RESEARCH ARTICLE

Application of an integrated physical exploration method based on Gaussian hybrid clustering in the study of river evolution patterns

Yu He, Hai Guo*, Yu Li, Qiang Li, Jie Peng

Geological Survey Institute of Hunan Province, Changsha, Hunan, China.

Received: December 10, 2024; **accepted:** March 21, 2025.

Qingjiang River located in Hubei Province, China is one of the tributaries of the Yangtze River. In recent years, some progress has been made in the control of Qingjiang River, but pollution and water pressure still exist. To meet the demand of the water conservancy department for the comprehensive management of Qingjiang River including Qingjiang current section, this research put forward the joint profile method and high-density electric method based on Gaussian hybrid clustering algorithm to explore the evolution law of river with the help of engineering cases. The results showed that, in the southern district, the S2 geophysical profile had only the joint profile method, which the section orientation was 120, joint fault plane was AO = 110 m. Although it was not very thick and roughly the same as the joint fault plane of AO = 200 m, the abnormal rendezvous parts were roughly the same. In the S3 intersection, in two directions, AO = 110 m and 200 m, respectively. The vertical distribution law of the two was consistent with each other. In the north district, joint section was AO = 110 m. Despite its measured thickness, its overall evolution characteristics were consistent with those of AO = 200 m. There was also a very good consistency in its corresponding parts. From the N3 joint profile view, in two directions, AO = 110 m and 200 m, respectively. The distribution law of the two in space was consistent too. When applying the high-density electric method to distinguish the river evolution characteristics, it should be based on the geological conditions and the exploration crossing tectonic or lithological boundaries. With the joint section orthogonal intersection as the key target, combined with the comparative analysis of the high-density electrical methods, the river evolution law in this region could be effectively located, which was important for the exploration of the similar areas.

Keywords: Gaussian hybrid clustering; integrated physical exploration; river evolution; water resources.

***Corresponding author:** Hai Guo, Geological Survey Institute of Hunan Province, Changsha 410014, Hunan, China. Email: 13786531070@163.com.

Introduction

Qingjiang River located in Hubei Province, China originating from Enshi region covers a basin area of about 24,000 square kilometers and is an important tributary of the Yangtze River. Its water resources are crucial to agriculture, industry, and drinking water in the surrounding areas with a beautiful ecological environment and attracting a large number of tourists. Pollution, overdevelopment, and flood risk in the Qingjiang River basin exacerbate water quality deterioration and ecological damage. Water pollution, flood prevention, and control and over-exploitation of water resources are the main problems faced at present. With resource utilization, ecological protection, and disaster prevention and control interrelated, the Qingjiang River needs comprehensive management. At present, water quality

monitoring, ecological restoration, flood management, and rational multidimensional management of water resources should be strengthened to ensure sustainable development [1]. River exploration provides basic data for integrated management and helps us to understand hydrological, ecological, and environmental conditions. The information obtained through exploration can optimize the management measures of water resources allocation, flood control and disaster prevention, ecological protection and so on, and realize the scientific and accurate comprehensive watershed management.

Currently, the understanding of river evolution law mainly includes geological drilling, high density surface wave, high density electric method, geological radar, elastic wave computed tomography such as Elastic Wave Computed Tomography, and collision formation method [2]. Comprehensive geophysical methods have been widely used in geological exploration, but the results have individual differences. He *et al*. applied physical experiments with the optimal physical test data processing technology as the entry point and studied the multi-object joint detection technology based on machine learning to realize the joint interpretation of the detection results of complex geological bodies [3]. In recent years, with the continuous development of machine learning methods, machine learning has been outstanding in data classification and data mining. Many scholars have introduced machine learning methods into civil engineering-related fields and achieved meaningful results by using machine learning as a means of data processing [4]. Deng *et al*. proposed a wind farm equivalent modelling method based on Gaussian hybrid model clustering for the problem of low accuracy and insufficient clustering basis of equivalent modelling of wind farms under the dynamic operation conditions of wind farms [5]. Dong *et al*. incorporated the idea of a shaded set into the three-branch Gaussian mixture model and proposed a three-branch Gaussian mixture clustering algorithm based on a shaded set [6]. The Gaussian mixture model (GMM) belongs to the machine learning "unsupervised learning" analysis method with unknown label or physical meaning of the training samples. Through the establishment of a Gaussian mixture model of unlabeled training samples, the intrinsic laws of the data can be revealed to achieve a series of unlabeled samples of the data classification process [7].

To meet the requirements of comprehensive management of Qingjiang River including the Qingjiang River volumetric section from the government water conservancy department, this research proposed a joint profile method and high-density electric method based on Gaussian hybrid clustering algorithm. Based on this proposed method, the river evolution law of Qingjiang River Basin was investigated with a view to providing a guiding significance to the research of the river evolution survey in similar areas.

Materials and methods

High-precision classification fusion method

To guarantee the correctness of the classification and fusion results, it is necessary to pre-process the data, i.e., to add the location information of the data points, to select the test data in the location intersection set of different tests as the valid data points, and to carry out the whole set of methods based on the valid data points. Data preprocessing included adding location information to different physical exploration test data and taking them in the location intersection set as valid data. With the help of the Gaussian mixture model, the physical exploration data set A (expressed as resistivity) was divided into k categories. The data points of different categories would obey different Gaussian distributions in principle. If no processing was done, there always existed the situation that the data points obeyed different Gaussian distributions at the same time, and these data points were called mixed data set. Propose Category-boundary algorithm was applied to classify and fuse the mixed data set with high

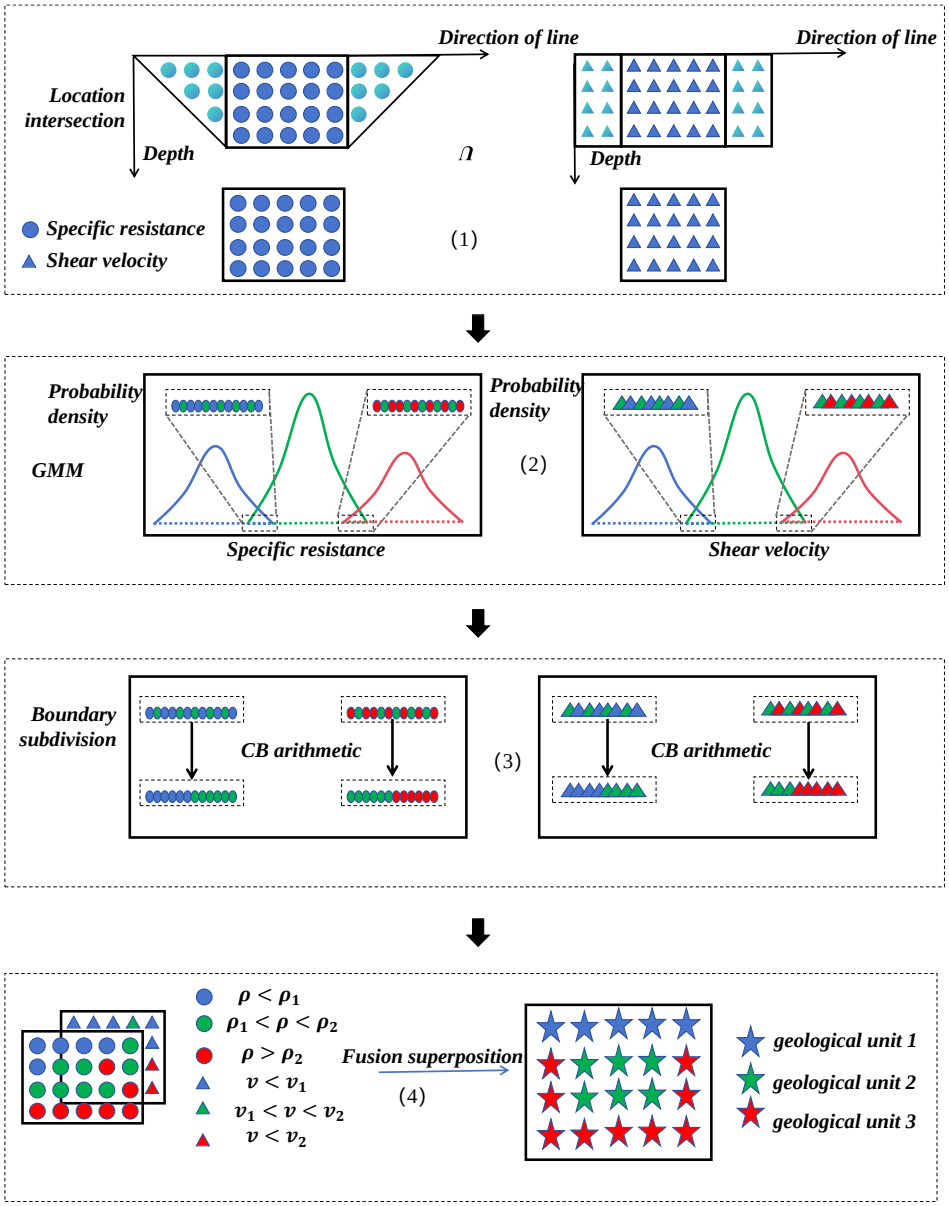

Figure 1. Classification fusion method based on improved Gaussian mixture model.

accuracy. The point set composed of these data points was named mixed data set. Category-boundary algorithm was then proposed to further divide the mixed data set to obtain the segmentation results of the improved Gaussian mixing model. A classification process was further introduced into the other data set B (represented by the shear wave velocity) and the other data set C (the data set collected by other methods of physical exploration) [8] (Figure 1). The expert experience was further introduced to superimpose and fuse the classification results of different physical exploration test data, and the river evolution prediction results could be obtained by integrating multiple physical exploration methods [9]. Comprehensive physical exploration techniques refer to the use of two or more karst exploration methods for analysis. As different methods have different accuracy and scope of application, the data sets that can be acquired for the same survey line also differ in depth range [10].

Gaussian mixture model

The mathematical prototype of a finite mixture model was described as that, let be a d-dimensional random variable, $p_i(x \mid \theta_i)$ (1, 2, 3, I m = …) be its corresponding probability density function. If a sample of Xi was mixed in a certain proportion and then any one of them was observed, and the result was recorded as X, the random variable X obeyed the mixed distribution and its probability density function.

$$\boldsymbol{p}(\boldsymbol{x}\mid\ \theta)=\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}}\pi_{\boldsymbol{i}}\boldsymbol{p}_{\boldsymbol{i}}\left(\boldsymbol{x}\mid\ \theta_{\boldsymbol{i}}\right)\boldsymbol{x}\in\boldsymbol{R}^{\boldsymbol{d}} \quad (1)$$

where π_i was the mixing ratio that satisfies the following.

$$\pi_{\boldsymbol{i}}\geq 0,\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}}\pi_{\boldsymbol{i}}=1 \quad (2)$$

$$\theta=\left(\pi_1,\pi_2,\ldots,\pi_{\boldsymbol{m}},\theta_1^{\cdot}\ ,\theta_2^{\cdot}\ ,\ldots,\theta_{\boldsymbol{m}}^{\cdot}\right)^{\cdot}\ ,\theta\in\Theta \quad (3)$$

$$\left.\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}}\pi=1,\pi_{\boldsymbol{i}}\geq 0,\theta_{\boldsymbol{i}}\in\Theta_{\boldsymbol{i}},\boldsymbol{i}=1,2,\ldots,\boldsymbol{m}\right\} \quad (4)$$

where θ was the parameter space. If the unique variable of the above formula was θ, it was called a family of mixed density parameters of the m- -branch. As a kind of finite mixture model, the parameter learning and modelling of Gaussian mixture model (GMM) are important in statistical learning and pattern recognition and have been widely used in many fields. In GMM, the data to be clustered were viewed as mixed probability distributions from multiple normal distributions, which represented different classes, so that the relevant parameters (mean and covariance matrices) of each normal distribution could be used as the class prototypes [11, 12]. If the d-dimensional random vector X had a finite mixture of normal distributions, its probability density function was as below.

$$\mathrm{p}(\mathrm{x}\mid\ \theta)=\sum_{\mathrm{i}=1}^{\mathrm{k}}\pi_{\mathrm{i}}\mathrm{p}_{\mathrm{i}}\left(\mathrm{x}\mid\ \theta_{\mathrm{i}}\right)=\sum_{\mathrm{i}=1}^{\mathrm{k}}\pi_{\mathrm{i}}(2\pi)^{-\frac{\boldsymbol{d}}{2}}\left|\Sigma_{\mathrm{i}}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\mathrm{x}-\mu_{\mathrm{i}}\right)^{\mathrm{T}}\Sigma_{\mathrm{i}}^{-1}\left(\mathrm{x}-\mu_{\mathrm{i}}\right)\right) \quad (5)$$

$$\mathrm{p}_{\mathrm{i}}\left(\mathrm{x}\mid\ \theta_{\mathrm{i}}\right)=(2\pi)^{-\frac{1}{2}}\left|\Sigma_{\mathrm{i}}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\mathrm{x}-\mu_{\mathrm{i}}\right)^{\cdot}\ \Sigma_{\mathrm{i}}^{-1}\left(\mathrm{x}-\mu_{\mathrm{i}}\right)\right)\mathrm{i}=1,2,\boldsymbol{k}_{\boldsymbol{k}} \quad (6)$$

where π_i was the mixing ratio.

$$\sum_{\mathrm{i}=1}^{\mathrm{L}}\pi_{\mathrm{i}}=1,\theta_{\mathrm{i}}=\left(\mu_{\mathrm{i}}^{\mathrm{T}},\overset{\mathrm{r}}{\Sigma}_{\mathrm{i}}\right)^{\mathrm{T}} \quad (7)$$

$$\theta=\left(\theta_1^{\mathrm{T}},\theta_2^{\mathrm{T}},\mathrm{L}\ ,\theta_{\mathrm{k}}^{\mathrm{T}},\lambda_1,\lambda_2,\mathrm{L}\ ,\lambda_2\right)^{\mathrm{T}} \quad (8)$$

There was a residual problem with data in practice, which was because much of the data in the real world was incomplete. Expectation-Maximization (E.M.) algorithm is a parameter estimation method for probability models containing hidden variables, which maximizes the likelihood function through iterative optimization and is especially suitable for cases where there are missing values or hidden variables in the data. The initialization process included random initialized model parameters or used some heuristics to set the initial value. The expected step (Step E) included the calculation of expected value of the hidden variable under the current parameter. Specifically, the posterior distribution of the hidden variables was calculated under the current model parameters, given the observed data, which was the expected value of each hidden variable at each sample point. The maximize step (Step M) was re-estimated the model parameters based on the hidden variable expectation obtained from the E step to maximize the likelihood function, then the parameters were updated by maximizing the expected log-likelihood function. The judgment convergence determined whether the parameter updated converges, usually by seeing if the log-likelihood function changed little or if the parameter changed within a certain threshold. If it was converged, the algorithm stopped, otherwise, it returned to step E and continued iteration. If θ(t) was the estimate of the a posteriori plurality after the start of the t + 1th iteration, the two steps of the t + 1th iteration were as follows. Step E included constructing Z away by taking the expectation of the conditional distribution of $\mathrm{p}(\theta \mid \mathrm{Y},\mathrm{Z})$ or $\log\ \mathrm{p}(\theta \mid \mathrm{Y},\mathrm{Z})$ with

respect to Z, i.e. $\boldsymbol{Q}(\theta|\ \theta^{(t)},\boldsymbol{Y}) = \boldsymbol{E}_{\boldsymbol{Z}}\left[\log \boldsymbol{p}(\theta|\ \boldsymbol{Y},\boldsymbol{Z})|\ \theta^{(t)},\boldsymbol{Y}\right] = \int \log[\boldsymbol{p}(\theta|\ \boldsymbol{Y},\boldsymbol{Z})]\boldsymbol{p}\left(\boldsymbol{Z}|\ \theta^{(i)},\boldsymbol{Y}\right)\boldsymbol{dZ}$, while step M involved Maximise $\boldsymbol{Q}(\theta|\ \theta(\boldsymbol{i}),\boldsymbol{Y})$, i.e. find a point θ(t+1) as

$$\boldsymbol{Q}\left(\theta^{(\mathrm{i}+1)}|\ \theta(\mathrm{i}),\mathrm{Y}\right) = \max_{\theta}(\theta|\ \theta^{(\mathrm{i})}) \tag{9}$$

After obtaining $\theta_{(t+1)}$, an iteration was created as

$$\theta_{(t)} \ \rightarrow \theta_{(t+1)} \tag{10}$$

The above E-steps and M-steps were Iterated up to the follows.

$$\left\|\theta^{(\boldsymbol{l}+1)} - \theta_{(\mathrm{i})}\right\| \tag{11}$$

$$\left\|\mathrm{Q}\left(\theta^{(\mathrm{l}+1)}\right)|\ \theta^{(\mathrm{i})},\mathrm{Y} - \mathrm{Q}\left(\theta^{(\mathrm{i})}|\ \theta\theta^{(\mathrm{l}},\mathrm{Y}\right)\right\| \tag{12}$$

The great likelihood method was used to estimate the parameters of the finite mixture model, and the E.M. algorithm became the easiest and fastest method for great likelihood estimation while optimizing the whole clustering process. By applying the E.M. algorithm to the finite mixture case by defining the unobservable data below

$$\boldsymbol{Z} = \left(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{\boldsymbol{n}}\right) \tag{13}$$

$$\boldsymbol{z}_{\boldsymbol{i}} \in \{1,\ldots,\boldsymbol{g}\} \tag{14}$$

$$\pi = \left(\pi_{\boldsymbol{i}};\boldsymbol{i}=1,\ldots,\boldsymbol{g}\right), \theta = \left(\theta_{\boldsymbol{i}};\boldsymbol{i}=1,\ldots,\boldsymbol{g}\right) \tag{15}$$

the process of estimating the parameter $\mathrm{s}(\psi,\psi^{(\mathrm{m})})\mathrm{Y}$ was as follows with the log-likelihood function of Y as:

$$\log(\mathrm{p}(\mathrm{Y}|\ \psi)) = \log\left(\prod_{\mathrm{i}=1}^{\infty}\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}\right),\mathrm{z}_{\mathrm{i}}|\ \psi\right) = \sum_{\mathrm{m}=1}^{\infty}\log\left(\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}|\ \mathrm{z}_{\mathrm{n}},\psi\right)\mathrm{p}\left(\mathrm{z}_{\mathrm{i}}|\ \psi\right)\right) \tag{16}$$

the conditional distribution density function of the unobservable data Z as:

$$\mathrm{p}\left(\mathrm{z}|\ \mathrm{X},\psi^{(\infty)}\right) = \prod_{\infty}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z},|\ \mathrm{x}_0,\psi^{(\infty)}\right) \tag{17}$$

where $\left(\mathrm{as}\ |\ \mathrm{X},\psi^{(\mathrm{m})}\right)$ was given by Bayes' formula below.

$$\mathrm{p}\left(\mathrm{z}_1|\ \mathrm{x}_{\mathrm{n}}\psi^{(*)}\right) = \frac{\mathrm{p}\left(\mathrm{x}_1|\ \mathrm{z}_0\psi^{(*)}\right)\mathrm{p}\left(\mathrm{z}_1|\ \psi^{(*)}\right)}{\mathrm{p}\left(\mathrm{z}_1|\ \psi^{(*)}\right)} = \frac{\pi\mathrm{z}_0\mathrm{np}\left(\mathrm{x}_1|\ \theta^{(*)}_{z_2}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\pi_{\mathrm{i}}^{(\mathrm{n})}\mathrm{p}\left(\mathrm{x}_1|\ \theta_{\mathrm{i}}^{(\mathrm{n})}\right)} \tag{18}$$

Then the following equation was obtained.

$$\begin{aligned}
&\mathrm{Q}\left(\gamma,\psi^{\mathrm{n}}\right) = \sum_{\mathrm{z}=1}\log(\mathrm{p}(\mathrm{Y}|\ \mathrm{v}))\mathrm{p}\left(\mathrm{z}|\ \mathrm{X},\psi^{(\mathrm{m})}\right)\\
&= \sum_{\mathrm{r}=\mathrm{i}}^{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\log\left(\pi_3\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}|\ \theta_{\mathrm{i}}\right)\right)\prod_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_1|\ \mathrm{x}_{\mathrm{i}},\psi^{(\mathrm{n})}\right)\\
&= \sum_{\mathrm{x}_1=1}^{\mathrm{N}}\mathrm{L}\sum_{\mathrm{s}=1}^{\infty}\sum_{\mathrm{i}=1}^{\mathrm{n}}\log\left(\pi_4\mathrm{p}\left(\mathrm{x}_1|\ \theta_{\mathrm{x}_1}\right)\right)\prod_{\mathrm{l}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_1|\ \mathrm{x}_{\mathrm{i}},\psi^{(\mathrm{n})}\right)\\
&= \sum_{\mathrm{s},1}^{\mathrm{L}}\mathrm{L}\sum_{\mathrm{s}=1}^{\infty}\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\infty}\delta_{\mathrm{j}=4}\log\left(\pi_j\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}|\ \theta_{\mathrm{j}}\right)\right)\prod_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_{\mathrm{i}}|\ \mathrm{x}_1,\psi^{(\mathrm{m})}\right)\\
&= \sum_{\mathrm{i}=1}^{\infty}\sum_{\mathrm{j}=1}^{\mathrm{n}}\log\left(\pi_j\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}|\ \theta_{\mathrm{f}}\right)\right)\sum_{\mathrm{n}=1}^{\mathrm{n}}\mathrm{L}\sum_{\mathrm{n}=1}^{\mathrm{n}}\delta_{\mathrm{k}=1}^{\mathrm{n}}\prod_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_{\mathrm{i}}|\ \mathrm{x}_{\mathrm{y}},\psi^{(\mathrm{n})}\right)
\end{aligned} \tag{19}$$

For fixed $\mathrm{i} \in \{1,\ldots,\mathrm{n}\}$ and $\mathrm{j} \in \{1,\ldots,\mathrm{g}\}$, there were:

$$\sum_{\mathrm{n}=1}^{\mathrm{n}}\sum_{\mathrm{n}=1}^{\mathrm{n}}\delta_{\mathrm{k}=1}^{\mathrm{n}}\prod_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_{\mathrm{i}}|\ \mathrm{x}_{\mathrm{y}},\psi^{(\mathrm{n})}\right) = \prod_{\mathrm{min}^2}\left(\sum_{\mathrm{n}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{z}_1|\ \mathrm{x}_{\mathrm{i}},\psi^{(\mathrm{n})}\right)\mathrm{p}\left(\mathrm{j}|\ \mathrm{x}_{\mathrm{e}}\psi^{(\mathrm{n})}\right) = \mathrm{p}\left(\mathrm{j}|\ \mathrm{x}_{\mathrm{c}},\psi^{(\mathrm{n})}\right)\right) \tag{20}$$

$$\sum_{\mathrm{n}=1}^{\mathrm{n}}\mathrm{p}\left(\mathrm{i}|\ \mathrm{x}_{\mathrm{n}},\psi^{(\mathrm{e})}\right) = 1 \tag{21}$$

and then, from the above equations, the following equation was obtained.

$$\begin{aligned}
&\boldsymbol{Q}\left(\Psi,\Psi^{(\omega)}\right) = \sum_{\boldsymbol{m}-1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}-1}^{\boldsymbol{n}}\log\left(\pi,\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{i}}|\ \theta_j\right)\right)\boldsymbol{p}\left(\boldsymbol{j}|\ \boldsymbol{x}_{\boldsymbol{J}},\Psi^{(\omega)}\right)\\
&= \sum_{\boldsymbol{i}-1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}-1}^{\boldsymbol{n}}\log(\pi,)\boldsymbol{p}\left(\boldsymbol{j}|\ \boldsymbol{x}_{\boldsymbol{i}},\Psi^{(\boldsymbol{m})}\right) + \sum_{\boldsymbol{m}-1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}-1}^{\boldsymbol{n}}\log\left(\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{i}}|\ \theta_j\right)\right)\boldsymbol{p}\left(\boldsymbol{j}|\ \boldsymbol{x}_{\boldsymbol{i}},\Psi^{(\boldsymbol{m})}\right)\\
&= \sum_{\boldsymbol{i}-1}^{\boldsymbol{n}}\omega_{\boldsymbol{i}}^{\boldsymbol{r}}\boldsymbol{I} + \sum_{\boldsymbol{m}-1}^{\boldsymbol{n}}\omega_{\boldsymbol{i}}^{\boldsymbol{r}}\boldsymbol{u}_{\boldsymbol{J}}(\boldsymbol{\theta})
\end{aligned} \tag{22}$$

where the Jth component of the vector was:

$$\omega_{\boldsymbol{q}} = \boldsymbol{p}\left(\boldsymbol{j}|\ \boldsymbol{x}_{\boldsymbol{i}},^{(\boldsymbol{m})}\right) = \frac{\pi_{\boldsymbol{j}}^{(\boldsymbol{m})}\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{i}}|\ \theta_{\boldsymbol{j}}^{(\boldsymbol{m})}\right)}{\sum_{\boldsymbol{i}=1}^{2}\pi_{\boldsymbol{i}}^{(\boldsymbol{m})}\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{i}}|\ \theta_{\boldsymbol{i}}^{(\boldsymbol{m})}\right)} \tag{23}$$

In step M, since π and θ were independent of each other, maximizing $\mathrm{Q}(\psi,\psi^{(\mathrm{m})})$ could be

done by maximizing the two terms in the above equation separately. Firstly, in the restriction $\sum_{j=1}^{n}\pi=1$, the maximum value of $\sum_{=1}^{\infty}\omega_i^{T}I$ was found with respect to π_j, i.e., differentiating with respect to π for $\sum_{i=1}^{\infty}\omega_i^{T}\boldsymbol{I}-\lambda\left(\sum_{\boldsymbol{m}=1}^{2}\pi,-1\right)$, while making it = 0, where λ was the Lagrange multiplier as below.

$$\frac{\partial}{\partial\pi_j}\left[\sum_{\boldsymbol{m}=1}^{\boldsymbol{n}}\omega_1^{r}\boldsymbol{l}-\lambda\left(\sum_{\boldsymbol{j}=1}^{\boldsymbol{s}}\pi_j-1\right)\right]=0 \tag{24}$$

This led to the equation (25) below.

$$\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}\omega_{\boldsymbol{g}}\frac{1}{\pi_j}-\lambda=0 \tag{25}$$

which was constrained by

$$\sum_{\boldsymbol{m}=1}^{2}\pi_j=1 \tag{26}$$

$$\lambda=\sum_{\boldsymbol{p}=1}^{\boldsymbol{n}}\sum_{\boldsymbol{n}=1}^{\infty}\boldsymbol{a}_{\boldsymbol{g}}=\boldsymbol{n} \tag{27}$$

$$\pi_j^{(\boldsymbol{e}+1)}=\frac{1}{\boldsymbol{n}}\sum_{\boldsymbol{n}=1}^{\infty}\omega_{\boldsymbol{g}} \tag{28}$$

$$\mathrm{p}\left(\mathrm{x}|\ \theta_{\mathrm{f}}\right)=\frac{1}{(2\pi)^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}}\exp\quad\left\{-\frac{1}{2}\left(\mathrm{x}-\mu_{\mathrm{s}}\right)^{\mathrm{x}}\Sigma_{\mathrm{j}}^{-1}\left(\mathrm{x}-\mu_{\mathrm{y}}\right)\right\} \tag{29}$$

$$\sum_{i=1}^{\infty}\theta_i\boldsymbol{u}_i(\theta)=\sum_{i=1}^{\infty}\theta_{ij=1}^{\infty}\left[-\frac{1}{2}\log\quad(2\pi)-\frac{1}{2}\log\quad(|\Sigma,|)-\frac{1}{2}\left(\boldsymbol{x}_i-\mu_{j\boldsymbol{e}}\right)\right] \tag{30}$$

This yielded an estimate of μ_j as follows.

$$\mu_j^{(\boldsymbol{m}+1)}=\frac{\sum_{\boldsymbol{i}=1}^{\infty}\omega_i\boldsymbol{X}_i}{\sum_{\boldsymbol{i}=1}^{\infty}\omega_i}=\frac{1}{\boldsymbol{n}\varepsilon_j^{(\boldsymbol{n}+1)}}\sum_{\boldsymbol{i}=1}^{\infty}\omega_i\boldsymbol{X}_i \tag{31}$$

$\sum_{\mathrm{i}=1}^{\infty}\omega_{\mathrm{i}}^{\cdot}\ \mu_{\mathrm{i}}(\theta)$ could also be written as:

$$\begin{aligned}\sum_{i=1}^{\infty}\left[\boldsymbol{u}_{\mathrm{s}}(\theta)^{\cdot}\right.&=\sum_{\mathrm{i}=1}^{\infty}\sum_{\mathrm{j}=1}^{\mathrm{s}}\omega_{\mathrm{s}}\left[-\frac{\boldsymbol{p}}{2}\log(2\pi)+\frac{1}{2}\log\left(\left|\Sigma_{\mathrm{j}}^{-1}\right|\right)-\frac{1}{2}\mathrm{lr}\left(\Sigma_{\mathrm{j}}^{-1}\left(\mathrm{x}_{\mathrm{i}}-\mu_{\mathrm{j}}\right)\left(\mathrm{x}_{\mathrm{j}}-\mu_{\mathrm{j}}\right)^{\mathrm{T}}\right)\right]\\&=-\frac{\boldsymbol{np}}{2}\log(2\pi)+\sum_{\mathrm{j}=1}^{\infty}\left[\frac{1}{2}\log\left(\left|\Sigma_{\mathrm{s}}^{-1}\right|\right)\sum_{\mathrm{i}=1}^{\infty}\omega_{\boldsymbol{g}}-\frac{1}{2}\sum_{\mathrm{i}=1}^{\infty}\omega_{\mathrm{j}}\,\mathrm{tr}\left(\Sigma_{\mathrm{j}}^{-1}\mathrm{N}_{\mathrm{G},s}\right)\right]\end{aligned} \tag{32}$$

By finding the Σ_j^{-1} partial derivative of $\sum_{\mathrm{i}=1}^{\varepsilon}\omega_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{u}_{\boldsymbol{i}}(\theta)$ and making the partial derivatives equal to 0, the following equation was yielded,

$$\begin{aligned}&\frac{1}{2}\sum_{\mathrm{i}=1}^{\infty}\omega_{\mathrm{G}}\left(2\Sigma_{\mathrm{j}}-\mathrm{diag}\left(\Sigma_{\mathrm{j}}\right)\right)-\frac{1}{2}\sum_{\mathrm{i}=1}^{\infty}\omega_{\mathrm{G}}\left(2\,\mathrm{N}_{\mathrm{G},\mathrm{j}}-\mathrm{diag}\left(\mathrm{N}_{\mathrm{i},\mathrm{j}}\right)\right)\\&=\frac{1}{2}\sum_{\mathrm{i}=1}^{\infty}\omega_{\mathrm{G}}\left(2\mathrm{M}_{\mathrm{G},\mathrm{j}}-\mathrm{diag}\left(\mathrm{M}_{\mathrm{G}}\right)\right)\\&=2\mathrm{S}-\mathrm{diag}(\mathrm{S})=0\end{aligned} \tag{33}$$

which gave a re-estimation of Σ_j as below.

$$\begin{aligned}\sum_{\boldsymbol{j}}^{(\boldsymbol{e}+1)}&=\frac{\sum_{\boldsymbol{i}=1}^{\infty}\Theta_{\boldsymbol{i}}\left(\boldsymbol{x}_{\boldsymbol{i}}-\mu_{\boldsymbol{j}}^{(\boldsymbol{e}+1)}\right)\left(\boldsymbol{x}_{\boldsymbol{i}}-\mu_{\boldsymbol{j}}^{(\boldsymbol{e}+1)}\right)^{\boldsymbol{T}}}{\sum_{\boldsymbol{i}=1}^{\infty}\theta 4}\\&=\frac{1}{\boldsymbol{n}\pi_{\boldsymbol{j}}^{(0+1)}}\sum_{\boldsymbol{i}=1}^{2}\grave{\mathrm{o}}_{\boldsymbol{g}}\left(\boldsymbol{x}_{\boldsymbol{i}}-\mu_{\boldsymbol{j}}^{(\boldsymbol{m}+1)}\right)\left(\boldsymbol{x}_{\boldsymbol{i}}-\mu_{\boldsymbol{j}}^{(\boldsymbol{m}-\boldsymbol{a})}\right)\end{aligned} \tag{34}$$

(2) Time series clustering algorithm

The clustering applied time series clustering algorithm based on local linear embedding (LLE) and Gaussian mixture model (GMM), which consisted of two steps including that, from the point of view of preserving the local structure of the dataset, each high-dimensional time series sample was represented as a linear combination of its k-nearest neighbors using LLE and reconstructed in the low-dimensional space to achieve dimensionality approximation while preserving the local geometrical structure of the dataset [13]. Then GMM was used to perform the clustering analysis from the point of view of probability distribution. The main steps of the algorithm were as follows [14].

The algorithm:

$$\boldsymbol{LLEGMM}(\mathrm{X},\mathrm{C},\mathrm{k},\mathrm{d}) \tag{35}$$

Embedding dimension d:

$$\mathrm{X}=\left\{\mathrm{x}_1,\mathrm{x}_2,\mathrm{L}\ ,\mathrm{x}_\mathrm{N},\mathrm{x}_\mathrm{i}\in\mathrm{R}^\mathrm{m}\right\} \tag{36}$$

The output of clustering results included 9 steps as step1 used the Principal Component Analysis (PCA) algorithm on dataset X to remove noise and redundancy; step2 constructed the set of nearest neighbors for any xi with k nearest neighbors xj; step3 calculated the local reconstruction weight vector of xi and formed the weight coefficient matrix (W) as below;

$$\mathrm{W}_\mathrm{i}=\frac{\mathrm{Z}_\mathrm{i}^{-1}1_\mathrm{k}}{1_\mathrm{k}^\mathrm{K}\mathrm{Z}_\mathrm{i}^{-1}1_\mathrm{k}} \tag{37}$$

step4 constructed matrix M and computed the first d + 1 eigenvalues and the corresponding eigenvectors of M with the low-dimensional embedding as Y;

$$\mathrm{Y}=\left\{\mathrm{y}_2,\mathrm{L}\ ,\mathrm{y}_{\mathrm{d}+1}\right\} \tag{38}$$

step 5 initialized Gaussian mixture model parameters to start iteration; step 6 (E-step) found the probability of each sample for each category; step 7 (M-step) optimized the model parameters of E-step to get new parameters; step 8 repeated E-step and M-step until the parameters converged or the maximum number of iterations was reached; step 9 clustered with the trained GMM model.

Simulated data clustering and variable selection

(1) Two-dimensional modelled data

Cluster analysis of simulation data was assumed according to Gaussian distribution and Gaussian distribution without any noise. Three categories of cluster analysis were identified including Cluster 1, Cluster 2, and Cluster 3 with a total of 120 observation samples and 50, 20, and 50 samples in each cluster, respectively. The Gaussian distribution was N (0, 1), N (2.5, 1), and N (5, 1). The mixed samples were clustered using the Gaussian mixture model, while the BIC criterion was used to select the number of clusters and penalty parameters, and the detection index of the clustering effect was ARI. The scatter plot of the mixed samples, the clustered samples of K = 3, and the BIC index plot of the Gaussian model with different covariance matrices in the BIC metrics plot for the number of clusters K=3 were shown in Figure 2.

(2) High-dimensional modelled data

High-dimensional data clustering was based on traditional methods. Due to the failure of the algorithm based on distance, high-dimensional data dimensionality was reduced. However, if the high-dimensional data dimensionality was reduced first before clustering, it might seriously damage the structure of the clusters. This study adopted the form of the penalty model to compress the parameters, which could be achieved while reducing the dimensionality of the side of the effect of the cluster and would not destroy the structure of the clusters themselves. Based on the Gaussian mixture model clustering, the model added a penalty term and adjusted the penalty parameters on the model parameters compression. Figure 3 illustrated the application of L penalty Gaussian hybrid model for high-dimensional data clustering analysis and variable selection. The penalty parameter A changed Gaussian hybrid model on the variable mean value of the compression of the trend. The clustering number of indicators BIC criterion was selected for the penalty parameter A as 2.5 and the number of clusters K as 3. The BIC was then to achieve the smallest value. As a result, the model could cluster accurately at small times, but it could not select the information variables effectively (Figure 3(1)). The compression of the mean parameter using the BIC criterion for K = 2 and A = 20 showed that all variables were compressed to 0 at the appropriate parameter values, which was not meaningful for the clustering of the model (Figure 3(2)).

An integrated physical detection method for river evolution based on Gaussian hybrid clustering

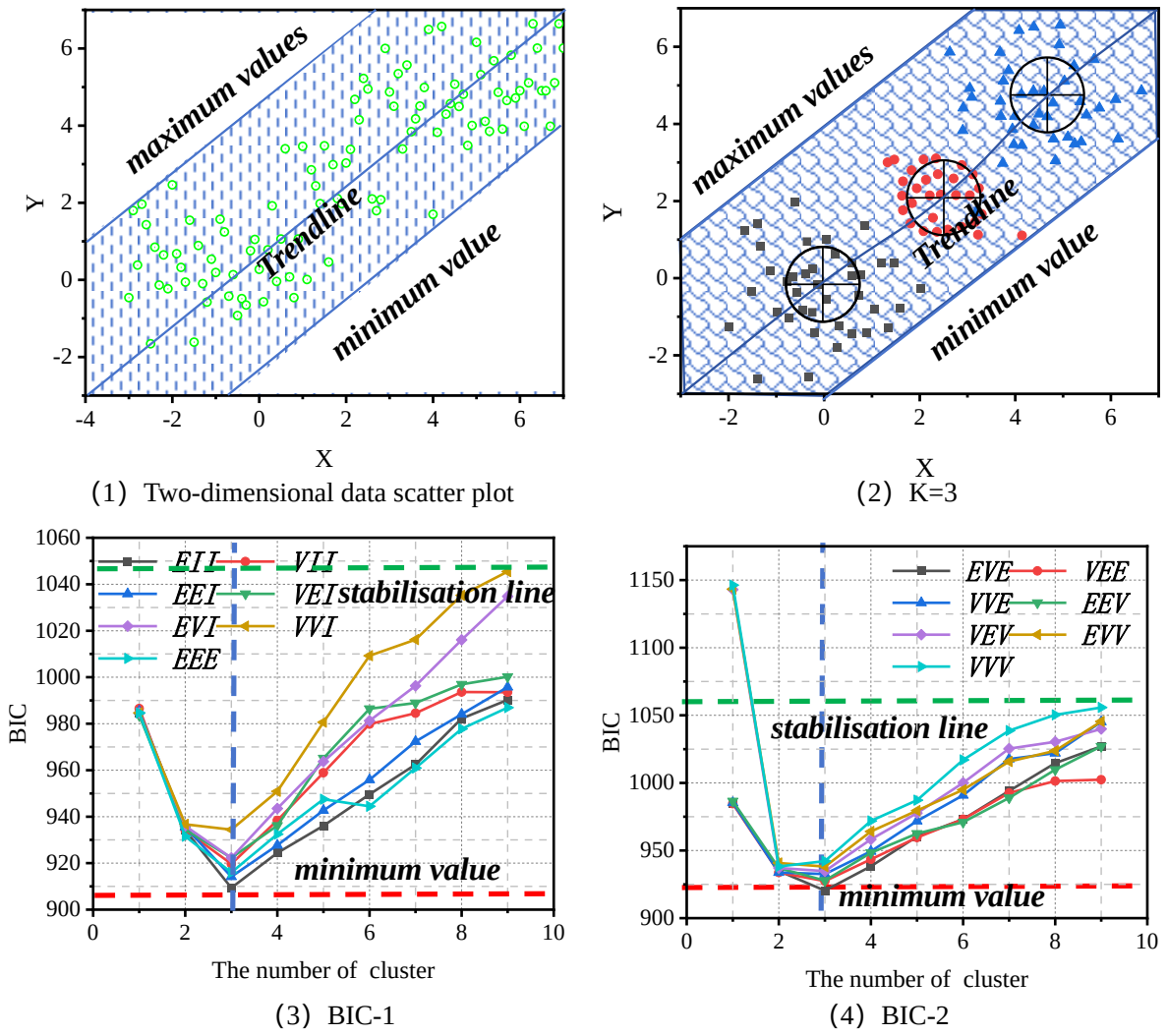

Figure 2. 2D simulation data clustering.

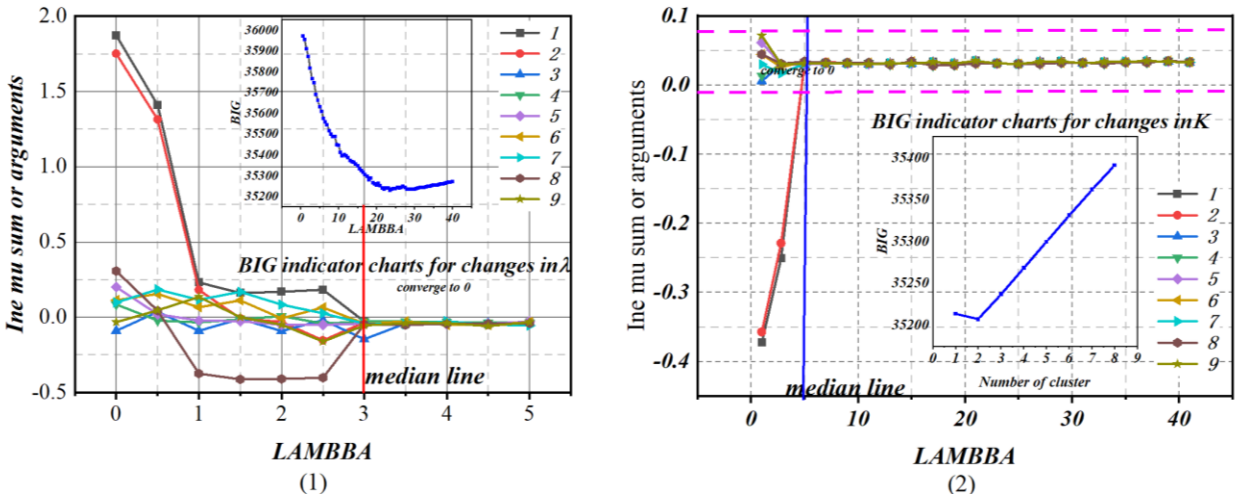

Figure 3. L-GMM clustering.

(1) Joint profiling

The joint rate profiling method is, in essence, a combination of two sets of three-pole devices of AMN and MNB. The AMNB electrodes were arranged along the measurement line, while the middle point of M.N. was the observation point. C was the poor-poor remote pole (vertically connected to A and B and O.C. = 3 - 5 times O.A.), and the power supply of A and C and the power supply of B and C were selected at the same observation point, respectively. Accordingly, the M.N. electrodes were measured to obtain the two sets of potential differences, i.e., the potential differences between the AMN and MNB devices. The two values of ρ_s^A and ρ_s^B were obtained after calculation, and the work of one observation point was completed. In the same way, after the point-by-point observation along the measurement line, the apparent resistivity values of ρ_s^A and ρ_s^B were obtained for all the measurement points. The two apparent resistivity curves of ρ_s^A and ρ_s^B on the same measurement line were plotted with the measurement line number as the horizontal axis and the ρ_s value as the vertical axis. The ρ_s value of each measurement point was as follows.

$$\rho_s^A = \frac{\boldsymbol{J}_{\boldsymbol{A},\boldsymbol{MN}}}{\boldsymbol{J}_{\boldsymbol{A},0}} \rho_{MN} \quad (39)$$

$$\rho_s^B = \frac{\boldsymbol{J}_{\boldsymbol{B},\boldsymbol{MN}}}{\boldsymbol{J}_{\boldsymbol{B},0}} \rho_{MN} \quad (40)$$

The basic working method was A, M, N, B three different types of line along the same line with O as the central symmetry. In work of A, M, N, B three different spacing, A was the large displacement in the same line. Each measuring point had two types of loop measurement to obtain the visual resistivity value ρ_s^A and ρ_s^B in two visual resistivity curves. The changing pattern of the resistivity curve caused by the low-resistance target body was studied. When the AMN and MNB triple devices were far away from the ρ_2 target body, the observed resistivity was the background resistivity, i.e., $\rho_s^A = \rho_s^B = \rho_1$. When AMN moved to the right, the current field of electrode A was "attracted" by the low resistance body ρ_2, and the power supply density JAMN at the M.N. near the low resistance body increased, so ρ_s^A increased, and then the maximum was obtained at a certain position. As AMN continued to move to the right, the "attraction" effect of the low resistance on the current field at pole A became inclined downward. The power density at M.N. JAMN began to decrease, so ρ_s^A decreased and obtained a minimum value at A certain position. Then, as the device continued to move to the right, the "attraction" effect of the low resistance on the current field of the A pole began to weaken, so JAMN began to increase, so ρ_s^A also increased until the device left the low resistance far enough (in fact, more than two times the distance of A.O.). ρ_s^A tended to the surrounding rock resistivity ρ_1. Similarly, the ρ_s^B curve could be qualitatively explained by the subsurface current distribution law without being repeated. Just above the low-resistance body, the current field of AMN and MNB devices was equivalent to the "inhalation" of the low-resistance body. Then, $\rho_s^A = \rho_s^B$ was the intersection position. When the target body was a high resistor, it would "repel" the current field and opposite to the current field and the low resistance target body, forming a reverse intersection point. The different underground resistivity anomalies corresponded to different curve forms and forward and antonogonal points in the joint curve. The essence of the joint profile method was to infer the plane position, conductive property, size, and buried depth of the target body by using the morphology, intersection nature, and intersection position of the measured joint curve to achieve the purpose of detection. The joint profile method was characterized by simple and intuitive, obvious abnormality, and strong resolution ability, which was often used to detect good conductive ore bodies, water-bearing fracture, fracture zones, karst channels, and so on.

(2) Basic theory of the high-density electrical method

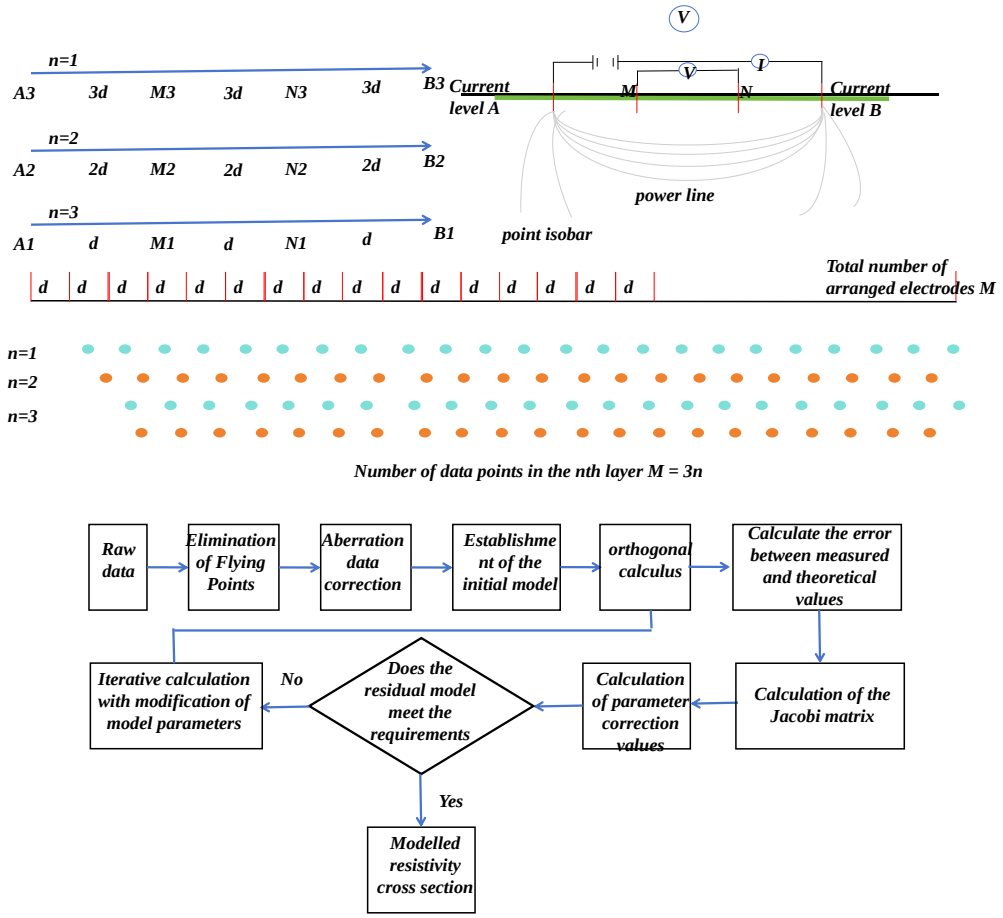

Figure 4. Schematic diagram of the working principle of the high-density electro fabricated Wenner device.

The high-density electrical method was an array-type direct-current resistivity exploration method that integrated the electrical bathymetry method and the electrical profiling method and was characterized by small point spacing, high data acquisition density, and high resolution. Its basic principle was the same as that of the conventional geoelectric method, which was based on the difference in electrical properties of the underground medium and calculated the resistivity parameter of the underground medium by observing the change of the artificially excited electric field on the ground surface, so as to deduce the geological structure of the underground and achieve the purpose of exploration. When observing in the field, the apparent resistivity of each measuring point was calculated by the formula below.

$$\rho_s = \boldsymbol{K} \cdot \frac{\Delta \boldsymbol{V}}{\boldsymbol{I}} \tag{41}$$

The DUK-4 cascade full-waveform high-density resistivity measurement system (Chongqing Geological Instrument Factory, Chongqing, China) was used for data acquisition. According to the working characteristics of each device of high-density electrical method, the Wenner device with strong interference ability and high productivity was selected for data acquisition in combination with the actual situation of the survey area. The two-dimensional apparent resistivity section data were obtained from the field measurement, while the two-dimensional inversion technique was used to obtain a more realistic geoelectric model (Figure 4). In this data processing, the original data were firstly edited

by the 2D DEM data analysis software to eliminate individual distorted data points and then inverted by the Gauss-Newton least-squares inversion method with smoothing constraints, which was based on the following equations (42) and (43).

$$\left(\boldsymbol{J}^{\boldsymbol{T}}+\mu \boldsymbol{F}\right)\boldsymbol{d}=\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{g} \quad (42)$$

$$\boldsymbol{F}=\boldsymbol{f}_{\boldsymbol{x}}\boldsymbol{f}_{\boldsymbol{x}}^{1}+\boldsymbol{f} \quad (43)$$

(3) Method application of high-density electrical method

In the process of finding groundwater, the main methods involved in the geophysical exploration methods were the resistivity profile method, resistivity sounding method, auxiliary methods, a natural potential method, and charging method. The anomalous area of low resistivity of the joint profile curve was speculated to be the position of the fissure fracture zone. When there was a fault fracture in the formation profile, the low resistance orthogonal point would appear in the joint profile curve. The tendency of the displacement of the different pole alignment curves was observed. Wenner arrangement, Schlenbel arrangement, and dipole arrangement belonged to the combined profile device. The obtained inversion map and contour map could intuitively show the underground electrical difference of the measurement line profile, which could distinguish the abnormal areas of high and low-resistance wells and speculated on geological conditions such as underground cracks and strata. The curve type of the electric depth depended on the nature of the ground electric section, and the change of the curve type could reflect the changing characteristics of the underground rock strata. The isotopic resistivity was often drawn to reflect the change of the visual resistivity in the vertical section where a measuring line passed.

(4) Method discussion of high-density electrical methods

The high-density electric method was a multi-functional electric method integrated with various electric methods. The field data could meet the geological exploration requirements of shallow within 150 m. The measurement section was generally inverted trapezoid, which could not fully reflect the overall geological characteristics of the area covered by the measuring line. The combined profile method could directly indicate the development of underground structure cracks and had a better effect on finding water content in the carbonated areas. However, its disadvantage was the long field measurement time.

Application of a geophysical approach for river evolution based on Gaussian mixed clustering

Geodesic work is the basis of comprehensive geophysical profile measurement. Lingrui S86 GPS RTK (South Surveying and Mapping Co., LTD., Guangzhou, Guangdong, China) was used for the measurement in this study. The CORS network joint test was adopted in the whole region with the central meridian longitude of 117 and the Gaussian 3-degree band coordinate system (Xi'an coordinate system in 1980, 1985 national elevation benchmark). The whole process of measurement strictly followed the Specification for Geophysical and Chemical Engineering Measurement (DA / T0153-2014) and Global Positioning System (GPS) Measurement Specification (GB / T18134-2009). The overall deployment of the geophysical detection line was based on the principle of the main NW and NE fault structures in the vertical area and the parallel profile verification or recourse structures. According to the relative position of the geophysical section in the study area, south (S) and north (N) shared the same section and point number. The southern area was Wangjia Village-Fangjia area located in the west of Madu Village with 7 S1 - S7 sections and Liuwu-Zhoucun area located in the west of abacus mouth with 5 N1 - N5 sections (Figure 5). On the basis of the original geological data, a comprehensive survey of the fault structure was carried out through the joint profile method. Further verification and comprehensive analysis of high-density electric method, electrographic sounding, transient electromagnetic, natural source wave

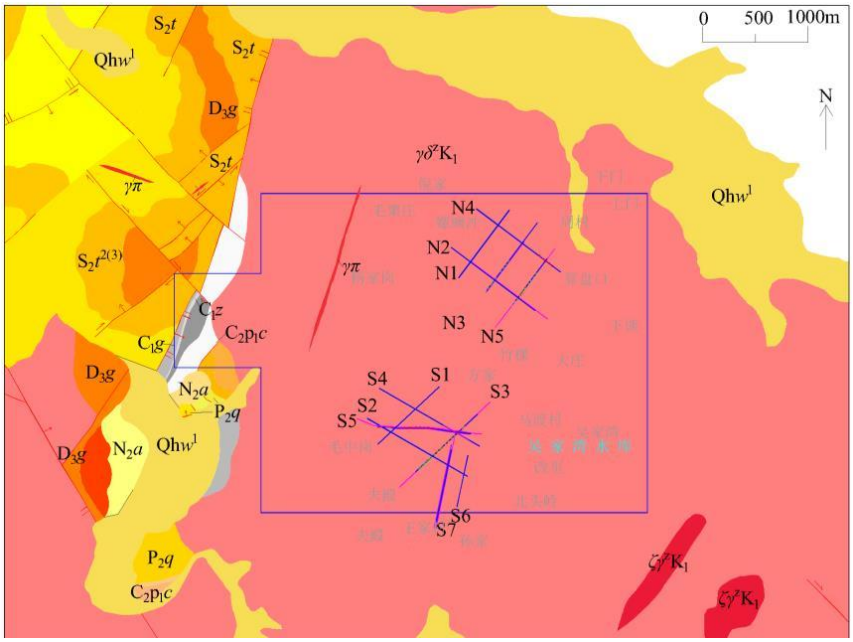

Figure 5. Actual material diagram of ground synthesis.

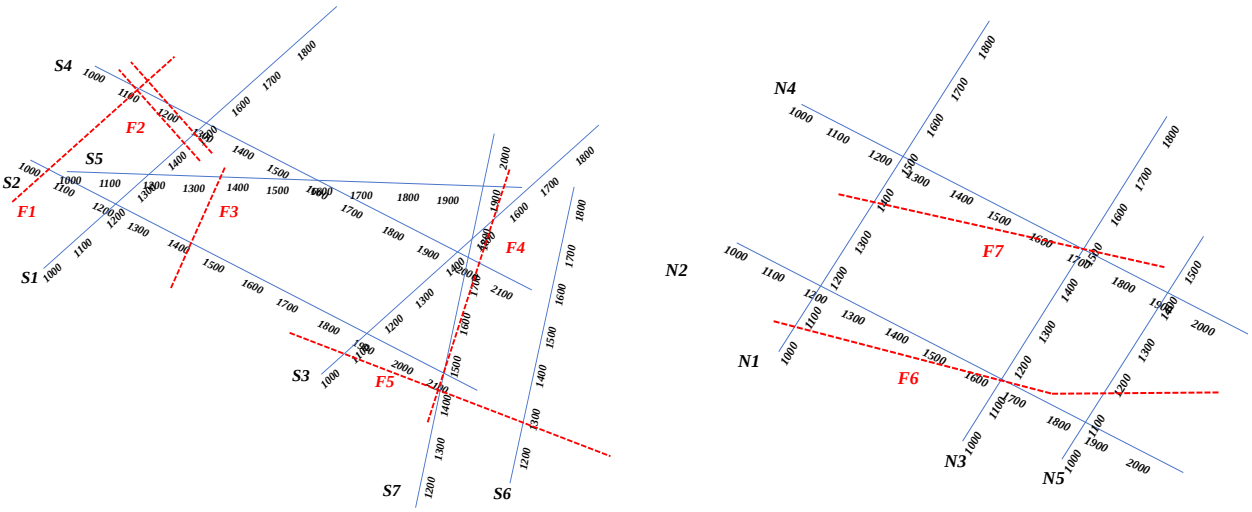

Figure 6. Comprehensive geophysical results.

exploration (micromotion) in the local abnormal positions with research significance were also performed. The original measurement and quality inspection observation of each comprehensive geophysical survey method was carried out in accordance with the relevant specifications. The optimal combination method of ground composite exploration in this area was joint profile method, micromotion, high density electrical method, and transient electromagnetic. Combined section and micromotion had good indicator effect on the fault structure, which could connect the normal point and the obvious low speed abnormal center as the key target. The micromotion reflected the fault structure, while the visual

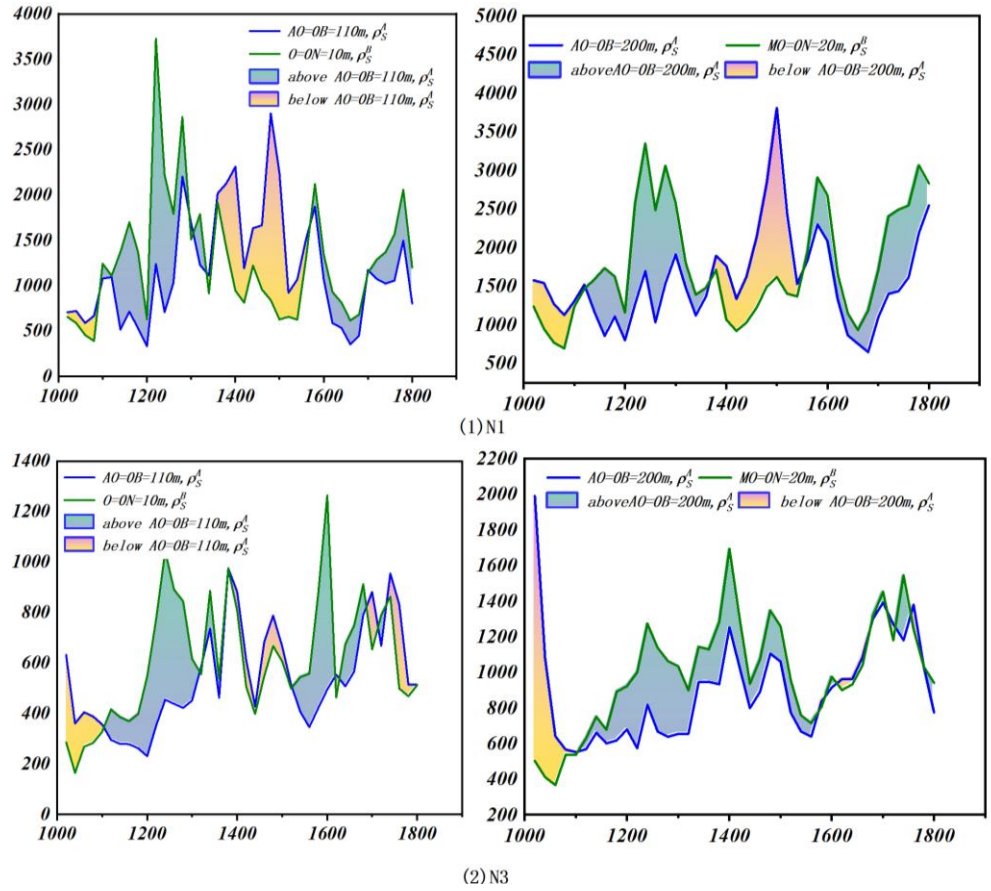

Figure 7. N1 and N3 joint profile results.

polarizer was not obvious to the water-content structure and could be deleted. High density electric method well reflected the deep and shallow faults, and its low resistance abnormal feature could indirectly indicate the relative depth of the fault structure. Transient electromagnetic was very sensitive to low resistance, which could be used as the verification and comparison of the fault structure. In addition, the practical application needed to screen the anomalies caused by the bedrock fissure. Based on the above geological data and the results of the ground synthesis, 4 NW and 3 NE fault structures were comprehensively inferred (Figure 6).

Results and discussion

Geo-evolutionary analysis

(1) The south part

The S2 material exploration profile had only the joint profile method with the orientation of the profile as 120°. The depth of the joint profile reflected by AO of 110 m, a polar distance, was relatively shallow, but in general, it was basically the same as that reflected by AO of 200 m, and the corresponding locations of the intersection anomalies were basically the same (Figure 7). Taking AO of 200 m as an example, there were two obvious anomalies of intersection points, which corresponded to points 1,030, 1,400, and 2,040, respectively. Corresponding to point 1,030, it was assumed that there was an F1 fracture structure here, and it was assumed that the F1 fracture structure was oriented towards NNE and spread along the S2 profile from point 1,030 to point 1,110 in the S4 profile. Corresponding to point 1,400, it was assumed that there was an F3 fracture structure here, and

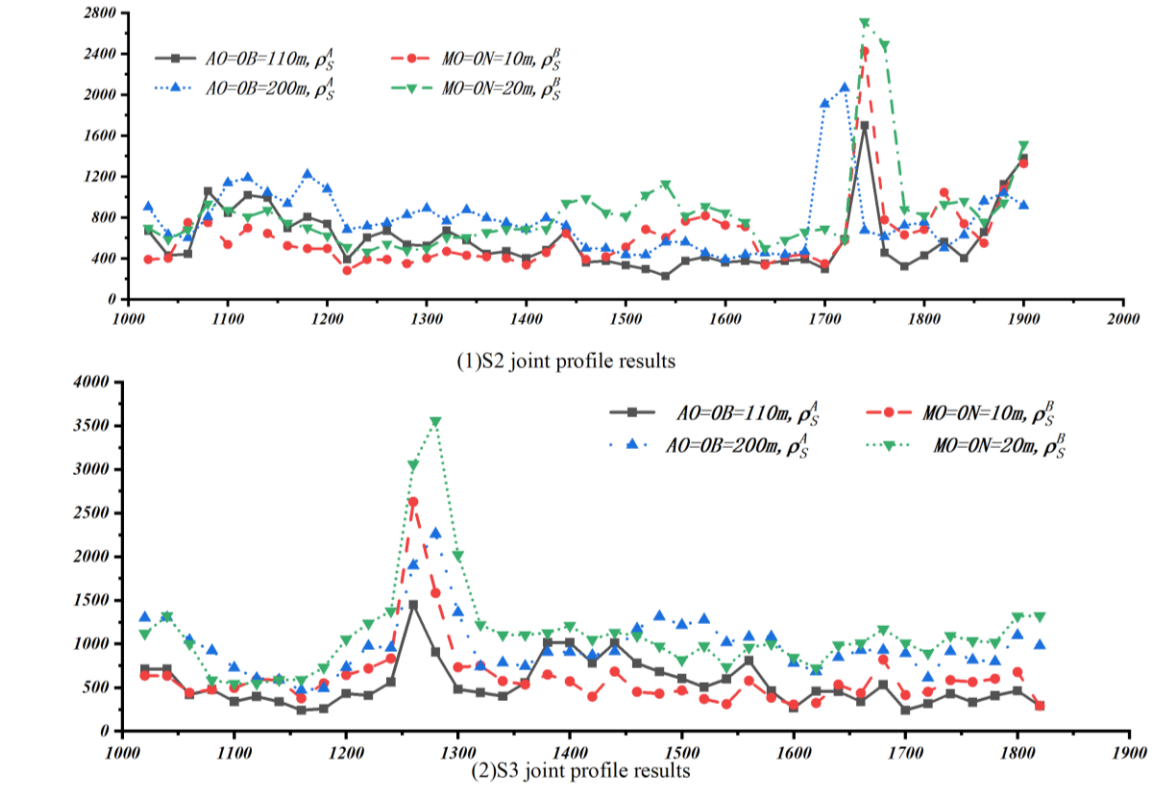

Figure 8. Map of joint profile results.

it was assumed that the F3 fracture structure was oriented towards NNE and spread along the S2 profile from point 1,420 to point 1,300 in the S5 profile. -F3 fracture structure was assumed to be along the NNE trend, spreading along the S2 profile at point 1,420 and the S5 profile at point 1,300. The S3 physical exploration profile was a comprehensive profile including the combined profile method, high density electric method, exciting electric bathymetry, and natural source surface wave survey (micromotion). The profile orientation was 47°. All the methods reflected different degrees of the inferred fracture structure, and the corresponding positions matched very well. From the S3 joint profile, the change characteristics of AO = 110 m and AO = 200 m were basically the same. The corresponding locations of the orthogonal anomalies were basically the same too. Taking AO = 200 m as an example, there were two obvious anomalies of intersection points, which corresponded to 1,120 points and 1,600 points, respectively. Corresponding to point 1,120, it was assumed that there was an F5 fracture here with an NWW trend and spread along the S6 profile at point 1,420, S7 profile at point 1,600, and S3 profile at point 1,120. Corresponding to point 1,600, it was assumed that there was an F4 fracture here with an NNE trend. The F4 fracture structure was assumed to be along with the NNE trend and spread along the S7 profile at point 1,600, the S4 profile at point 1,900, and the S3 profile at point 1,600 (Figure 8).

(2) The north part

In the north zone (N), there were five physical exploration profiles from N1 to N5, of which N2, N3, N4, and N5 were comprehensive profiles. Based on the complete interpretation of the results, the single physical exploration profile N1 and the comprehensive physical exploration profiles N3 and N5 were taken for elaboration. The results showed that N1 had only the combined profile method with the orientation of

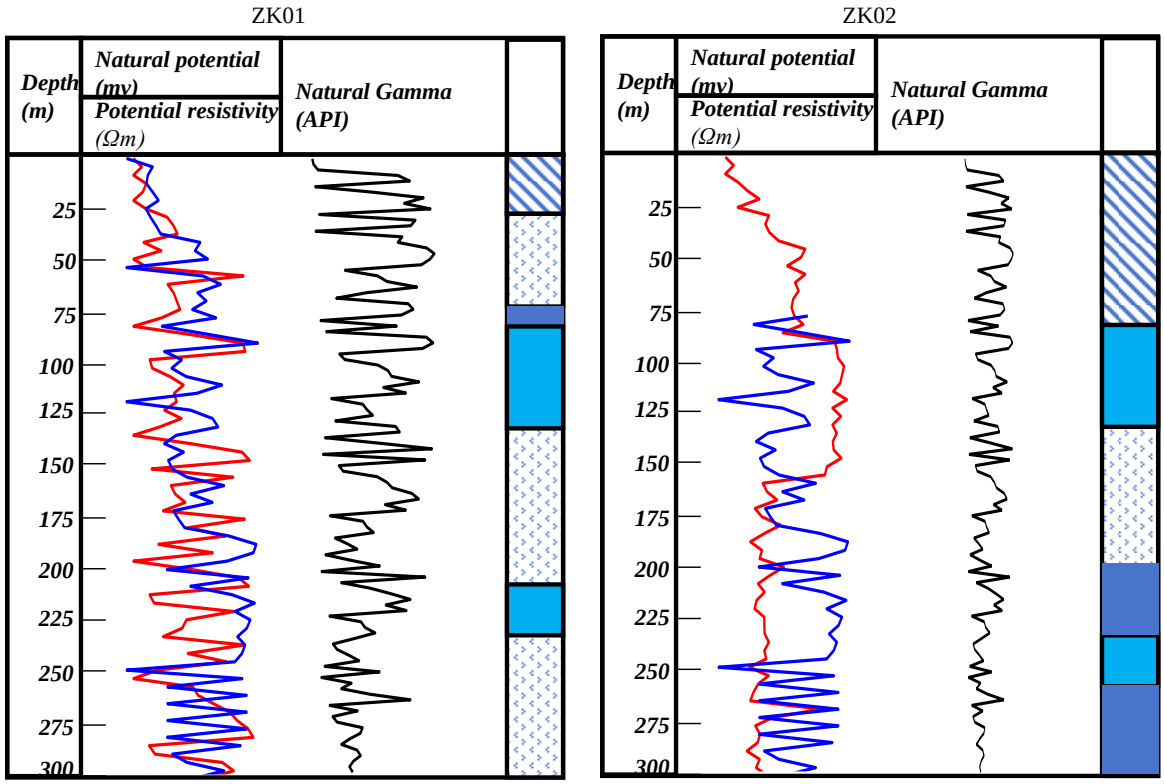

Figure 9. Hydrological results.

the profile as 37°. The measured depth of the combined profile reflected by AO = 110 m was relatively shallow, but the overall characteristics of the changes were basically the same as those reflected by AO = 200 m, and the corresponding positions of the anomalies of the orthogonal points basically coincided with each other. Taking AO = 200 m as an example, low resistance anomalies occurred at about 1,080 points and 1,540 points. Corresponding to point 1,080, it was interpreted that there might be an F6 fault structure with an NWW trend and spread along the N5 profile at point 1,090, N2 profile at point 1,680, N3 profile at point 1,080, and N1 profile at point 1,080. Corresponding to point 1,540, it was interpreted that there might be an F7 fault structure with a low resistance anomaly here. Corresponding to point 1,540, it was interpreted that there might be an F7 fracture structure here with an NWW trend and spread along the N4 profile at point 1,810, the N3 profile at point 1,520, and the N1 profile at point 1,540. The N3 geological profile was a comprehensive profile with an orientation of 37°. The anomalies reflected by each method were in good agreement with the inferred fracture structure. From the joint N3 profile, the change characteristics of AO = 110 m and AO = 200 m were basically the same, and the corresponding locations of the quadrature anomalies were basically the same as well. Take AO = 200 m as an example, there were two low-resistance anomalies corresponding to 1,080 points and 1,520 points, respectively. Corresponding to point 1,080, it was assumed that there was an F6 fracture structure here along the NWW trend, spreading along the N5 profile at point 1,090, N2 profile at point 1,680, N3 profile at point 1,080, N1 profile at point 1,080. Corresponding to point 1,520, it was assumed that there was an F7 fracture structure here along the NWW trend and spread along the N4 profile at point 1,810, the N3 profile at point 1,520, and the N1 profile at point 1,540 (Figure 7).

Hydrological evolution analysis

(1) ZK01 Hydrological logging

The actual hole depth of ZK01 was 300.56 m, and the hydrological logging methods basically reached the bottom of the hole. The logging curves and comprehensive interpretation results were shown in Figure 9. Taking the 150 – 200 m section corresponding to ZK01 as the reference, the values of apparent resistivity, natural potential, and natural gamma were generally stable around 350 – 400 Ω-m, 20 – 25 mv, and 250 – 300 API, respectively, which were reflections of intact granodiorite. For ZK01, the logging value of intact granodiorite was the normal background field, and the aquifer could be divided into two types according to the characteristics of its hydrological logging anomalies, which included one about 75 – 85 m, and the other about 85 m and above. The hydrological logging curves of 75 – 85 m demonstrated obvious low resistivity, obvious rightward deviation of the natural potential, and low natural gamma anomalies with the corresponding amplitude of change of about -200 Ω-m, 10 mv, and -40 API, respectively, which might be due to the relatively broken of this section and could be interpreted as fracture structure, showing good water permeability and strong water enrichment. About 85 – 138 m and 256.5 – 300 m, the hydrographic curve showed low resistivity, natural potential right deviation anomalies, which could be interpreted as the development of bedrock fissures, showing good water permeability and weak water enrichment.

(2) ZK02 Hydrological logging

The actual depth of ZK02 was 194.40 m. Unlike ZK01, ZK02 was seriously broken in the middle and collapsed at the bottom. The maximum depth of hydrological logging was 142.06 m, and the logging curve and comprehensive interpretation results were shown in Figure 9. ZK02 hydrological logging reflected objectively and clearly, and the anomalous characteristics were obvious, which could be used as a typical case of hydrological logging for identifying aquifers in granite areas. Corresponding to about 52 – 82 m, the apparent resistivity, natural potential, and natural gamma values were very stable about 750 – 800 Ω-m, -2 - 2 mv, and 420 - 440 API, respectively, which were reflections of intact granodiorite. For ZK02, with the logging value of intact granodiorite as the normal background field, the aquifer could be classified into three kinds of aquifers according to its hydrological logging anomalies including one about 34 – 52 m, which exhibited obvious low resistivity and obvious rightward deviation of natural potential and was interpreted as a weathered fissure, showing good water permeability and weakly enriched water. The second one was about 82 - 96.5 m and 130 - 142 m, which exhibited obvious low resistivity, obvious right deviation of natural potential, and low natural gamma anomaly, corresponding to the amplitude of change of about -400 Ω-m, 30 mv, and -100 API, respectively. The lifting drilling process brought out a large number of conglomerate-like crushed material, interpreted as a fracture structure, showing that the permeability of the water was good for strong water enrichment. The third one was about 96.5 – 130 m, which showed low resistivity and natural potential of the anomaly of the right deviation and could be interpreted as the development of bedrock fissures, indicating good permeability for weakly water enrichment.

Conclusion

In addition to the absence of backbone water storage projects in the upper reaches of the Qingjiang River, there are also serious water obstructions in the downstream drop holes and serious siltation or bank failure in some sections of the river, which makes the comprehensive management of the Qingjiang River system particularly urgent. This research proposed a comprehensive physical exploration method based on a Gaussian hybrid clustering algorithm. The river evolution law of the Qingjiang River Basin was then explored. The category-boundary algorithm was based on the Gaussian mixture model to further refine the classification results to realize the automatic division of the test data

of the opposite wave method and electric method. The feature matrix was extracted from the time and variable dimensions and then clustered by GMM from the perspective of probability distribution. The proposed algorithm had two parameters, r and s, and it was time-consuming to manually select the parameters and might not be able to obtain the global optimum, so it was worthwhile to further research on how to select the optimal parameters in an adaptive way. The joint dissection work was arranged on the tectonic belt of aeromagnetic interpretation, and the measurement results identified the P1 tectonics spreading in the east-west direction, which was reflected as the negative pole H3-1 in the aeromagnetic results. The tectonics was nearly upright with wide width, and there was a high-resistivity rock body exposed or shallowly buried in the north side. Taking the intersection point of joint dissection as the key target and combining it with the comparative analysis of the high-density electric method could effectively locate the river evolution law in this area, which was of guiding significance to the exploration and research in similar areas.

Acknowledgements

This research was supported by the project “Research on Electromagnetic Modeling and High-Efficiency Detection Techniques for Different Types of Geothermal Resources in Hunan Province” (Grant No. HNGSTP202521).

References

1. Jaaz ZA. 2019. Clustering algorithms subjected to K-means and Gaussian mixture models on multidimensional data sets. Probabilistic Models. 7(2):448-457.
2. Huang ZK, Chau KW. 2008. A new image thresholding method based on the Gaussian mixture model. Appl Math Comput. 205(2):899-907.
3. He X, Cai D, Shao Y, Bao H, Han J. 2010. Laplacian regularized Gaussian mixture model for data clustering. IEEE T Knowl Data En. 23(9):1406-1418.
4. Liu Z, Song YQ, Xie CH, Tang Z. 2016. A new clustering method of gene expression data based on multivariate Gaussian mixture models. Signal Image Video P. 10:359-368.
5. Deng J, Zhang Y, Li Y, Xia N, Qi ZH, Gao T. 2024. A dynamic equivalent modelling method for doubly-fed wind farms based on Gaussian mixture model clustering. J Sol Energ-T ASME. 45(01):342-350.
6. Dong X, Wan RX, Miao ZQ, Yue XD. 2023. A three-branch Gaussian hybrid clustering algorithm based on shadow sets. Journal of Guangxi University (Natural Science Edition). 48(04):958-971.
7. Maugis C, Celeux G, Martin-Magniette M. L. 2009. Variable selection for clustering with Gaussian mixture models. Biometrics. 65(3):701-709.
8. Malsiner-Walli G, Frühwirth-Schnatter S, Grün B. 2016. Model-based clustering based on sparse finite Gaussian mixtures. Stat Comput. 26(1):303-324.
9. Patel E, Kushwaha DS. 2020. Clustering cloud workloads: K-means *vs* Gaussian mixture model. Procedia Computer Science. 171:158-167.
10. Ouyang M, Welsh WJ, Georgopoulos P. 2004. Gaussian mixture clustering and imputation of microarray data. Bioinformatics. 20(6):917-923.
11. Su T, Dy JG. 2007. In search of deterministic methods for initializing K-means and Gaussian mixture clustering. Intell Data Anal. 11(4):319-338.
12. Viroli C, McLachlan GJ. 2019. Deep Gaussian mixture models. Stat Comput. 29:43-51.
13. Yang MS, Lai CY, Lin CY. 2012. A robust E.M. clustering algorithm for Gaussian mixture models. Pattern Recognition. 45(11):3950-3961.
14. Yang J, Liao X, Yuan X, Llull P, Brady DJ, Sapiro G, *et al*. 2014. Compressive sensing by learning a Gaussian mixture model from measurements. IEEE Trans Image Process. 24(1):106-119.