien-

tific foundations for effective groundwater system management and protection.

3 Conclusions

1. Few-shot learning demonstrates considerable potential for groundwater prediction, particularly in data-scarce areas or regions with limited historical observation records. Through algorithm optimization, data augmentation, and model structure improvement, it is entirely possible to enhance prediction accuracy and generalization capability. The verification and generalization tests conducted in this study confirm that the improved model can quantitatively predict groundwater inflow in the survey area. Based on the modeling results from the Taihang Mountain piedmont plain in Shijiazhuang City to the inland plain area of Hengshui City in the North China Plain, the model with optimized input neurons exhibits superior generalization capabilities. Notably, the incorporation of comprehensive parameters effectively mitigates the effects of mineralization factors, significantly enhancing the model's capacity for generalization.

2. Small sample learning has proven highly practical for groundwater prediction through algorithmic innovation and advanced data processing strategies. In the future, it will be essential to further integrate physical mechanisms with multi-source data and to explore cutting-edge technologies such as transfer learning. These developments can improve the model's robustness, reliability and interpretability in scenarios where data availability is severely limited.

3. It must be emphasized that predictive modeling in the field of geophysical prospecting remains constrained by the inherent multi-solution nature of subsurface geophysical problems. Currently, only a limited number of geophysical prospecting methods are suitable for use in predictive modeling, and these alone are insufficient to address the wide range of geological complexities encountered in practice. Therefore, future research should embrace a more diverse and integrated approach. By integrating multiple geophysical techniques, researchers can gather a more comprehensive and robust set of relevant parameters. This holistic strategy will not only improve the accuracy and reliability of predictive models but also broaden their applicability to complex hydrogeological settings. Ultimately, these advancements will enable planners and practitioners in hydrogeological surveying and land-use planning to make more informed decisions, thereby supporting the sustainable management and development of groundwater resources.

Ackonwledgements

This work was financially supported by the National Natural Science Foundation of China (No. 42277460), the Fundamental Research Project of Institute of Hydrogeology and Environmental Geology, Chinese Academy of Geological Sciences (No. SK202210).

References

Almuhaylan MR, Ghumman AR, Al-Salamah IS, et al. 2020. Evaluating the impacts of pumping on aquifer depletion in arid regions using MODFLOW, ANFIS and ANN. Water, 12(8): 2297. DOI: 10.3390/w12082297

Batu V. 1998. Aquifer hydraulics: A comprehensive guide to hydrogeologic data analysis. John Wiley & Sons.

Bear J. 2012. Hydraulics of groundwater. Courier Corporation.

Behzad M, Asghari K, Coppola Jr EA. 2010. Comparative study of SVMs and ANNs in aquifer water level prediction. Journal of Computing in Civil Engineering, 24(5): 408-413. DOI: 10.1061/(ASCE)CP.1943-5487.0000043

Cheng W, Dong F, Tang R, et al. 2022. Improved combination weighted prediction model of aquifer water abundance based on a cloud model. ACS Omega, 7(40): 35840−35850. DOI：10.1021/acsomega.2c04162.

Davydycheva S, Rykhlinski N, Legeido P. 2006. Electrical-prospecting method for hydrocarbon search using the induced-polarization effect. Geophysics, 71(4): G179-G189. DOI: 10.1190/1.2217367

Devi TG, Rajkumar G, Srinivasan A, et al. 2022. Radial basis function neural network and salp swarm algorithm for paddy leaf diseases classification in Thanjavur, Tamilnadu geographical region. Acta Geophysica, 70(6): 2917–2932. DOI：10.1007/s11600-022-00865-w.

Dong Y, Ma Y. 1996. Pattern recognition of hte characteristics of AE source using neural network proceedings of the 14th word conference on NDT v. 4, New Delhi.

Duragasi AR, Sukumaran JV. 2023. Geo-electrical resistivity–based study to estimate the aquifer parameters—a study in parts of Yamuna Nagar District, Haryana State, North India. Arabian Journal of Geosciences, 16(10): 584. DOI: 10.1007/s12517-023-11691-9

Ghosh R, Sutradhar S, Das N, et al. 2023, A comparative evaluation of GIS based flood susceptibility models: A case of Kopai river basin, Eastern India. Arabian Journal of Geosciences, 16(11): 591. DOI：10.1007/s12517-023-11693-7.

Goldman M, Neubauer F. 1994, Groundwater exploration using integrated geophysical techniques. Surveys in Geophysics, 15: 331–361. DOI：10.1007/BF00665814.

Gyeltshen S, Kannaujiya S, Chhetri IK, et al. 2023. Delineating groundwater potential zones using an integrated geospatial and geophysical approach in Phuentsholing, Bhutan. Acta Geophysica, 71(1): 341–357. DOI：10.1007/S11600-022-00856-X.

Hálek Václav, Jan Švec. 2011. Groundwater hydraulics. 7. Elsevier.

Hansen R. 1997. Feature recognition from potential fields using neural networks. Geophysics, 62(3): 806–817. Available on https://www.pqdtcn.com/thesisDetails/E91274A10FEB4108D4FA530524A984F4

Hayashi Y, Buckley JJ, Czogala E. 1993. Fuzzy neural network with fuzzy signals and weights. International Journal of Intelligent Systems, 8(4): 527–537. DOI：10.1002/int.4550080405.

Hiskiawan P, Chen CC, Ye ZK. 2023. Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures. Arabian Journal of Geosciences, 16(10): 581. DOI：10.1007/s12517-023-11690-w.

Janič P, Jadlovská S, Zápach, J, et al. 2019. Modeling of underground mining processes in the environment of MATLAB/Simulink. Acta Montanistica Slovaca, 24(1): 1335−1788.

Krenker A, Bešter J, Kos A. 2011. Introduction to the artificial neural networks. Artifiical Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18.

Li S, Bian H, Zhang D, et al. 2024, Research on pore structure and classification evaluation of tight oil reservoirs based on fractal theory. Acta Geophysica, 1-11. DOI: 10.1007/s11600-024-01299-2

Liu GH, Zhang XM, Jia XM. 2007. New technology of groundwater re-sources electrical prospecting. Beijing: Earthquake Press.

Liu L, Chen J, Xu L. 2008. Realization and application research of BP neural network based on MATLAB, in Proceedings 2008 International Seminar on Future BioMedical Information Engineering, IEEE, 130–133. DOI：10.1109/FBIE.2008.92.

Liu S, Li X, Wang H. 2011. Hydraulics analysis for groundwater flow through permeable reactive barriers. Environmental Modeling & Assessment, 16: 591–598. DOI：10.1007/s10666-011-9268-0.

Lin HT, Ke KY, Chen CH, et al. 2010. Estimating anisotropic aquifer parameters by artificial neural networks. Hydrological Processes, 24(22): 3237−3250. DOI：10.1002/hyp.7750.

Lin J, Lin T, Ji Y, et al. 2013. Non-invasive characterization of water-bearing strata using a combination of geophysical techniques. Journal of Applied Geophysics, 91: 49−65. DOI：10.1016/j.jappgeo.2013.02.002.

Ling CP, Sun YJ, Yang LH, et al. 2007. Prediction of inrush water of mine with pore water yield based on BP artificial neural network. Hydrogeology & engineering geology, 34(5): 55–58. DOI：10.16030/j.Cnki.Issn.1000-3665.2007.05.010.

Lohani A, Krishan G. 2015. Groundwater level simulation using artificial neural network in southeast Punjab, India. Journal of Geology and Geosciences, 4(3): 206. DOI：10.4172/2329-6755.1000206.

Maiorov K, Vachrusheva N, Lozhkin A. 2021. Solving problems of the oil and gas sector using machine learning algorithms. Acta Montanistica Slovaca, 26(2): 327−337. DOI：10.46544/AMS.v26i2.11.

Maiti S, Erram VC, Gupta G, et al. 2012. ANN based inversion of DC resistivity data for groundwater exploration in hard rock terrain of western Maharashtra (India). Journal of Hydrology, 464: 294−308. DOI：10.1016/j.jhydrol.2012.07.020.

Mnasri M, Amiri A, Nasr IH, et al. 2023. Integrated geophysical approach for ore explo-

ration: Case study of Sidi Bou Aouane–Khadhkhadha Pb–Zn province–Northern Tunisia. Geophysical Prospecting, 71. Advanced Techniques, Methods and Applications for an Integrated Approach to the Geophysical Prospecting, 1772–1791. DOI：10.1111/1365-2478.13338.

Patra H, Adhikari SK, Kunar S. 2016. Groundwater prospecting and management, Springer. DOI:10.1007/978-981-10-1148-1

Persico R. 2023. Introduction to the special issue on 'Advanced techniques, methods and applications for an integrated approach to the geophysical prospecting'. Geophysical Prospecting, 1693–1695. DOI：10.1111/1365-2478.13423.

Sanchez-Vila X, Guadagnini A, Carrera J. 2006. Representative hydraulic conductivities in saturated groundwater flow. Reviews of Geophysics, 44(3): RG3002. DOI：10.1029/2005RG000169.

Saravanan KS, Bhagavathiappan V. 2024. Prediction of crop yield in India using machine learning and hybrid deep learning models. Acta Geophysica, 1-20. DOI: 10.1007/s11600-024-01312-8

Scanlon BR. 2000. Uncertainties in estimating water fluxes and residence times using environmental tracers in an arid unsaturated zone. Water Resources Research, 36(2): 395-409. DOI: 10.1029/1999WR900240

Sivakrishna K, Ratnam DV, Sivavaraprasad G. 2022. Support vector regression model to predict TEC for GNSS signals. Acta Geophysica, 70(6): 2827–2836. DOI：10.1007/s11600-022-00954-w.

Song H, Mu H, Xia F. 2018. Analyzing the differences of brackish-water in the Badain Lake by geophysical exploration method. Journal of Groundwater Science and Engineering, 6(3): 187−192. DOI：10.19637/j.cnki.2305-7068.2018.03.004.

Song H, Xia F, Mu H, et al. 2020. Study on detecting spatial distribution availability in mine goafs by ultra-high density electrical method. Journal of Groundwater Science and Engineering, 8(3): 281−286. DOI：10.19637/j.cnki.2305-7068.2020.03.008.

Xia F, Song H, Wang M, et al. 2019. Analysis of prospecting polymetallic metallogenic belts by comprehensive geophysical method. Journal of Groundwater Science and Engineering, 3: 237−244. DOI：10.19637/j.cnki.2305-7068.2019.03.004.

Yilmaz OS. 2022. Flood hazard susceptibility areas mapping using Analytical Hierarchical Process (AHP), Frequency Ratio (FR) and AHP-FR ensemble based on Geographic Information Systems (GIS): A case study for Kastamonu, Türkiye. Acta Geophysica, 70 (6): 2747−2769. DOI：10.1007/s11600-022-00882-9.

Zhong S, Wang Y, Zheng Y, et al. 2021. Electrical resistivity tomography with smooth sparse regularization. Geophysical Prospecting, 69 (8-9): 1773−1789. DOI：10.1111/1365-2478.13138.