Comprehensive Study on the Electrical Characteristics and Full-Spectrum Tracing of Water Sources in Flooded Coal Mines

Donglin Dong [1,2] , Qiang Wu[1,2]

[1]College of Geoscience and Surveying Engineering, China University of Mining & Technology-Beijing, Beijing 100083, China

[2]Inner Mongolia Research Institute, China University of Mining and Technology (Beijing), Ordos 017004, China

Abstract

This study proposes a novel spectral tracing technique to identify inrush water sources in Donghuantuo coal mine. Electrical characteristics of the mine floor were analyzed, and spectral data from main aquifers were used to build a source database. A CSSOA-optimized random forest model (CSSOA-RF) was developed, achieving 100% accuracy in tests. The method enables rapid and reliable identification of single and mixed water sources, offering technical support for mine safety and water hazard prevention. While effective in this relatively simple geological setting, further validation is needed for broader application.

Keywords: Identification of Inrush water Sources; Spectral Tracing Technology; Chaotic Sparrow Search Optimization Algorithm (CSSOA)

Introduction

Traditional methods for studying inrush water, which rely on inrush water data and physical parameters, have proven effective in distinguishing relatively simple water sources. However, in cases where the water quality is complex and physical parameters are similar, or where the water quality of different inrush sources is alike, traditional methods often fail. This necessitates the use of alternative approaches for identifying inrush water sources (Sun 1965; Sun and Zheng 1996). By analysing the basic components of aquifers and the chemical composition of water samples from inrush points, suitable discriminant functions are selected, and discriminant formulas are established based on specific criteria. These formulas are then used to classify unknown samples, employing methods such as fuzzy mathematics, grey relational analysis, extension identification, Geographic Information Systems (GIS), support vector machines, and artificial neural networks (Panagopoulos *et al.* 2016; Zhang *et al.* 2019; Ju and Hu 2021).

In recent years, with the rapid advancement of computer technology, neural networks have gradually been used to identify inrush water sources in mines. Common artificial neural network models include BP neural networks, RBF neural networks, ELM, and Elman neural networks, which are valuable for accurately identifying multiple water sources in mines. Deep learning algorithms represent an evolution of artificial neural networks, with common approaches including deep neural network (DNN) analysis, convolutional neural network (CNN) analysis, and probabilistic neural network (PNN) analysis. Scholars often use these methods to determine the origin of water samples. Optimization algorithms such as genetic algorithms, ant colony algorithms, and particle swarm algorithms are frequently used for function optimization and combinatorial optimization of discriminant models. For example, Zhang Di applied a genetic algorithm-optimized support vector machine to identify inrush water sources, which improved the accuracy of parameter selection for support vector machines in inrush water source identification. However, this method still has drawbacks, such as tracing detection times exceeding one hour, inability to trace mixed water and its proportions, and loss of the optimal rescue window post-inrush.

This study aims to explore a new method that abandons the traditional approach of concentration testing. Instead, it focuses on the rapid and accurate identification of inrush water sources, proposing a novel method for identifying mixed inrush water sources.

Methods

Study Area and Water Sampling

Donghuantuo mine is located in the city of Tangshan in North China's Hebei Province. The southeast wing of the mine extends 13.5 km in strike length and 3 km in dip width, while the northwest wing stretches 8 km in strike length and 0.5 km in dip width. The mining area covers 40.5 km². The terrain within the mining boundary is quite flat, with no rivers traversing the area. Additionally, there are no surface water systems within the Donghuantuo mining boundary, as illustrated in Fig. 1. The Donghuantuo mining field hosts multiple aquifers, including the Quaternary alluvial pore-confined aquifer, the Carboniferous–Permian sandstone fissure-confined aquifer, and the Middle Ordovician limestone karst fissure-confined aquifer. These aquifers are divided into seven aquifer groups, as shown in Fig. 1.

In the preliminary phase of this study on spectral tracing of inrush water sources, four water samples were collected from the primary aquifers of the Donghuantuo mine. These samples were taken from the 12–2–14–1 aquifer, the 5 coal roof aquifer, the Quaternary aquifer, and the Ordovician limestone aquifer, with each sample measuring 1000 mL. Each of the four samples was labeled and numbered accordingly.

Transient Electromagnetic Method for Inrush Water Sources

In this study, the enhanced TEM67 transient electromagnetic method was employed. The transmitter used was the TEM57-Mk2, augmented with two TEM67 power modules, capable of reaching an emission voltage of up to 240 V and a maximum emission current of 28 A. This configuration allowed the use of large-sized or large effective area transmitter coils, achieving greater exploration depths. The table below compares the transmission power of the TEM67 and the enhanced TEM67. It shows that with a transmission frame size of 2000 m × 2000 m (equivalent to an exploration depth of 2000 m), the magnetic moment M (I × S) of the enhanced TEM67 can reach up to 36,363,636 A·m².

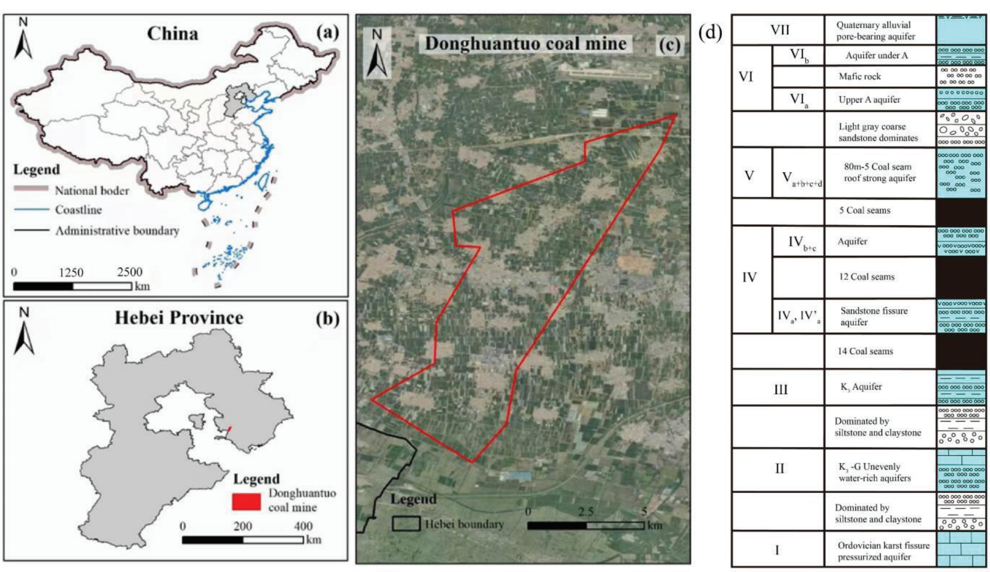

***Figure 1** Geographical location of the study area.*

Chaos Sparrow Search Optimization Algorithm

The sparrow search algorithm (SSA) faces challenges such as a tendency to get trapped in local optima, a lack of randomness in the search process, and slow convergence. To address these issues, researchers including Xin introduced the tent chaotic sequence and Gaussian distribution into the SSA, forming the chaos sparrow search optimization algorithm (CSSOA). This method uses tent chaotic mapping during the population initialization stage to maintain a uniform distribution of initial individuals. When the population exhibits convergence or divergence, chaotic perturbations and Gaussian mutations are applied to alleviate local optima problems.

Random Forest (RF) Algorithm

Random forest (RF) is an ensemble learning algorithm commonly applied to classification tasks. It constructs a large number of decision trees from training samples to form a collective discriminative model, which is then used for classifying unknown samples. In UV–Visible spectrum classification, RF can effectively handle non-linear relationships and, through its built-in feature importance metrics, select the most informative features from large datasets, thus enabling efficient processing of high-dimensional spectral data while ensuring classification accuracy.

Electrical Characteristics of Water-Rich Inrush Water Source

Electrical Characteristics of Water-Rich Inrush Water Sources at the 20223 and 3015 Working Face

The current retreating working faces are 20223 and 3015. To ensure safe and efficient mining, it is necessary to conduct electrical resistivity surveys on the roof and floor of these working faces to accurately delineate their water-bearing properties. The enhanced TEM67 transient electromagnetic method was used to detect water-rich areas at the 20223 and 3015 working faces. These surveys were complemented by data from four drill holes and tunnel exposure data. The water-bearing characteristics of the roof and floor of these working faces are illustrated in the following Fig. 2.

According to Fig. 2, a relatively low-resistivity zone, designated as DF-1, was identified within the 0 to 80 m vertical detection range of the floor at the 20223 working face. Analysis of related hydrogeological data suggests that this zone corresponds to a sandstone fissure-confined aquifer. Comparisons with the fault lines in the profile and cross-section diagrams indicate that this zone is likely connected by faults, warranting focused investigation. Verification through four drill holes confirmed that this is indeed a fault zone.

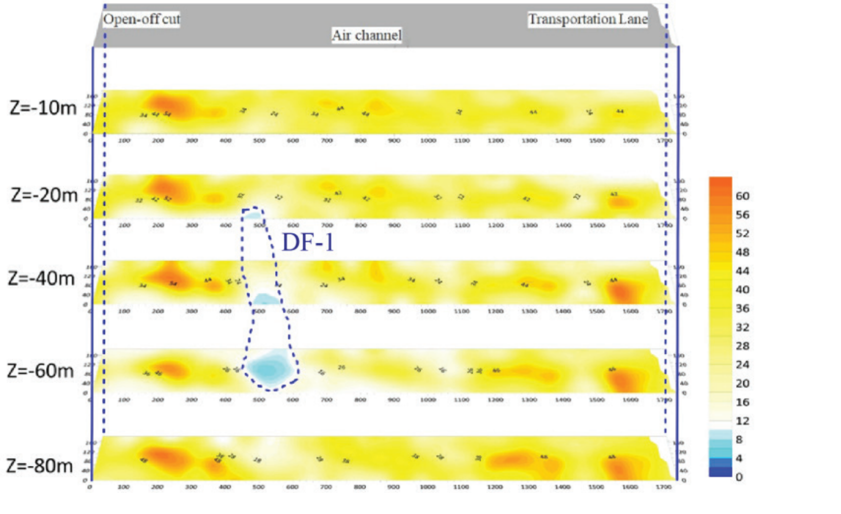

Figure 2 *3D Electrical resistivity tomography (ERT) image of the 20223 working face floor.*

Spectral Characteristics of Various Inrush Water Sources

A vacuum filtration apparatus was used to filter each water sample solution through a 0.45 μm PTFE hydrophilic filter membrane to remove particulate impurities. The filtered water samples were then dispensed into sample bottles that had been washed with ultrapure water and air-dried. To increase the sample size for subsequent machine learning, each type of water sample was divided into eight sample bottles, resulting in eight sets of spectral data for each type of water sample. To ensure accuracy, each sample was measured three times, and the arithmetic mean of these measurements was used as the spectral curve data for that sample. For ease of data processing and result interpretation, samples were assigned hierarchical numbers, e.g., the first sample of water sample type 1 was labeled 1–1, and its two repeated measurements were labeled 1–1–1 and 1–1–2, respectively. In total, 32 sets of water samples were prepared for subsequent classification model training in this tracing study.

The prepared samples were subjected to spectrophotometric measurement. After the instrument completed its self-check and warm-up, the measurement parameters of the spectrophotometer were set: the starting wavelength was set to 320 nm, the ending wavelength to 650 nm, and the scanning interval to 1.0 nm. Zero-line calibration was then performed. Each sample was measured sequentially, with three repeated measurements. The data were then imported into a computer using a USB drive and labeled accordingly. After measuring each type of sample, the cuvette and other instruments were cleaned with ultrapure water. Upon completing the measurements of all water samples, the repeated measurement data for each sample were organized and averaged arithmetically, resulting in 32 sets of single-source spectral data, as shown in Fig. 3.

Analysis of Fig. 4 reveals significant differences in the spectral data of the four aquifer types within the measured wavelength range of 320–650 nm. The absorbance

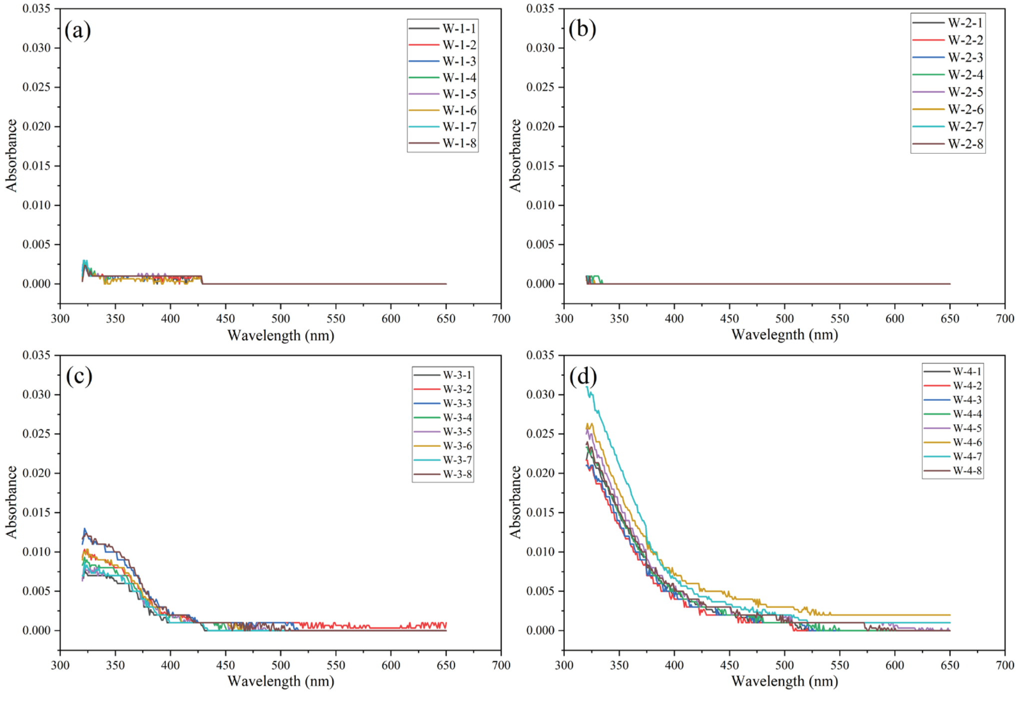

__Figure 3__ Original spectral data of each aquifer. (a) 12-2~14-1 aquifer water samples, (b) 5 coal roof aquifer water samples, (c) quaternary aquifer water samples, (d) Ordovician limestone aquifer water samples.

fluctuations for water samples W-1 and W-2 are minimal, distributed within the range of 0 to 0.005, with most absorbance values being 0. Specifically, the absorbance for W-1 drops to 0 after 429 nm and remains unchanged, while for W-2, it remains at 0 beyond 335 nm. In contrast, water samples W-3 and W-4 exhibit significant absorbance variations, with both reaching maximum absorbance values around 320 nm and then gradually decreasing to their minimum values. The absorbance range for W-3 is between 0 and 0.015, while for W-4, it is between 0 and 0.0325. The absorbance variations for all four water samples are concentrated in the ultraviolet region (<400 nm), aligning with the visual characteristics of colorless and transparent water samples.

Water samples W-3 and W-4, which are from shallower locations, experience more frequent water–rock interactions compared to the deeper W-1 and W-2 samples. As a result, W-3 and W-4 have higher ion concentrations and more complex ionic compositions, leading to more pronounced spectral variations. The distinct spectral differences among the four water samples ensure the accuracy of the subsequent machine learning tracing model's classification and identification.

Based on the hydrogeological conditions, hydraulic connections, and actual geological situations of the four main aquifers sampled from the Donghuantuo mine, the study investigated the mixing of the Quaternary water sample W-3 and the 12–2~14–1 aquifer water sample W-1. The preparation method for these mixed samples was consistent with that of the single-source samples. W-1 and W-3 samples were mixed in ratios of 1:9, 2:8, ..., 8:2, and 9:1. To ensure sufficient training samples for the subsequent tracing model, 10 samples were prepared for each mixing ratio, resulting in a total of 90 mixed spectral data entries, as shown in Fig. 4.

Construction and Evaluation of the Tracing Identification Model

Using the preprocessed sample data, we constructed the CSSOA-RF spectral tracing identification model based on the Python language. The CSSOA algorithm was employed to optimize the key parameters of the RF model, achieving adaptive parameter optimization for the tracing identification model. This phase involved steps such as da-taset partitioning, CSSOA algorithm optimization, and the construction of the CSSOA-RF tracing identification model.

The parameters of the RF model, including *n_estimators, max_depth, min_samples_leaf*, and *min_samples_split*, were optimized based on the training set. The CSSOA algorithm's objective function was set to the average error obtained from five-fold cross-validation on the training set. The population size was set to 60, with a discoverer ratio of 0.7 and a warning reconnaissance ratio of 0.2. The optimization process is illustrated in Fig. 5. After 67 iterations, the model achieved the minimum average error value of 0.1596, which remained stable. Therefore, the parameter combination at this point was chosen as the optimal parameter values for the RF model.

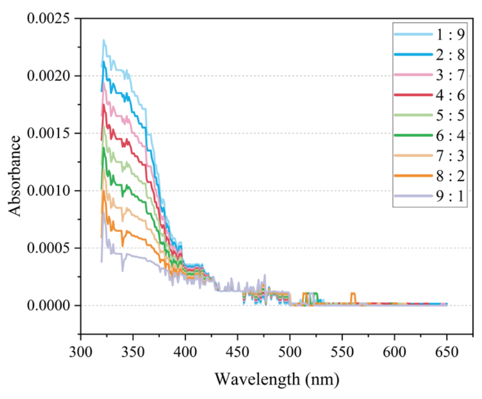

Figure 4 Original spectral data of different aquifers.

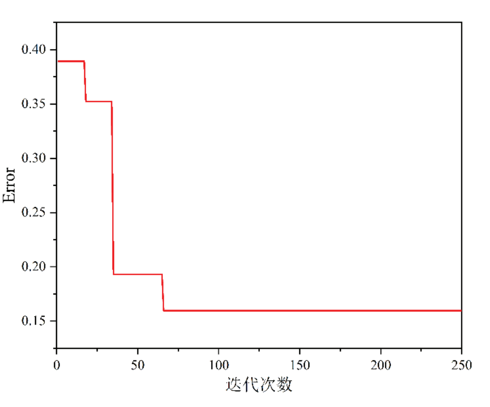

Figure 5 CSSOA parameter optimization curve.

Conclusion

CSSOA-RF Spectral Tracing Identification Model: By using the chaos sparrow search optimization algorithm (CSSOA) to optimize the key parameters of the random forest (RF) model, the CSSOA-RF spectral tracing identification model was constructed. The CSSOA algorithm, through the introduction of chaotic perturbations and Gaussian mutations, overcame the limitations of traditional sparrow search algorithms that tend to get trapped in local optima. This improved optimization efficiency and the model's global search capability. The optimized RF model achieved optimal parameter selection, providing reliable technical support for tracing identification. In the test set, the CSSOA-RF model demonstrated excellent classification performance, achieving 100% accuracy, with only one sample misclassified, indicating the model's outstanding classification ability and generalization performance. The confusion matrix visualization further confirmed the model's accuracy and stability in practical applications. This result validates the effectiveness and practicality of spectral tracing technology in identifying mine inrush water sources.

Innovative and Reliable Method for Inrush Water Source Identification: This research provides a new, more accurate, and reliable method for identifying inrush water sources, addressing the shortcomings of traditional methods in handling complex water quality conditions. The model helps in quickly identifying inrush water sources in coal-bearing regions of North China, reducing disaster losses and enhancing mine safety. Additionally, the spectral database and CSSOA-RF model developed in this study offer valuable references for future related research.

Thanks to the National Key Research and Development Program 2023YFC30121 for funding this study.

References

Sun HL (1965) Characteristics of karst water inrush channels in coal seam fractures. China Coal Soc 2: 51–57.

Sun FY, Zheng SY (1996) Research on the characteristics of mine water quality and its application in identifying water sources. Coal Sci. Technol 10: 27–30.

Panagopoulos G, Angelopoulou D, Tzirtzilakis E, Giannoulopoulos P (2016) The Contribution of Cluster and Discriminant Analysis to the Classification of Complex Aquifer Systems. Environ Monit Assess 188: 591. doi:10.1007/s10661-016-5590-y

Zhang H, Xing H, Yao D, Liu L, Xue D, Guo F (2019) The Multiple Logistic Regression Recognition Model for Mine Water Inrush Source Based on Cluster Analysis. Environ. Earth Sci 78–612. doi:10.1007/s12665-019-8624-2

Ju Q, Hu Y (2021) Source Identification of Mine Water Inrush Based on Principal Component Analysis and Grey Situation Decision. Environ Earth Sci 80: 157. doi:10.1007/s12665-021-09459-z