Academic Journal of Emerging Technologies
ISSN Print: 3104-4417
ISSN Online: 3104-4425
CD ISSN: 3104-4433
https://doi.org/10.63313/AJET.9010

Research on Mineral Inversion Method Based on Elemental Logging

Ruonan Zhang[1.2]
[1]School of Earth Sciences and Engineering, Xi 'an Shiyou University, Xi 'an, 710065, China
[2]Shaanxi Key Laboratory of Petroleum Accumulation Geology, Xi 'an Shiyou University, Xi 'an, 710065, China
Email: 15836901839@163.com

How to cite this paper: Zhang, R. N. (2025). Research on Mineral Inversion Method Based on Elemental Logging. Academic Journal of Emerging Technologies, 1(2), 35-47. ISSN Print: 3104-4417; ISSN Online: 3104-4425.
https://doi.org/10.63313/AJET.9010
Published: 2025-10-13

Abstract

The elemental data and mineral data obtained from elemental logging and X-ray diffraction were used to invert the mineral content through five methods. The oxide closure model method is based on the theoretical stratigraphic model, and the core condition of setting the sum of the mass percentages of clay minerals, carbonates, and the quartz-feldspar-mica mixture (QFM) to 100% is used for modeling and solving. The non-negative least squares method, truncated singu-lar value decomposition method, and linear programming method are based on the conversion relationship model between elements and minerals proposed by Herron. Through the analysis of the lithological characteristics of the block, the element content inversion mineral content model is established based on the matrix solution method. The XGBoost machine learning algorithm automatically learns and captures the complex relationships between samples from the data samples to solve the problem. Taking well x as an example, the mineral content was calculated through the five inversion methods. The comparison of the cal-culation results with the whole rock analysis results shows that the variation trends of each method with the depth of the stratum are basically consistent with the whole rock analysis results. Among the four mineral inversion meth-ods, the processing result of the XGBoost machine learning algorithm is the best, with a mean square error of 0.91; the processing effects of the non-negative least squares method and the truncated singular value decomposition method are second, with mean square errors of 0.88 and 0.81, respectively; the mean square error of the linear programming method is relatively low, only 0.79.

Keywords

Elemental logging; X-ray diffraction; Mineral inversion; XGBoost; Conversion coefficient

1. Introduction

Element logging (XRF) excites the core, cuttings and other geological samples obtained from drilling with primary X-rays, causing the samples to produce fluores-

cent X-rays. Subsequently, by analyzing the number of pulses of the fluorescent X-rays in the detector, the types and contents of elements in the samples can be obtained. Currently, element logging instruments can measure the content data of more than 17 elements in cuttings.

X-ray diffraction (XRD) irradiates cuttings samples with X-rays to obtain various diffraction patterns. By analyzing the diffraction patterns, crystal structure diagrams can be obtained. Based on the correspondence between minerals and crystal structures, the types and contents of minerals can be obtained. X-ray diffraction logging can obtain the content information of more than 30 minerals in the strata[1].

Studies on the abundance of elements in the Earth's crust show that although there are more than 90 elements in the Earth's crust, the relative average contents of elements are extremely uneven. The most abundant element O can be 10^{17} times greater in content than the least abundant element Rn. If the elements are arranged in the order of decreasing Clarke value, the first two most widely distributed elements (O and Si) account for 76.5% of the total mass of the Earth's crust, and the first 10 elements (O, Si, Al, Fe, Ca, Na, K, Mg, Ti, and Mn) account for 99.58%. The mass of the remaining elements does not exceed 0.5% of the total mass of the Earth's crust. Similarly, although more than 2,200 minerals have been discovered in the Earth's crust rocks, only about 10 types of minerals are common. When the chemical composition of minerals is relatively stable, the percentage content of each element in the minerals remains basically unchanged. This is the prerequisite for converting elements into minerals and also the prerequisite for selecting indicator elements of minerals. Therefore, as long as the contents of these main elements are accurately measured, the types and contents of minerals in the Earth's crust rocks can be identified. Based on this, the content of minerals can be quantitatively obtained by establishing the conversion relationship between elements and minerals.

Through the identification of rock thin sections and the statistical analysis of X-ray diffraction results of a certain block of strata, it is known that the elemental content differences of different rock components in the metamorphic rocks of the target layer section in this area are relatively obvious. The main lithology is sandstone and mudstone. By applying the most optimized analysis method, the mineral composition content of the strata can be quantitatively obtained.

2. Data Processing

Since the data obtained from elemental logging represents the compound mass fraction of a single element in the formation, normalization processing should be carried out before calculation. Based on the distribution pattern of elements in nature, the types of minerals in whole rock analysis, and the XRF element content, the main influencing elements are selected. The oxide forms of the seven main XRF measured elements, namely silicon (Si), aluminum (Al), iron (Fe), calcium (Ca), sodium (Na), magnesium (Mg), and potassium (K), are normalized separately according to silicon

dioxide (SiO_2), tricalcium aluminate (Al_2O_3), ferrous carbonate ($FeCO_3$), calcium carbonate ($CaCO_3$), potassium oxide (K_2O), magnesium carbonate ($MgCO_3$), and sodium oxide (Na_2O), and then the single element mass percentage content is calculated and processed.

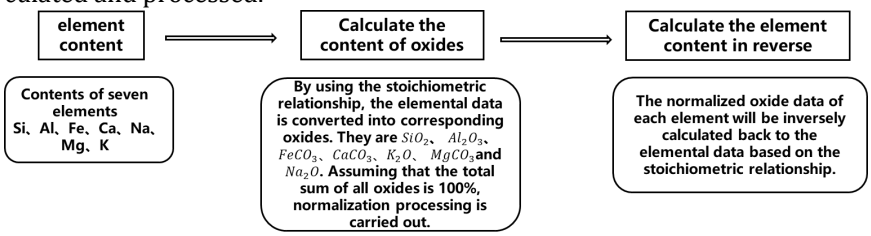

Figure 1. Flowchart of Element Logging Data Preprocessing Process

The data obtained through X-ray diffraction represent the compound mass fractions of each mineral in the stratum. Before establishing the model for the mineral inversion method, based on the main mineral types of the metamorphic rock granulite and gneiss stratum, normalization processing was carried out for quartz, potassium feldspar, plagioclase, illite, chlorite, and calcite.

$$\bar{x} = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

3. Mineral inversion method

3.1. Oxide closed model

The principle of the oxide closure model method is to assume that the sum of the clay mineral content, carbonate content, and the mass percentage of the mixture of quartz, feldspar, and mica (QFM) in the composition of the mineral is 100%, and then inversely solve for the three types of minerals. Previous studies have proved that the content of clay minerals has a good positive correlation with the content of aluminum, quartz has a good correlation with silicon, and carbonate rocks have a good correlation with calcium. Therefore, the content of these elements can be used to determine the corresponding mineral content. The mineral content determined by this method is the mineral content of the rock skeleton, and the unit is weight percentage rather than volume percentage. The oxide closure model method is simple in calculation, but the mineral types inversely solved have a large limitation and can only roughly divide and calculate the three major types of minerals[2].

(1) Estimating clay content

It is known that the correlations between Al, Fe, K, Na and clay content are quite obvious. A calculation model for clay content can be established through mathematical statistical methods.

$$Wclay = 15.3 + 2.61 \times WAl + 5.48 \times WFe - 9.13 \times WK - 2.96 \times WNa \quad (2)$$

(2) Estimating carbonate content

Carbonate minerals mainly include calcite and dolomite. The main constituent element of calcite is calcium (Ca), while the main constituent elements of dolomite are calcium (Ca) and magnesium (Mg).

$$WCar = 3.67 + 5.44 \times WCa - 2.49 \times WMg \tag{3}$$

(3) Estimating the content of QFM

The remaining part of the formation is the sandy part, which mainly consists of quartz, feldspar and mica (QFM).

$$WQFM = 100 - WClay - WCar \tag{4}$$

The calculation results of mineral content are shown in Figure 2.

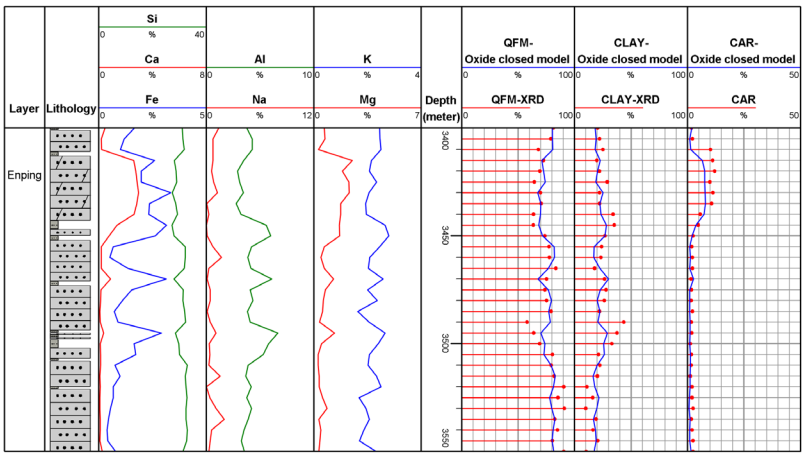

Figure 2. Calculation results of the oxide closed model

3.2. Mathematical Modeling

Establishment of the relationship between elements and minerals: Due to the different rock types of metamorphic rocks and the varying but relatively fixed contents of each element, it is necessary to establish a quantitative relationship between the two first by calculating the mineral content based on the element content. Herron et al.[3] conducted a large number of experiments and analyses. They first analyzed the contents of the main elements in the sample rock cores, then determined the contents of the main minerals through XRD analysis, and finally statistically analyzed the quantitative relationship between these elements and minerals. After conducting experiments, analyses, statistics and research on a large number of rock cores worldwide, Herron proposed a relationship formula for converting element content to mineral content:

$$E_i = \sum_j^n C_{ij} M_j \tag{5}$$

In the formula, Ei represents the content of the i-th element, expressed as a percentage; Cij indicates the content of the i-th element in the j-th mineral, which is the conversion coefficient between the element content and the mineral content, and is also one of the key points of this method; Mj represents the content of the

j-th mineral, expressed as a percentage. By inverting the formula, the desired mineral percentage content can be obtained using the element percentage content.

$$[M] = [C]^{-1}[E] \tag{6}$$

In the formula, [E] is the matrix composed of the contents of each element; [M] is the matrix composed of the contents of each mineral; $[C]^{-1}$ is the inverse matrix of the conversion coefficient matrix [C].

Table 1. Herron General Conversion Coefficient

	Na	Mg	Al	Si	K	Ca	Fe
Quartz	0	0	0	46.7	0	0	0
K-feldspar	0	0	9.69	10.09	14.05	0	0
Plagioclase	8.78	0	19.396	32.06	0	14.406	0
Illite	0	0	17.81	23.42	5.15	0	0
Chlorite	10	7.82	4.34	13.56	0	0	25.97
Dolomite	0	12.9	0	0	0	28	0

3.2.1 Non-negative Least Squares Method

Element content inversion mineral content basic model [E]=[C][M], that is, solving the equation system Ax=b (where $A \in C^{m\times n}$, $b \in C^m$ is given, and $x \in C^n$ is the unknown to be solved), when the equation system is a contradictory one, the optimal solution needs to be solved. Although the overdetermined equation system does not have an exact solution, by constructing an error function, the total error can be minimized to obtain an optimal solution[4]. The least squares method (NNLS) is the most commonly used method for solving the optimal solution of the equation system, that is, if $\hat{x}$ satisfies:

$$\| A\hat{x} - b \|^2 = \min_{x\in c^n} = \| Ax - b \|^2 \tag{7}$$

then $\hat{x}$ is called the least squares solution of the overdetermined equation system.

Due to the negative values of the mineral content calculated by the least squares method, and in actual situations, the mineral content cannot be negative, therefore, the least squares method is optimized and the mineral content is inverted using the non-negative least squares method[5]. The non-negative least squares method is a constrained version of the least squares method, requiring that all coefficients cannot be negative, which is more in line with the actual situation. The calculation results of the mineral content are shown in Figure 3.

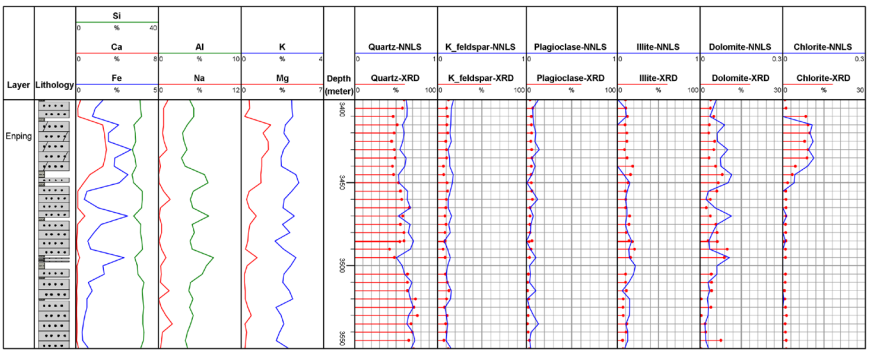

Figure 3. Results of Least Squares Calculation

3.2.2 Truncated Singular Value Decomposition Method

The conversion relationship between elements and mineral contents can be expressed as Ax = b (where $A \in C^{m\times n}$, $b \in C^m$ is given, and $x \in C^n$ is to be found). The least squares solution of this relationship can be represented as:

$$X_{is} = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T b \tag{8}$$

Performing singular value decomposition on the conversion coefficient matrix yields:

$$A = \mathrm{USV}^T = \sum_{i=1}^{n} s_i\, u_i v_i^T \tag{9}$$

In the formula, $\mathrm{U} = [u_1, \cdots, u_n] \in C^{m\times n}$ and $\mathrm{V} = [v_1, \cdots, v_m] \in C^{n\times n}$ are both orthogonal matrices; u_i and v_i represent the i-th column of U and V respectively; S = diag(s_1, s_2, ..., s_n), where s_i $(i = 1, 2, \cdots, n)$ are the singular values of A. Substituting formula (9) into formula (8) yields:

$$X_{ls} = \mathrm{V}(\mathrm{S}^T\mathrm{S})^{-1}\mathrm{S}^T\mathrm{U}^T b = \frac{\sum_{i=1}^{n} u_i^T b}{s_i v_i} \tag{10}$$

The corresponding mean square error is:

$$MSE(X_{ls}) = \sigma_0^2 (A^T A)^{-1} = \frac{\sigma_0^2 \sum_{i=1}^{n} 1}{s_i^2} \tag{11}$$

From the above derivation, it can be seen that the least squares method and the singular value decomposition method (TSVD) are equivalent. When matrix A is a singular matrix, there will be one or more singular values of matrix A that are close to 0. From equation (10), it can be known that when the singular values are very small or close to 0, the mean square error of the least squares solution will become very large. That is to say, the solution at this time is not the optimal solution[6]. This requires removing those singular values that are very close to 0. By this means, we can reduce the mean square error of the solution obtained by the least squares method and optimize the solution, increasing the adaptability of the solution. If there are n - k singular values close to 0 in the singular values obtained by the decomposition of matrix A, the solution obtained by removing these singular values can be expressed as:

$$X_{TSVD} = \frac{\sum_{i=1}^{k} u_i^T b}{s_i v_i} \tag{12}$$

In truncated singular value decomposition, the truncation parameter k is used to determine which k singular values, left singular vectors, and right singular vectors in the singular value decomposition are to be retained, thereby performing a low-rank approximation of the original matrix and achieving the goals of reducing ill-conditioning and optimizing the solution. This paper uses the L-curve method to determine the truncation parameter k. The L-curve method is a method proposed by Hanson for determining regularization parameters. This parameter is determined by comparing the curves of the residual modulus $\|Ax_k - b\|$ and $\|x_k\|$ in the log-log scale. Usually, the inflection point of the L-curve can be represented by the point with the maximum curvature, and $\eta = \|x_k\|^2,\ \rho = \|Ax_k - b\|^2,\ \mu = \log\eta,\ v = \log\rho$. The curvature of the L-curve is defined as:

$$c(k) = 2\frac{v'\mu'' - v''\mu'}{((v')^2 + (\mu')^2)^{\frac{3}{2}}} \tag{13}$$

According to the singular value decomposition theorem, the matrix $A = \sum_{i=1}^{n} s_i\, u_i v_i^T$. Through the Tikhonov regularization method, we can obtain:

$$x_k = \sum_{i=1}^{n} f_i \frac{u_i^T b}{s_i} v_i \tag{14}$$

In the formula, $f_i = \frac{s_i^2}{s_i^2 + k^2}$. Then there is:

$$\eta = \|x_k\|^2 = \sum_{i=1}^{n} \left(f_i \frac{u_i^T b}{s_i}\right)^2 \tag{15}$$

$$\rho = \|Ax_k - b\|^2 = \sum_{i=1}^{n} \left((1 - f_i) u_i^T\ b\right)^2 \tag{16}$$

Let$\beta_i = u_i^T b$. Then, from $\mu' = \frac{\eta'}{\eta}$, $v' = \frac{\rho'}{\rho}$, we obtain:

$$\mu' = -\frac{4}{k}\sum_{i=1}^{n}(1 - f_i)\, f_i^2 \frac{\beta_i^2}{s_i^2} \tag{17}$$

$$v' = \frac{4}{k}\sum_{i=1}^{n}(1 - f_i)^2\, f_i \beta_i^2 \tag{18}$$

$$\mu'' = \frac{d}{dk}\frac{\eta'}{\eta} = \frac{\eta''\eta - (\eta')^2}{\eta^2} \tag{19}$$

$$v'' = \frac{d}{dk}\frac{\rho'}{\rho} = \frac{\rho''\rho - (\rho')^2}{\rho^2} = -2k\eta' - k^2\eta'' \tag{20}$$

Substituting formulas (17), (18), (19), and (20) into formula (13), we obtain:

$$c(k) = 2\frac{\eta\rho}{\eta'}\frac{k^2\eta'\rho + 2k\eta\rho + k^4\eta\eta'}{(k^2\eta^2 + \rho^2)^{\frac{3}{2}}} \tag{21}$$

In the above formula, k represents the most appropriate truncation value. The calculation results of mineral content are shown in Figure 4 below.

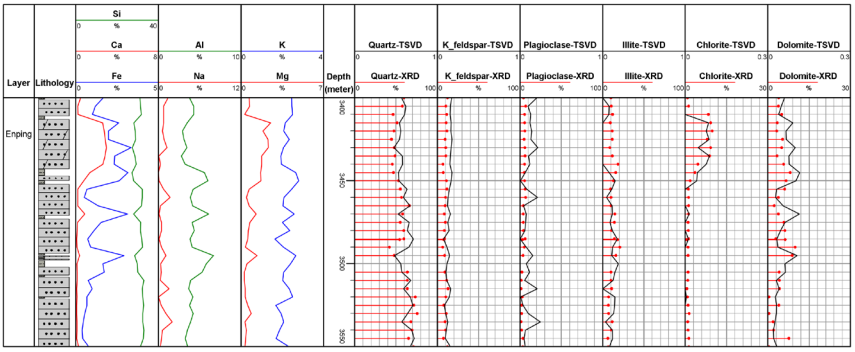

Figure 4. Results of Truncated Singular Value Decomposition Method

3.2.3 Linear Programming Method

Linear programming (LP) is a mathematical theory and method that studies the extremum problems of linear objective functions under linear constraints. Due to certain errors in measurement and analysis processes, there is a residual between the true content of stratum elements and the measured values. Based on the conversion relationship between element content and mineral content $\sum_j^n C_{ij}M_j$, we can obtain:

$$\begin{cases} C_{11}M_1 + C_{12}M_2 + \cdots + C_{1n}M_n = E_1 \\ C_{21}M_1 + C_{22}M_2 + \cdots + C_{2n}M_n = E_2 \\ \cdots\cdots \\ C_{m1}M_1 + C_{m2}M_2 + \cdots + C_{mn}M_n = E_m \end{cases} \tag{22}$$

Among them, E_i represents the content of the i-th element, where i = 1, 2, ... m ; C_{ij} represents the corresponding relationship between the known element content and the content of the mineral to be determined; M_j is the content of the j-th mineral to be determined. Since there will be certain errors between the measured values of stratum elements and the actual element contents, this is referred to as the residual. Assuming that the residual is represented by Δ, formula (22) can be modified as:

$$\begin{cases} C_{11}M_1 + C_{12}M_2 + \cdots + C_{1n}M_n + \Delta_1 = E_1 \\ C_{21}M_1 + C_{22}M_2 + \cdots + C_{2n}M_n + \Delta_2 = E_2 \\ \cdots\cdots \\ C_{m1}M_1 + C_{m2}M_2 + \cdots + C_{mn}M_n + \Delta_m = E_m \end{cases} \tag{23}$$

Then, the issue of conversion between elements and minerals can be transformed into an optimization problem under the condition of minimum residual sum of squares, as shown in the following equation:

$$\min S = \sum_{i=1}^m (E_i - C_{i1}M_1 - \cdots - C_{in}M_n)^2 = \sum_{i=1}^m \Delta_i^2 \tag{24}$$

Taking into account the actual situation, the mineral content in the strata is all greater than or equal to 0, and the total mineral content is less than or equal to 1, that is:

$$\begin{cases} M_1,\ M_2,\ \cdots,\ M_n \geq 0 \\ M_1 + M_2 + \cdots + M_n \leq 1 \end{cases} \tag{25}$$

Therefore, the method of solving problems through linear programming can trans-

form the problem-solving process into: using Equations (23) and (25) as constraints, and solving the minimum value problem of the objective function Equation (24). Taking into account the quadratic function:

$$\min S = \sum_{i=1}^{m} \Delta_i^2 \tag{26}$$

The calculation is quite difficult. Let$\boldsymbol{\Delta_i = \alpha_i - \beta_i}$ where $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$ are both non-negative. Then the function to be solved becomes:

$$\min S = \sum_{i=1}^{m} (\alpha_i + \beta_i) \tag{27}$$

The constraints are as follows:

$$\begin{cases} C_{11}M_1 + C_{12}M_2 + \cdots + C_{1n}M_n + \alpha_1 - \beta_1 = E_1 \\ C_{21}M_1 + C_{22}M_2 + \cdots + C_{2n}M_n + \alpha_2 - \beta_2 = E_2 \\ \cdots\cdots \\ C_{m1}M_1 + C_{m2}M_2 + \cdots + C_{mn}M_n + \alpha_m - \beta_m = E_m \\ M_1,\ M_2 \cdots M_n \geq 0 \\ M_1 + M_2 + \cdots M_n \leq 1 \\ \alpha_1,\ \alpha_2,\ \cdots,\ \alpha_m \geq 0 \\ \beta_1 . \beta_2 \cdots \beta_m \geq 0 \end{cases} \tag{28}$$

In summary, the basic steps for solving the mineral content using linear programming are as follows: (1) Determine the main mineral types based on the full rock analysis data from X-ray diffraction; (2) Obtain the content of elements Ei from the elemental logging data; (3) Based on the general conversion relationship and the conversion coefficient matrix trained using regional data; (4) Establish the objective function based on regional constraints. The calculation results of mineral content are shown in Figure 5.

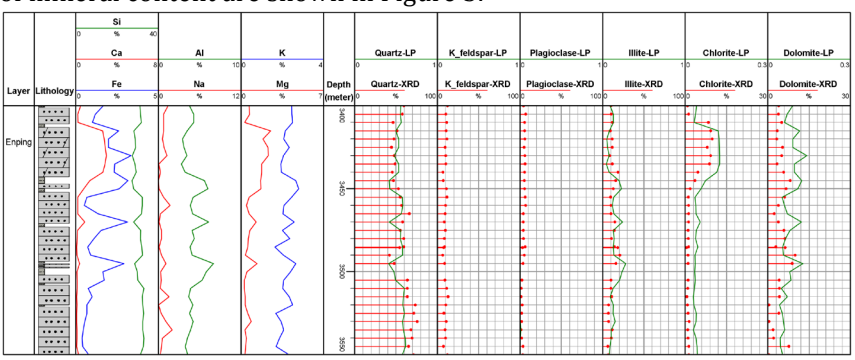

Figure 5. Results of Linear Programming Calculation

3.3. Machine learning model

The XGBoost (Extreme Gradient Boosting) algorithm is an efficient ensemble learning algorithm that has undergone a series of optimizations based on the Gradient Boosting Decision Tree (GBDT). The XGBoost algorithm can automatically learn and capture the complex relationships among data samples, and is widely used in regression, classification, and other problems. The core of the XGBoost algorithm is to construct multiple decision trees serially, and each new tree attempts to correct the

residuals of the previous tree. It uses the second-order Taylor expansion to approximate the target function, improving the convergence speed of the model, and incorporates regularization terms to prevent overfitting. The XGBoost algorithm employs some efficient optimization strategies to reduce the correlation between trees, enhancing the generalization ability of the model, and can better handle sparse data through the sparse sensing algorithm. Through these mechanisms, the XGBoost algorithm can extract effective features from the input element content data, handle the complex relationships between minerals and elements, and accurately invert the mineral content.

The objective function of the XGBoost algorithm is composed of the loss function and the regularization term:

$$\mathcal{L} = \sum_{i=1}^{n} l\left(y_i,\ \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{29}$$

In the formula, $l\left(y_i,\ \hat{y}_i\right)$: loss function, which measures the error between the predicted value and the true value; $\Omega(f_k)$: regularization term for a single tree, in the form of $\gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$, where T is the number of leaf nodes of the tree, w_j is the predicted value of the j-th leaf node, and γ and λ are regularization coefficients.

The generation of each new tree f_k is achieved through a greedy algorithm to split the feature space, and at each step, it fits the negative gradient of the previous model. For the t-th iteration, that is, when constructing the t-th tree, the current prediction of the model is $\hat{y}_i^{(t-1)} = \sum_{k=1}^{t-1} f_k\left(x_i\right)$, then the objective of the t-th tree to be fitted is:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i,\ \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \tag{30}$$

In the formula, $g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i,\ \hat{y}_i^{(t-1)}\right)$ represents the first-order derivative of the loss function with respect to the current average estimate; $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i,\ \hat{y}_i^{(t-1)}\right)$, representing the second-order derivative. The structure of the tree f_t is determined by optimizing this approximate objective, including the splitting point, the number of leaf nodes, and the weights w_j of the leaf nodes.

Overall, the function that the XGBoost algorithm aims to achieve is that for the input sample x (element content data), the final predicted value ŷ (mineral content data) is the sum of the prediction values of K trees:

$$\hat{y}_i = \sum_{k=1}^{K} f_k\left(x_i\right) \tag{31}$$

In the formula, f_k represents the kth decision tree, and x_i is the i-th sample.

Compared with methods such as neural networks[7], the XGBoost algorithm is more rigorous in mathematical theory and is supported by statistical theory knowledge. The extensive practical research in the study has shown that the XGBoost algorithm has the advantages of efficiency and accuracy.

The contents of 7 elements (Si, Al, Fe, Ca, K, Na, Mg) in the stratum were selected as input features, and the contents of 6 minerals (quartz, potassium feldspar, plagioclase, illite, chlorite, and calcite) were selected as output features. The dataset was divided into a training set and a test set. The training set consisted of 80% of the

samples randomly selected from the dataset, and the test set consisted of 20% of the samples. Using the XGBoost algorithm, the mineral content was inversely calculated based on the element content. The calculation results of the mineral content are shown in Figure 6 below.

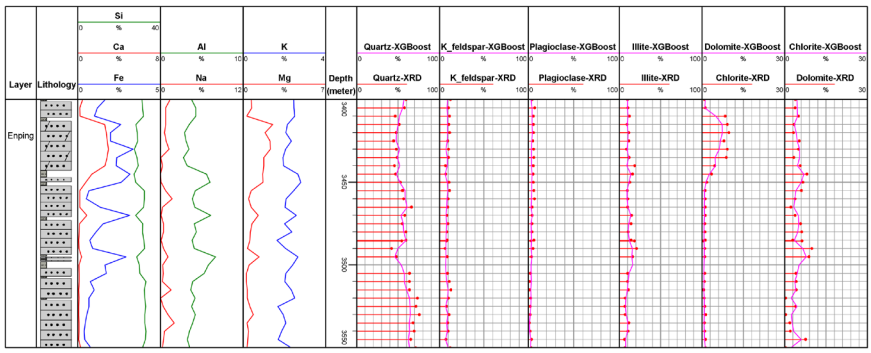

Figure 6. Results of XGBoost machine learning method calculation

4. Comparison of Results

The oxide closed model can only calculate the content of the three major types of minerals in sedimentary rocks, and there is a significant inconvenience in the comparison and inversion accuracy with other methods. Therefore, only the comparison of mathematical methods and machine learning algorithms is conducted here. The mineral contents calculated by the mathematical methods and the machine learning algorithms are compared with the actual laboratory full-rock analysis results, and the inversion result curves are shown in Figure 7. The red bar-shaped lines in the figure are the mineral contents obtained from the laboratory full-rock analysis, the blue curve is the mineral content calculated by the non-negative least squares method, the black curve is the mineral content calculated by the truncated singular value decomposition method, the green curve is the mineral content calculated by the linear programming method, and the pink curve is the mineral content obtained by the XGBoost machine learning algorithm. It can be seen that the mineral content results calculated by the four methods are basically consistent with the results obtained from the laboratory core analysis with respect to the variation with the depth of the strata. Among them, the contents of plagioclase and chlorite are slightly different, which may be due to the inaccurate model design and the low mineral content, resulting in larger errors; the inversion result of the linear programming method is biased, which may be due to the improper selection of conversion coefficients, and the subsequent model application can be continuously improved. According to the mean square error table of the four mineral inversion methods (see Table 2), it can be known that the inversion accuracy of the machine learning model is the best, reaching 0.91; the non-negative least squares method and the truncated singular value decomposition method have the second-best accuracy, which are 0.88 and 0.81 respectively; the linear programming method has the lowest accuracy, only

0.79. Overall, the XGBoost machine learning algorithm results are the best, especially for quartz and potassium feldspar, the calculation results are the most consistent with the core analysis; the least squares method and the truncated singular value decomposition method are second. The quartz content is the most consistent with the full-rock analysis results, which may be related to the relatively simple elemental composition of the minerals.

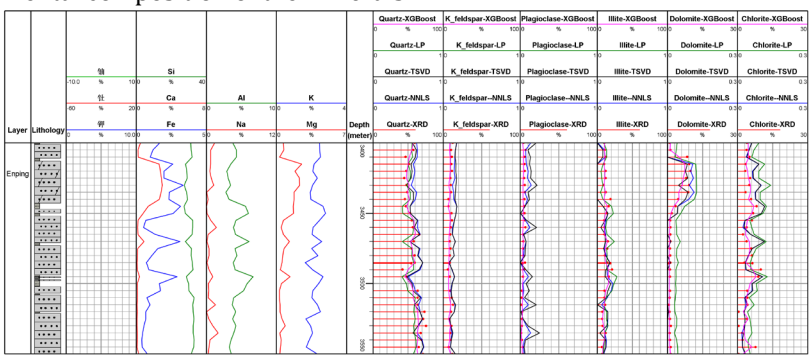

Figure 7. Comparison of results from four mineral inversion methods

Table 1. Comparison of Accuracy of Four Mineral Inversion Methods

	XGBoost	NNLS	TSVD	LP
MSE	0.91	0.88	0.81	0.79

5. Conclusion

(1) The oxide closure model method is simple to calculate and has wide applicability. It can be used in most sedimentary rock formations, but it can only roughly classify the contents of three major types of minerals. (2) The least squares method is mainly applicable when the number of equations is greater than the number of unknowns, that is, when the number of element types is greater than that of mineral types. It is used to inversely calculate the contents of 6 minerals based on the data of 7 elements in the text. At this time, the conversion equation becomes an overdetermined equation. The non-negative least squares method constrains all coefficients not to be negative, which is more in line with the actual geological conditions. (3) The truncated singular value decomposition method can solve the problem of the conversion coefficients being a singular matrix. This method requires that the number of element types in the equation be no less than that of mineral types. When solving the optimal solution, the singular values with very small values and 0 values are directly removed to reduce the mean square error of the solution and improve the accuracy and reliability of the solution. However, due to the truncation of singular values, some characteristic points are ignored, resulting in distorted calculation results. (4) The linear programming method calculates the mineral contents by setting the objective function and constraints. This method does not have specific re-

quirements for the number of elements and minerals. Theoretically, it is more suitable for solving practical problems. The key point of this method is to establish the optimal mathematical model, and a large amount of work is required for calculation, analysis and comparison. (5) The XGBoost machine learning algorithm conducts correlation analysis on the elements constituting the minerals, selects the elements with better correlation with the minerals as inputs, and establishes an algorithm model using the main minerals as outputs. The training set and test set ratios are divided to calculate the mineral contents. With a large amount of data support, better inversion results may be obtained.

References

[1] Lei J., Wang S. S., Yang Y. Application of cuttings logging digital technology in Tarim Oilfield[J]. Mud Logging Engineering, 2017, 28(4): 1-6, 129.

[2] Huang, Q. J., Sun, J. M., & Wang, H. Q. (2019). Comparative study on multiple methods for determining mineral content by elemental logging. Well Logging Technology, 43(5), 445-451.

[3] Herron, M. M. (1986). Mineralogy from geochemical well logging. Clays and Clay Minerals, 34(2), 204-213.

[4] Sun, J. M., Jiang, D., & Yin, L. (2014). A new method for determining mineral content by formation elemental logging. Natural Gas Industry, 34(2), 42-47.

[5] Zhu, B. Y. (2019). Multi-mineral inversion with least squares method under constraints in Matlab. China Petroleum and Chemical Standard and Quality, 39(6), 158-160.

[6] Guo, R. (2017). Study on mineral calculation method of elemental capture spectroscopy logging (Master's thesis). Xi'an Shiyou University, Xi'an. Retrieved from

[7] Zhang, T. (2016). Study on the relationship between elements and magmatic rocks, sedimentary rocks, and sedimentary rock minerals using artificial neural networks (Master's thesis). Northwest University, Xi'an. Retrieved from