ORIGINAL PAPER

Geochemical prospectivity mapping through factor analysis and maximum entropy model: a case study in the Mila Mountain copper deposit, southern Tibet

Youhua Wei[1] · Binbin Li[2] · Ke Guo[1]

Abstract

The complexity of geochemical patterns in surficial media necessitates considering symbiotic combinations of elements when identifying geochemical anomalies in geochemical prospecting. In the present study, the geochemical sample data were "opened" by logarithmic ratio transformation, and factor analysis was then carried out to obtain 9 factors, which represented different combinations of element. Among them, the low-value area in F2 (SiO_2-Al_2O_3-K_2O-Ba-Be-Na_2O-Cu) corresponded mainly to the copper mineralization zone in the study area. Then, the maximum entropy model for predicting the potential distribution of copper deposits was established with the factor score as a comprehensive index based on the maximum entropy principle. The spatial association of individual ore-controlling variables with the occurrences of copper deposits was investigated by response curves, and the relative importance of each ore-controlling variable was determined by jackknife analysis in the MaxEnt model. The effectiveness of the proposed method was verified by analyzing stream sediment geochemical samples from the Mila Mountain region, Tibet. The results indicated that the F2 score was the most important ore-controlling variable. The performance of the model was evaluated by AUC, Kappa, and TSS. The AUC, maximum Kappa, and maximum TSS values were 0.863, 0.606, and 0.657, respectively. The results show that the model can effectively combine multisource geospatial data with copper mineralization, according to the geological background and favorable mineralization conditions, initially identify several prospecting targets, and provide a scientific basis for subsequent copper exploration in the study area.

Keywords Compositional data · Multivariate geochemical analysis · Maximum entropy · Data fusion · Mila Mountain region

✉ Binbin Li
Libinbin@cwnu.edu.cn

Youhua Wei
weiyouhua@cdut.edu.cn

Ke Guo
guoke@cdut.edu.cn

[1] Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, People's Republic of China

[2] School of Mathematics and Information, China West Normal University, No.1, Shida Road ShunQing Distrct, Nanchong, Sichuan 637002, People's Republic of China

Introduction

In recent years, researchers have identified potential minerals mainly by extraction of weak anomalies in geological, geophysical, and geochemical information (Cheng 2012a, b, 2021). In processing regional geochemical data, complex geological and geographical background conditions have posed difficulties for screening geochemical anomalies: the areas with high element contents may not have minerals, and areas with low element contents, by contrast, may be potential mineral areas. Thus, the detection of weak geochemical anomalies is one of the main challenges in geochemical exploration (Xiong et al. 2018; Zuo et al. 2019; Wang and Zuo 2020).

The processes of element migration and enrichment do not occur for a single element, but follow the affinity characteristics of elements and the matching variations among elements, which lead to the numerical correlation of elements and

clustering in space (Cong et al. 2012). Previous studies have shown that the distribution patterns of geochemical anomalies are usually synthetic multi-element anomalies, mainly because some elements tend to show similar geochemical behaviors under specific geological conditions; thus, certain specific symbiotic combinations of elements appear in the final geological products (Wang 2018; Gao 2019). Therefore, the results of multiple geochemical synthesis anomalies are more reliable than those of single-element anomalies. Common information fusion models include logical regression (Mejía-Herrera et al. 2015; Yousefi et al. 2014), random forest (Zhang et al. 2019a, b; Sun et al. 2019; Hong et al. 2021), and deep learning (Moeini and Torab 2017; Zhang et al. 2021), all of which can identify geochemical anomalies against complex geological backgrounds and effectively integrate multivariate information. However, selecting appropriate models plays a crucial role in the regression analysis (Ramezanali et al. 2019). Due to the complexity of the algorithm, the efficiency of data processing is low in random forest and deep learning models (Chen and Wu 2017; Rahimi et al. 2021; Zuo et al. 2021; Zuo et al. 2019). Maximum entropy model is simple to calculate, can quickly and efficiently process data, and has high prediction accuracy, which is suitable for solving classification problems (Phillips et al. 2006; Phillips and Jane 2013). This model has been successfully applied in various fields, such as natural language processing (Dong et al. 2012), economic prediction (Xu et al. 2014), environmental evaluation (Biazar et al. 2020; Biazar and Ferdosi 2020; Aghelpour et al. 2020; Jahangir et al. 2021; Yang et al. 2021; Yang et al. 2020; Azareh et al. 2019), geographical distribution of animal and plant species (Wang et al. 2017), and mineral exploration (Liu et al. 2018; Li et al. 2019, 2021). Liu et al. (2018) applied the maximum entropy model to the potential distribution of orogenic gold deposits based on quantitative critical metallogenic processes in the Tangbale-Hatu belt, western Junggar, China. Li et al. (2019) used the maximum entropy model to predict the metallogenic prospect of the Mila Mountain region in Tibet, and the model considered both positive and negative factors related to mineralization.

In the present study, the copper polymetallic mineralization in the Mila Mountain region, southern Tibet is taken as the research object. The process and results of metallogenic prediction are discussed by using component data factor analysis and the maximum entropy model, to provide a reference for the wide application of machine learning algorithm in mineral resource evaluation and prediction.

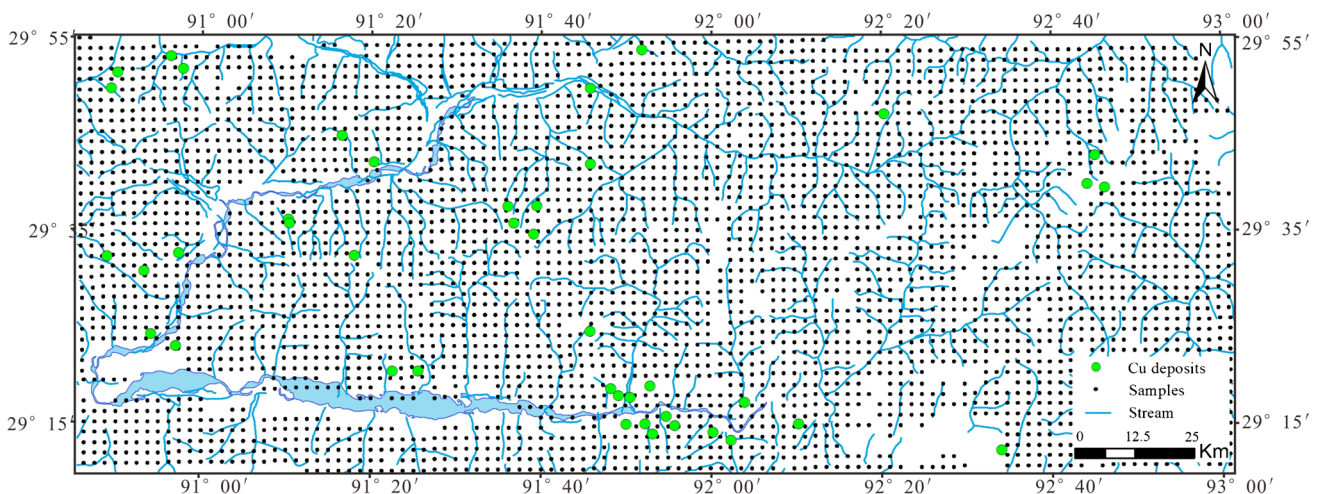## Study area and geochemical data

Study area

The study area is located in the middle and southern parts of the Gangdese-Nyainqentanglha (terrane) plate in the Tethys structural area of Tibet (29°10′N–29°55′N; 90°45′E–93°00′E), and is one of the most famous copper metallogenic regions in the world (Song et al. 2018; Lin et al. 2017a, b, 2019). The study area is a deeply incised alpine region, with the terrain high in the north and low in the south, high in the east and low in the west. Most of the mining areas are higher than 4500 m above sea level, and some mountaintops are covered with snow year-round, often forming modern glaciers. The climate in the region is a typical plateau continental climate, with a distinct dry season and a rainy season, and rainy season from June to September. The vertical zonation of temperature and vegetation is obvious, with long sunny days, low temperatures, large temperature differences between day and night, short frost-free periods and many snowfall days. The weather is relatively fair from April to September every year, and is suitable for field operations. The streams in the study area are well developed, including mainly the Yarlung-Tsangpo River, the Lhasa River and secondary streams.

Magmatic rocks in the research area widespread, including large deep intrusions and thick volcaniclastic rock layers, mainly distributed north of the Yajiang fault. These igneous rocks are important constituents of the Gangdese volcanic-magmatic arc (Lang et al. 2012). The intrusive rocks from a major component of the magmatic rocks in the Gangdese area, and represent the products of plate subduction-collision events during the evolution of middle-Cenozoic Tethyan tectonics. The general trend of the fault structure in the study area is nearly east-west, which plays an important role in mineralization. The Bouguer gravity anomaly is 350–550 mgL, and has a gradient gradually decreasing from south to north. The anomaly value (ΔT) of the aeromagnetic pole is -300–550 nT. Most of the single positive magnetic anomaly strips are oriented nearly east-west, and the positive and negative magnetic anomalies alternate to form strips (Zhang et al. 2019a, b). The anomalies in geochemical elements such as Cu, Mo, Pb, Zn, Au, and Ag are generally distributed nearly east-west, and the characteristics of element association are regular from south to north (Wang et al. 2010). The study area is very rich in minerals, including ferrous metal minerals, nonferrous metal minerals, precious metal minerals, fuel minerals, building materials, non-metallic minerals, and geothermal resources (Zheng et al. 2016; Yang and Hou 2009). Among them, nonferrous metal (copper, lead-zinc, etc.) building materials and geothermal resources are the dominant assets in the area (Figure 1).
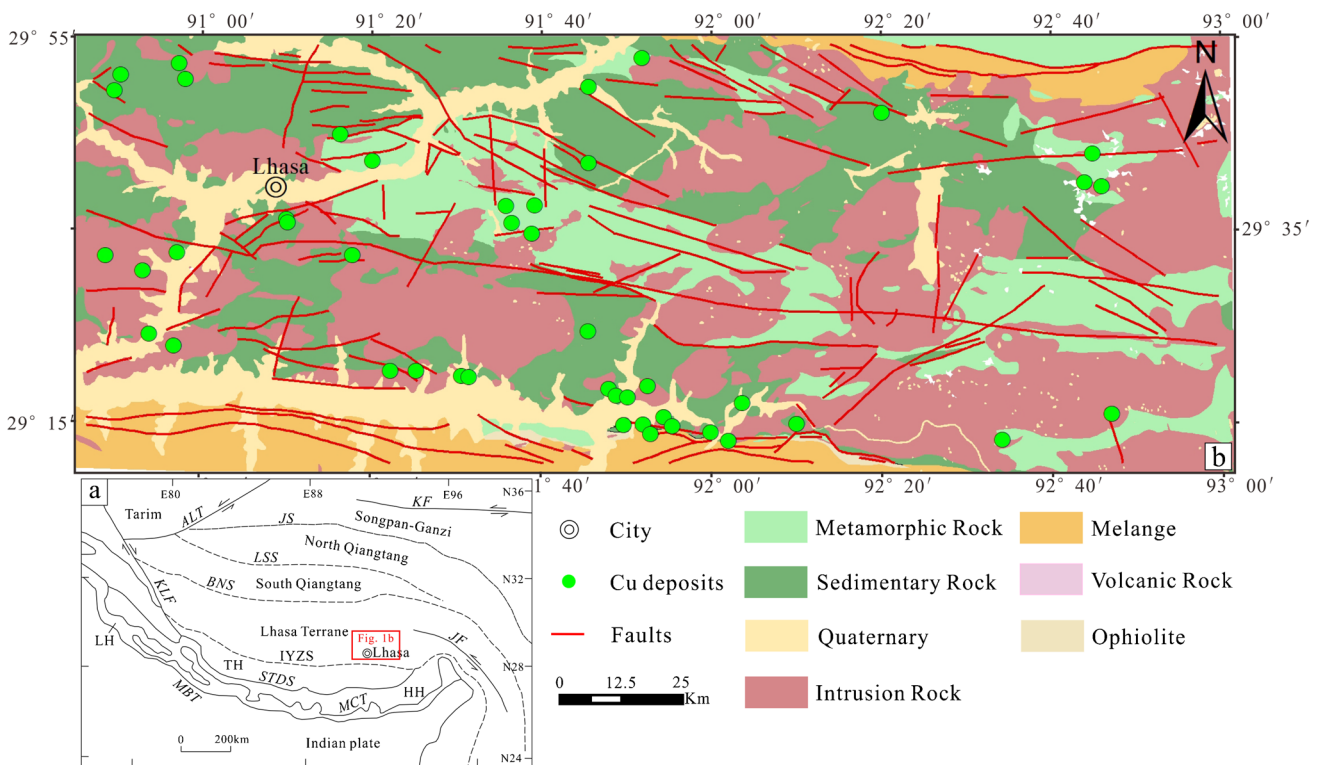
Fig. 1 **a** Tectonic sketch map showing the location of the study area (after Yin and Harrison 2000; Zheng et al. 2021). JS, Jinshajiang suture; LSS, Longmucuo–Shuanghu suture; BNS, Bangonghu–Nujiang suture; IYZS, Indus–Yarlung Zangbo suture; STDS, south Tibetan detachment system; MCT, main central thrust; MBT, main boundary thrust; ALT, Altyn Tagh fault; KF, Kunlun fault; KLF, Karakoram fault; JF, Jiali fault. TH, Tethys Himalaya; HH, Higher Himalaya; LH, Lesser Himalaya. **b** Generalized geological and deposits distribution map of the study area (Li et al. 2021)

Fig. 2 Geochemical sampling locations

Geochemical data

The geochemical dataset utilized in this study was collected as part of the Chinese National Geochemical Mapping project. The actual sampling area of the stream sediment survey was 12,290 km^2, and 4141 effective samples were collected (Figure 2) (Xie et al. 1997). The sampling locations were arranged according to specifications, and located at the bottom of modern rivers and riverbeds, or at the bottom of the seasonal flow or in the main

channel conducive to gravel deposition and mixed accumulation of various particle sizes. The sampling material was mainly sand, and one sample was combined after multipoint collection within 20–50 m near the sampling point. Samples with particle sizes less than 0.22 mm (60 mesh) were collected using 60-mesh stainless steel sieves and the sample weights were greater than 200 g. All samples were analyzed after drying and sifting. 39 elements such as Cu, Pb, Zn were quantitatively analyzed by the foam adsorption graphite furnace atomic absorption method, AC arc-emission spectroscopy, inductively coupled plasma atomic emission spectroscopy, inductance coupled plasma mass spectrometry, the alkali melting-catalytic electrode method, atomic fluorescence spectroscopy, etc. (Xie et al. 1997, 2008; Wang et al. 2011; Xie 2008). The specific test methods, detection limits and other information for each element are shown in Table 1, and detailed information about the data quality and sampling strategy can be found in these studies (Xie et al. 1997).

Methods

Log-ratio approach

The commonly used methods of log-ratio transformation include additive log-ratio (alr), central log-ratio (clr), and isometric log-ratio (ilr) (Aitchison 1986; Carranza 2011; Egozcue et al. 2003). The output result of the alr transformation depends mainly on the choice of denominator, so it has strong subjectivity. The clr transformation takes the geometric mean of all variables as the denominator, which can effectively improve upon the alr transformation, and the vectors before and after the transformation correspond one to one; thus, the clr transformation can be used to explain the statistical results based on the clr transformation data using the original variable. However, the clr transformation has the problem of data collinearity, which makes it impossible to use the ordinary least squares regression method (Carranza 2011; Filzmoser et al. 2009b; Zuo et al. 2013). The ilr transformtion solves the problem of data collinearity in clr transformation, and retains all the advantages. The ilr approach transforms the component data in simplex space into real numbers in Euclidean space (Egozcue et al. 2003). The clr transformation is applied to process the regional geochemical data. It involves a transformation from the simplex sample space to D-dimensional real space. A compositional random vector $X = (x_1, x_2, \cdots, x_D)$ can be defined as follows (Buccianti and Grunsky 2014; Egozcue et al. 2003):

Table 1 Specific test methods and detection limits for each element (Wang et al. 2011)

Element	Unit	Detection limit	Analytical method
Ag	µg/g	0.02	AAS
As	µg/g	1	AFS
Au	µg/g	0.0003	AAS
B	µg/g	5	AES
Ba	µg/g	50	XRF
Be	µg/g	0.5	AES
Bi	µg/g	0.1	AFS
Cd	µg/g	0.05	AAS
Co	µg/g	1	XRF
Cr	µg/g	15	XRF
Cu	µg/g	1	XRF
F	µg/g	100	ISE
Hg	µg/g	0.0005	AFS
La	µg/g	30	XRF
Li	µg/g	5	AAS
Mn	µg/g	30	XRF
Mo	µg/g	0.4	POL
Nb	µg/g	5	XRF
Ni	µg/g	2	XRF
P	µg/g	100	XRF
Pb	µg/g	2	XRF
Sb	µg/g	0.1	AFS
Sn	µg/g	1	AES
Sr	µg/g	5	XRF
Th	µg/g	4	XRF
Ti	µg/g	100	XRF
U	µg/g	0.5	LIF
V	µg/g	20	XRF
W	µg/g	0.5	POL
Y	µg/g	5	XRF
Zn	µg/g	10	XRF
Zr	µg/g	10	XRF
Al_2O_3	%	0.05	XRF
CaO	%	0.05	XRF
Fe_2O_3	%	0.05	XRF
K_2O	%	0.05	XRF
MgO	%	0.05	XRF
Na_2O	%	0.05	XRF
SiO_2	%	0.1	XRF

XRF, X-ray fluorescence; *AAS*, Atomic absorption spectroscopy; *AES*, Atomic emission spectroscopy; *AFS*, Atomic fluorescence spectrometry; *ISE*, Ion specific electrode; *POL*, Polarography; *LIF*, laser induced fluorescence

$$clr(X) = (y_1, y_2, \cdots, y_D) = \left(log \frac{x_1}{\sqrt[D]{\prod_{i=1}^{D} x_i}}, \cdots, log \frac{x_D}{\sqrt[D]{\prod_{i=1}^{D} x_i}} \right) \quad (1)$$

Where, $\sqrt[D]{\prod_{i=1}^{D} x_i}$ is the geometric mean.

Factor analysis

Factor analysis is one of the most commonly used multivariate statistical methods for dimensionality reduction of datasets. It is an effective method to realize high-dimensional data visualization in low-dimensional space based on variance and covariance matrices, and has been widely used in geochemical data processing (Meigoony et al. 2014; Hoseinpoor and Aryafar 2016; Wu et al. 2020; Yousefi et al. 2014; Zuo et al. 2013; Filzmoser et al. 2009a, b, c). Factor analysis can combine multiple related variables into a single variable, thereby reducing the dimension of a dataset to irrelevant principal components based on the covariance or correlation of the variables (Jolliffe 2002; Reimann et al. 2005; Zuo 2011). In the process of geochemical data processing, each element dataset is divided into several nonobservable factors through factor analysis. These factors extract the main information of the original variables and represent some inherent features of the original dataset, such as complex geologic origin and mineralization process (Johnson and DW 2002; Muller et al. 2008; Zuo 2011).

Maximum entropy model

Maximum entropy theory reflects a basic principle of nature: systems contain both constraints and freedom and always tend toward the maximum degree of freedom under the constraints, that is, maximum entropy (Phillips et al. 2006). Therefore, under known conditions, the system with the largest entropy is most likely to be close to its real state. Specifically, for an event, we often know only part of its situation and know nothing about other aspects of its situation. When building a model for the event, we should attempt to fit the known part to make the model conform to the known situation; for the unknown parts, the uniform distribution is maintained, and the entropy is the largest of the event at this time (Li 2012).

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of discrete random variables, $p_i = P(X = x_i)$ be the probability distribution of $x_i (i = 1, 2, \cdots, n)$, and the entropy can be explicity written as:

$$H(X) = -\sum_{i=1}^{n} p_i log p_i \tag{2}$$

And the entropy satisfies the following condition

$$0 \le H(X) \le log n \tag{3}$$

Where n is the total number of values available in X, and the right equality holds when X is uniformly distributed.

The maximum entropy principle originates from statistical mechanics (Jaynes 1957), which obtains the maximum entropy model (MaxEnt) when applied to classification problems. We assume that the classification model is a conditional probability distribution $P(Y|X)$, $X \in \chi \in R^n$ is the input data, $Y \in y \in R^n$ is the output data, x and y represent sets of input data and sets of output data, respectively. The model represents the output of Y with a conditional probability P for a given input data X. The maximum entropy model can be defined as follows (Li 2012).

Let $C \equiv \left\{ P \in \mathrm{P} \middle| E_P(f_i) = E_{\tilde{P}}(f_i), (i = 1, 2, \cdots, n) \right\}$ be the set of models that satisfy all constraints, $H(P) = -\sum_{x,y} \tilde{P}(x)P(y|x)lnP(y|x)$ be the conditional entropy defined on the conditional probability distribution $P(Y|X)$. Then, the model with the largest conditional entropy $H(P)$ in the model set C is called the maximum entropy model.

Where $E_{\tilde{P}}(f_i) = \sum_{x,y} \tilde{P}(x, y)f_i(x, y)$ is the expected value of n eigenfunctions $f_i(x, y)$ about the empirical distribution $\tilde{P}(x, y)$, and $E_P(f_i) = \sum_{x,y} \tilde{P}(x)P(y|x)f_i(x, y)$ is the expected value of n eigenfunctions $f_i(x, y)$ about $P(Y|X)$ and the empirical distribution $\tilde{P}(X)$.

The process of solving the maximum entropy model is a learning process for the model, which can be formalized into a constrained optimization problem (Li 2012).

Let $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ be the training dataset and $f_i(x, y), i = 1, 2, \cdots, n$ be the eigenfunction. The learning of the maximum entropy model is equivalent to the constrained optimization problem:

$$\max_{P \in C} H(P) = -\sum_{x,y} \tilde{P}(x)P(y|x)lnP(y|x) \tag{4}$$

$$\text{s.t. } E_P(f_i) = E_{\tilde{P}}(f_i) i = 1, 2, \cdots, n \tag{5}$$

$$\sum_y P(y|x) = 1 \tag{6}$$

According to the optimization problem, the maximum problem is rewritten to the equivalent minimum problem:

$$\min_{P \in C} -H(P) = \sum_{x,y} \tilde{P}(x)P(y|x)lnP(y|x) \tag{7}$$

$$\text{s.t. } E_P(f_i) - E_{\tilde{P}}(f_i) = 0 i = 1, 2, \cdots, n \tag{8}$$

$$\sum_y P(y|x) = 1 \tag{9}$$

The solution derived from the above constrained optimization problem is the solution of maximum entropy model learning. However, the expectation of the empirical distribution is usually not equal to the real situation,

but only approximates the real situation. If the solution is strictly based on the above constraint conditions, overfitting of the training data can easily be caused in the learning process. Therefore, the constraint conditions can be appropriately relaxed in the actual solution, and equation $E_P(f_i) - E_{\tilde{P}}(f_i) = 0 i = 1, 2, \cdots, n$ can be replaced by $E_P(f_i) - E_{\tilde{P}}(f_i) \leq \beta_i$, β_i is a constant, called regularization multiplier (Dudík et al. 2004). The problem of overfitting can be effectively avoided by solving the above constrained optimization problems (Phillips et al. 2004, 2006).

Model evaluation

Model evaluation is an important component of machine learning. In this paper, the MaxEnt model is evaluated by the area under the receiver operating characteristic curve (AUC), Cohen's maximum kappa coefficient and true skill statistics (TSS). These three metrics are calculated based on the specificity and sensitivity of the prediction model (Li et al. 2019).

Sensitivity (*SEN*) refers to the percentage of positive samples correctly identified by the classifier, and is used to measure the ability of the classifier to identify positive samples.

$$SEN = \frac{true positives}{\text{true positives} + \text{flase negatives}} \quad (10)$$

Specificity (*SPE*) is the percentage of negative samples correctly identified by the classifier, and is used to measure the ability of the classifier to identify negative samples.

$$SPE = \frac{true negatives}{\text{true negatives} + \text{false positives}} \quad (11)$$

The receiver operating characteristic curve (ROC) is plotted with the false positive rate (*FPR*) as the horizontal axis and the true positive rate (*TPR*) as the vertical axis.

$$TPR = \frac{true positives}{\text{true positives} + \text{flase negatives}} \quad (12)$$

$$FPR = \frac{false positives}{\text{true negatives} + \text{false positives}} \quad (13)$$

The AUC is a quantitative evaluation index independent of the threshold, and the greater value indicates a better model classification effect. If the ROC is connected by points with coordinate $\{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$, AUC can be expressed as:

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \quad (14)$$

The kappa coefficient refers to the accuracy of prediction relative to random occurrence, and is influenced by the incidence of distribution points and the threshold (Cohen 1960), it can be expressed as:

$$Kappa = \frac{(TP + TN) - (expected correct)_{ran}}{\text{Total} - (expected\ correct)_{ran}} \quad (15)$$

W   h   e   r   e   ,
$(expected correct)_{ran} = \frac{1}{N}[(TP + FN)(TP + FP) + (TN + FN)(TN + FP)]$, TP, TN, FP and FN are true positive, true negative, false positive, and false negative values, respectively.

TSS represents the ability of predictive results to distinguish between "yes" and "no", independent of the incidence of distribution points, but influenced by thresholds, while $TSS = SEN + SPE -1$ (Allouche et al. 2006). AUC, TSS, and kappa statistics have different responses to the incidence and threshold of distribution points (Table 2), and can be combined to better evaluate the performance of the model (Swets 1988; Araújo et al. 2005; Coetzee et al. 2009).

Results and discussion

Factor analysis for compositional data

In nature, almost all environmental data have the characteristics of component data. Due to the closure effect of component data, traditional multivariate analysis has certain limitations in processing this kind of data (Carranza 2011; Filzmoser et al. 2009a; Zuo et al. 2013; Zuo 2014). Geochemical data are typically compositional data, where the sum of all elements is a fixed value (e.g. 100%) (Aitchison 1986). Aitchison (1986) proposed that the study of component data should focus on the proportional relationships between components rather than the component itself (Aitchison 1986; Filzmoser et al. 2009b; Zuo et al. 2013), thus proposing two classic log ratio transformation methods for "open" component data, namely, the additive log ratio (alr) transformation and central log ratio (clr) transformation, making traditional statistical methods also applicable to the analysis of transformed data (Aitchison 1986; Aitchison et al. 2000; Carranza 2011; Filzmoser and Hron 2009; Chen et al. 2016; Chen et al. 2018).

Table 2 Measurement standards for AUC, kappa and TSS

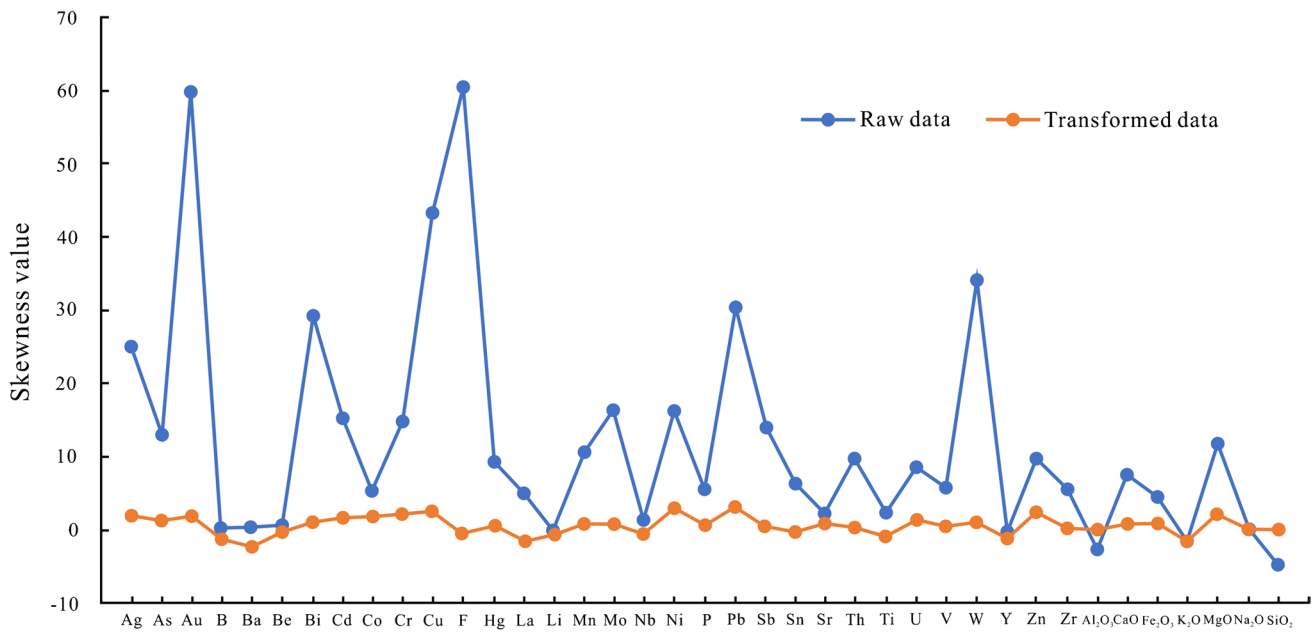| Evaluation index	Fail	Poor	Medium	Good	Very good
AUC		0.5–0.6	0.6–0.7	0.7–0.8	0.8–0.9	0.9–1.0
Kappa	0–0.4	-	-	0.4–0.75	0.75–1.0
TSS	0–0.5	-	-	0.5–0.8	0.8–1.0

Fig. 3 Skewness comparison diagram of data distribution

The clr transformation was carried out on the data for 39 geochemical elements, and the skewness value of the data distribution was then calculated after clr transformation, and compared with the skewness value of the raw data distribution (Figure 3). The skewness value of the data distribution of each element after clr transformation was significantly reduced, indicating that the transformed data distribution was closer to the normal distribution.

Factor analysis was carried out on clr transformed data, and the appropriate factor was determined by eigenvalues greater than 1 and a cumulative variance contribution rate greater than 70%. The element combination reflected by the orthogonal rotation factor load matrix was more reasonable and interpretable than other combinations. Therefore, the maximum variance method was used to classify the element combinations by an orthogonal rotation factor loading matrix. According to Table 3, the 39 variables were attributed to 9 factors, these 9 factors accounted for 73.035% of the total variance of the raw 39 variables. The variance contribution reflects the ability of each factor to explain the total variance of the original variable. A higher value indicates a more important corresponding factor. Then, the factor load after rotation is sorted, and the results obtained are shown in Table 4.

The combination of elements represented by each factor indicates different geochemical significance. The score distribution map of each factor shows the geological significance represented by each factor (Figure 4). F1 (Zr-La-Th-Nb-Y-Zn-Ti-Cd-Ag) is the element combination of the felsic granite distribution area. The low

Table 3 Characteristic values and cumulative variance contribution rates

Component	Initial eigenvalues			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.991	17.927	17.927	5.197	13.324	13.324
2	6.263	16.058	33.985	5.184	13.293	26.618
3	5.076	13.016	47.001	3.621	9.284	35.901
4	3.027	7.760	54.762	3.504	8.984	44.885
5	1.904	4.882	59.644	3.298	8.457	53.342
6	1.636	4.196	63.840	2.442	6.260	59.603
7	1.439	3.689	67.529	2.069	5.305	64.908
8	1.124	2.881	70.410	1.857	4.761	69.668
9	1.024	2.625	73.035	1.313	3.366	73.035

Extraction method: principal component analysis

**Table 4** Orthometric rotating factor loading matrix

Element	F1	F2	F3	F4	F5	F6	F7	F8	F9
Zr	0.864	−0.045	0.094	0.124	0.032	0.044	0.088	0.110	−0.009
La	0.771	0.202	−0.078	0.117	−0.115	−0.079	0.158	0.023	0.064
Th	0.747	−0.046	0.003	0.033	−0.124	−0.135	−0.022	−0.249	−0.033
Nb	0.682	0.159	0.334	0.070	0.044	0.015	−0.074	0.111	0.171
Y	0.673	0.174	0.238	0.248	0.066	0.153	−0.140	0.282	0.216
Zn	−0.557	−0.104	0.134	0.225	0.259	0.174	−0.145	−0.443	0.285
Ti	0.551	−0.218	0.130	0.098	0.051	0.470	0.144	0.353	0.227
Cd	−0.515	−0.202	−0.026	0.367	0.093	0.001	−0.325	−0.369	0.176
Ag	−0.425	−0.206	0.117	0.337	−0.193	0.079	−0.277	−0.336	−0.146
SiO_2	0.100	0.874	−0.206	−0.029	0.002	0.071	−0.142	0.093	0.049
Al_2O_3	−0.117	0.861	−0.125	0.119	0.227	0.118	0.168	0.079	0.021
K_2O	0.130	0.845	0.101	0.298	−0.193	−0.061	−0.025	−0.006	0.025
Ba	0.029	0.802	0.034	0.207	−0.051	0.153	0.160	−0.025	−0.009
Be	0.236	0.678	0.059	−0.115	−0.150	−0.168	0.057	−0.061	0.077
Na_2O	0.061	0.650	−0.631	0.157	0.050	−0.015	0.109	0.109	0.051
Cu	−0.354	−0.444	−0.076	−0.015	0.043	0.246	−0.013	−0.113	0.356
CaO	−0.030	−0.080	−0.817	−0.182	0.007	0.017	0.188	0.158	0.107
Sr	−0.076	0.392	−0.787	0.019	0.003	0.071	0.348	0.026	0.029
B	0.268	0.015	0.667	−0.117	−0.347	0.228	−0.042	0.150	0.011
Sn	0.402	−0.028	0.568	−0.069	−0.132	−0.007	0.047	0.060	0.137
Li	−0.192	0.484	0.540	−0.096	−0.172	0.083	0.259	0.141	0.081
Ni	−0.161	−0.155	0.140	−0.869	0.097	0.230	−0.110	0.041	0.040
Cr	0.033	−0.259	0.046	−0.854	0.131	0.208	−0.113	0.097	0.032
MgO	−0.276	−0.047	−0.381	−0.672	0.401	0.109	0.087	0.085	0.010
Co	−0.331	−0.077	0.010	−0.605	0.574	0.211	−0.003	0.068	0.043
Fe_2O_3	0.078	−0.225	−0.092	−0.346	0.747	0.192	0.240	0.063	-0.086
Mn	−0.365	−0.025	−0.063	0.009	0.724	−0.018	−0.110	−0.045	0.089
V	0.192	−0.167	−0.143	−0.239	0.703	0.267	0.267	0.108	−0.050
Sb	−0.003	−0.363	0.244	0.117	−0.519	0.276	−0.245	−0.027	−0.267
As	−0.301	−0.302	0.341	0.236	−0.356	−0.004	−0.102	0.010	−0.249
W	0.077	−0.213	0.039	0.146	−0.158	−0.751	−0.029	0.007	0.091
Mo	−0.440	−0.032	−0.001	0.351	0.113	−0.643	0.055	0.088	0.026
U	0.241	0.176	−0.197	0.210	−0.043	−0.583	0.091	−0.023	−0.078
Bi	0.029	−0.231	0.250	0.249	−0.282	−0.468	−0.142	−0.280	0.267
P	0.016	−0.013	−0.179	0.103	0.368	−0.029	0.752	0.066	−0.061
F	0.169	0.185	−0.119	0.078	−0.012	0.025	0.694	0.094	0.195
Pb	−0.152	−0.108	0.069	0.160	−0.168	0.015	−0.168	−0.818	0.036
Hg	−0.003	−0.114	0.188	0.084	−0.297	0.229	−0.327	0.439	0.020
Au	−0.163	−0.122	0.010	0.064	−0.036	0.080	−0.106	0.020	−0.765

Extraction method: principal component analysis.

value area in F2 (SiO_2-Al_2O_3-K_2O-Ba-Be-Na_2O-Cu) is the main copper mineralization zone in the study area, and there are many ore deposits in the low value area. F3 (CaO-Sr-B-Sn-Li) is the element combination associated with the alkaline magmatic rocks. F4 (Ni-Cr-MgO-Co) mainly reflects the composition characteristics of ultrabasic rocks and ophiolite, which is the Cr-Ni mineralization associated with ultramafic rocks. The

low value regions in F5 (Fe_2O_3-Mn-V-Sb-As) mainly reflect the mineralized zone associated with mineralized fracture structures. F6 (W-Mo-U-Bi) reflects the high-temperature mineralization enrichment zone of ore-forming elements, granite interior and exterior contact zones and zones where intermediate to felsic veins developed. The elements in F7 (P-F) are not indicative of the mineralization process. F8 (Pb-Hg) is closely
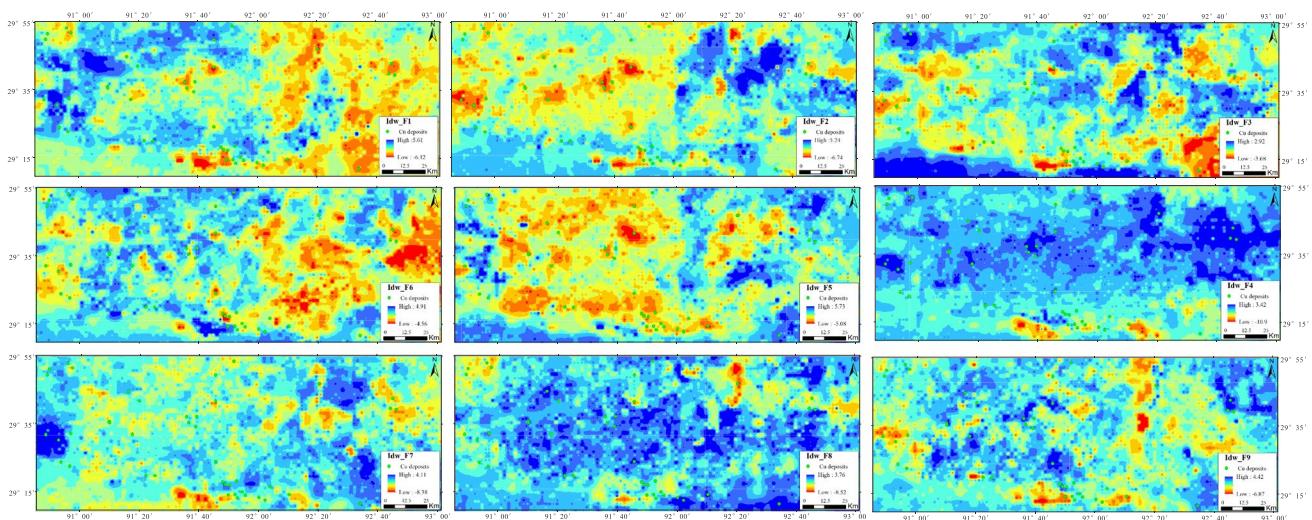
Fig. 4 The score distribution map of each factor

related to the widespread granite in the area, and F9 (Au) is the stable mineral display of placer gold under the condition of a supergene deposit.

MaxEnt model parameter analysis

The maximum entropy (MaxEnt) model is a probability distribution prediction model (Ratnaparkhi 2016). The prediction results depend on the sample point data and the environmental variable data affecting the sample point data. The probability of distribution of the sample data in other parts of the study area is determined according to the weight of the environmental variable data, and the point data with quantitative values are converted into data with probability values. In metallogenic prediction, Max-Ent obtains the prediction model by calculating the non-linear relationship between the geographical coordinates of known deposits and ore-controlling variables in the study area, and then uses this model to simulate the possible distribution of target minerals in the target area. In this study, MaxEnt v3.3.3 software was used to establish metallogenic prediction model.

When using MaxEnt software for modeling, the complexity of the model has a significant impact on the prediction effect of the model. An excessive number of variables will increase the complexity of the model. In addition, when the ore-controlling factors in the model have high collinearity, the model may be overfit. Therefore, the MaxEnt model was constructed with the factor scores in the factor analysis results. Furthermore, when the model parameters are improperly set, overfitting or redundancy of the model may also appear (Kong et al. 2019). Relevant studies show that the over-fitting of the model is mainly controlled by the setting of the

regularization multiplier, also known as β (Elith et al. 2011), and the comprehensive performance of the model with a β value between 2 and 4 is the best (Radosavljevic and Anderson 2014; Kong et al. 2019). Therefore, different β values (from 1 to 5 with a step size of 0.5) are tested to find the optimal β value for a particular model (Li et al. 2019; Wang et al. 2017). In the present study, we use the AUC value to evaluate the model, calculate the AUC values corresponding to different β values, and draw ROC curves (Figure 5). Therefore, the model is optimal when $\beta = 2$.

MaxEnt model for geochemical anomaly optimization

The MaxEnt model is constructed with the 9 groups of factor scores as the input parameters. Through a literature review combined with mineralization prediction theory, MaxEnt software randomly selects 75% of the known deposits to establish the training model, and verifies the model accuracy with the remaining 25% of the known deposits. The number of replicates is set to 20 to reduce uncertainties caused by outliers (Li et al. 2019). At the same time, the contribution of each ore-controlling variable to the mine distribution is determined by the Jackknife test brought by MaxEnt version 3. 3.3 software. The MaxEnt parameters are set as follows: the output format is "Logistic", the output file type is "ASC", the features are set to "Auto features" after removing "Threshold features", the regularization multiplier is set to "2", the replicated run type is "Bootstrap" (Wang et al. 2017; Li et al. 2021), the maximum number of iterations is set to "5000" (Phillips 2005), and the threshold rule is applied to select "10 percent training presence" (Kramer-Schadt

Fig. 5 The ROC curve corresponding to the different β values
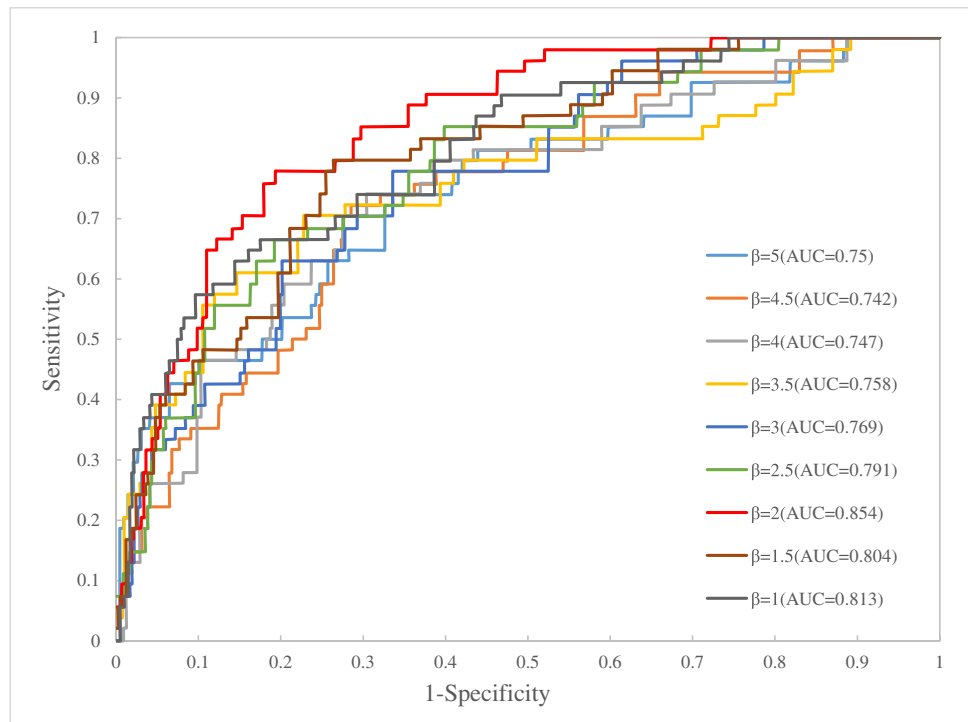
et al. 2013). Finally, response curves are created, and the jackknife is drawn to measure variable importance. The resulting data output by MaxEnt software is in ASCII format, and the ArcToolbox toolkit of ArcGIS is used to convert it to a grid, so that the results can be displayed in ArcGIS, and its value is between 0 and 1 (Figure 6).

The AUC value, kappa coefficient and TSS values of the final prediction results are calculated, the AUC = 0.863, the maximum Kappa = 0.606, and the maximum TSS = 0.657. All the evaluation indexes of the model show that the model has a good ability to identify the favorable/unfavorable areas of copper mineralization. Therefore, we infer that the model is reliable and acceptable, and more accurate than a random model. Figure 6 shows that most of the copper deposits are spatially consistent with the high anomalous probability zone marked in red. The figure illustrates that the model successfully connects the probabilities of multivariate geochemical anomalies with the known copper mineralization.

Table 5 shows the contribution rate of each factor score to the model, and Figure 7 shows the response curve of each factor score in the modeling. Among them, the F2 factor score is the most important ore-controlling variable to explain the occurrence of the copper deposits, reflecting the contribution to the deposit from the perspective of element composition and accounting for 29.3% of the total contribution rate. From the F2 response curve, as the factor score increases, the success rate of predicting

copper gradually decreases, which is consistent with the conclusion that copper has a negative load in F2. F5 (25.8%) and F4 (13.5%) are the second and third most important ore controlling factors, respectively.

To further eliminate the interaction between ore-controlling variables, the Jackknife analysis built in the MaxEnt model is used to test the importance of the impact of ore-controlling variables on the metallogenic process (Figure 8). The longer the blue bar, the more important the variable is to the distribution of the deposits. F2, F5, and F9 are important variables affecting the distribution of the deposit. The shorter the green bar, the more information the variable has than other variables have, and the variable has a great influence on the distribution of the deposit. F1, F2, and F5 have more unique information for the prediction of favorable metallogenic areas, and they are indispensable. The predictions of the MaxEnt model are consistent with the factor analysis results.

Regional Cu resource potential

The magmatic activity in the research area was frequent and large-scale, and the collision between the Asian and Indian plates formed a series of complex folds, thrust faults, and transpressional faults. The copper deposits in the study area are generally controlled by deep and large fault structures, nappe-slip structures and strike-slip structures, which are easily mineralized (Hou et al. 2006; Liu et al. 2020). Deep
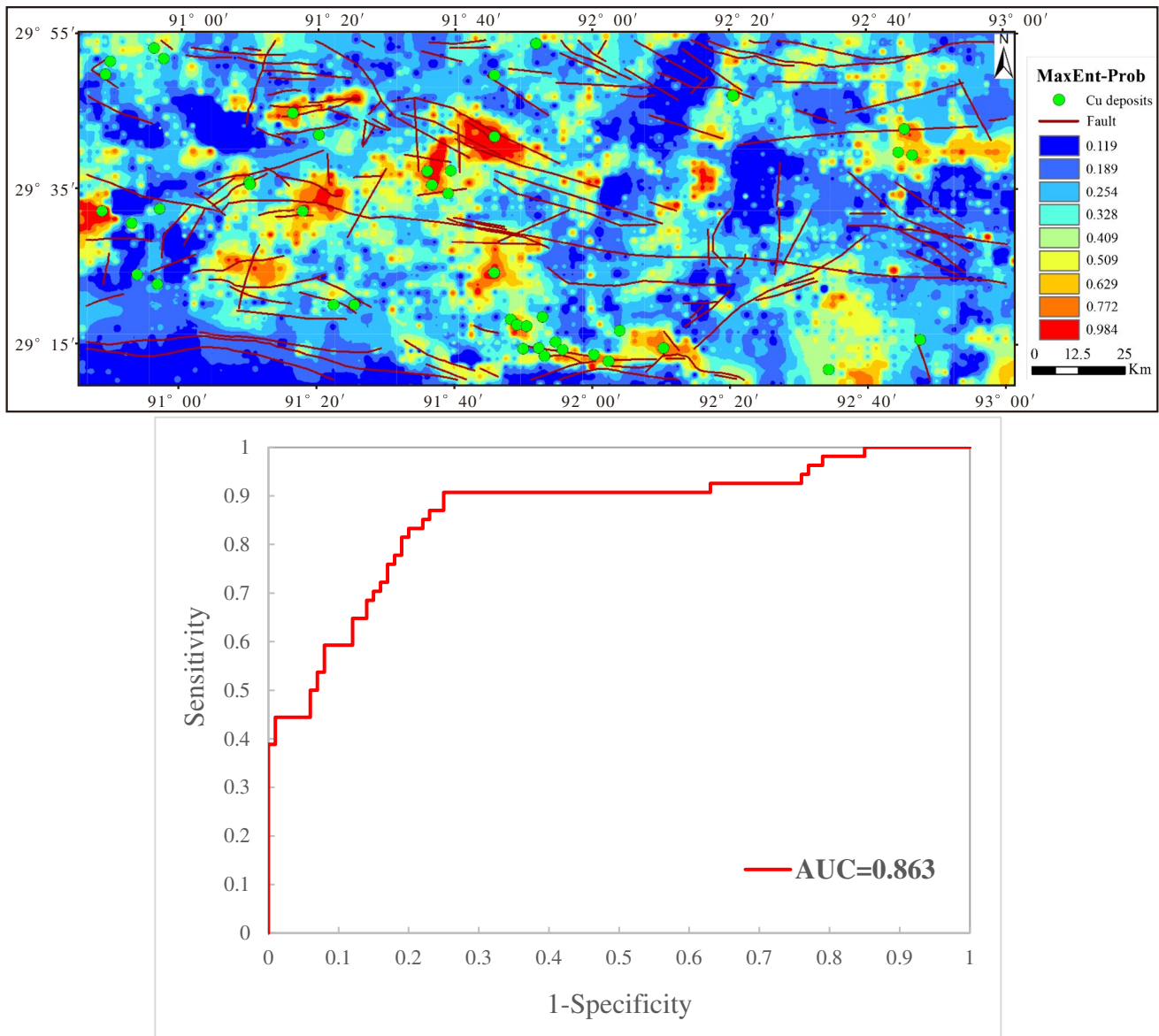
Fig. 6 Copper prospectivity map produced by optimized factor scores using MaxEnt model and the ROC curve

and large faults are usually rock and ore guiding structures, and copper deposits are distributed along main faults or branch faults (Li and Rui 2004; Liu et al. 2020). The mineralization corresponds with the anomaly probability, and their locations are clearly seen on the geochemical maps in Figure 6. In fact, the areas with high anomaly probability are always located at the edges of intermediate-felsic intrusions and the sides of a fault, or in the vicinity of fault intersections, which are favorable metallogenic regions in the study area. These results are consistent with the previous findings in the study area (Zuo et al. 2009; Zuo 2011). In addition, based on the geological background and suitable ore-forming conditions, such as intermediate-felsic intrusions, intersection of faults, and favorable sedimentary rocks, several

preliminary prospecting targets are proposed in Figure 9. However, these predicted prospects still need to be further investigated with more information.

Table 5 The contributions of ore-controlling variables

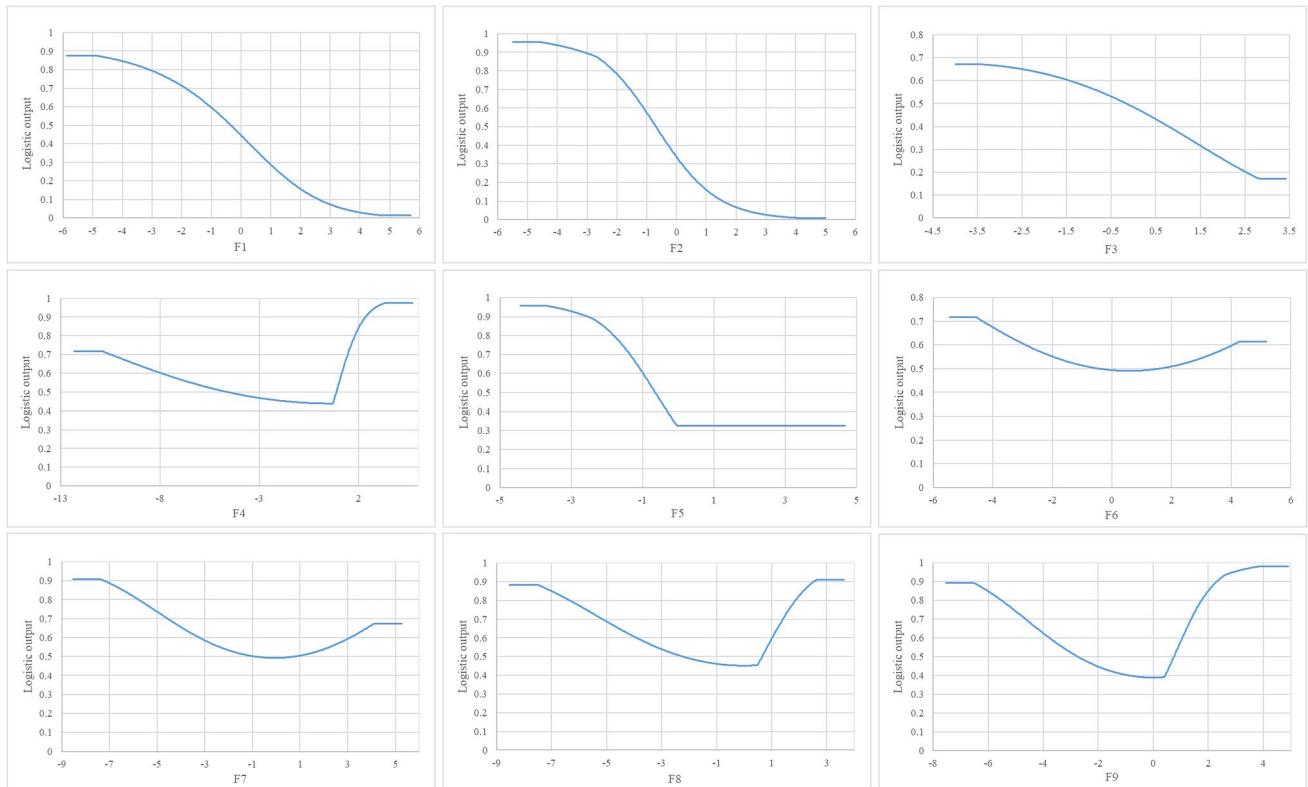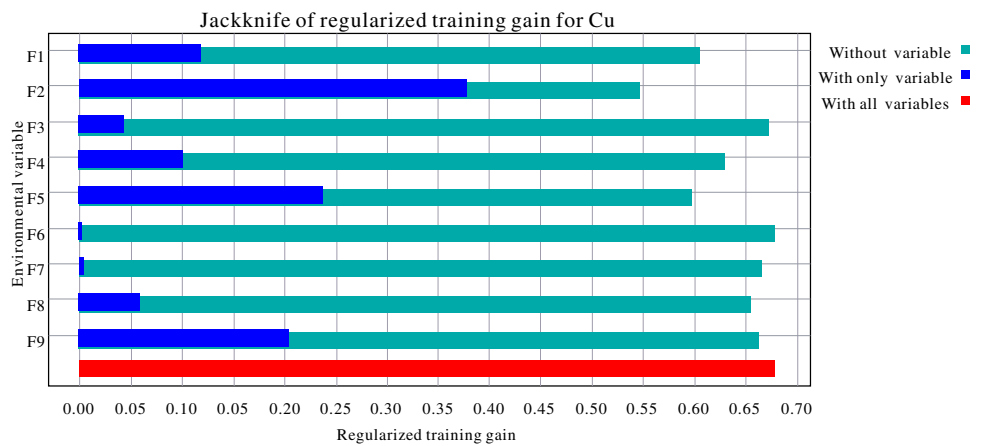| Variable	Percent contribution	Variable	Percent contribution
F1	7.80%	F6	2.30%
F2	29.30%	F7	2.70%
F3	7.60%	F8	5.60%
F4	13.50%	F9	5.50%
F5	25.80%		

Fig. 7 Response of Cu to factor scores

**Fig. 8** Evaluation of the relative importance of factor scores variables by Jackknife test

Conclusions

(1) Combinations of elements as grouped by factor analysis can reflect the symbiotic and origin relationships of elements. Compared with the original complex element aggregation, the factor analysis results are clearer and simpler. By combining the distribution area and geological characteristics of the corresponding factor combinations, directions for future geological work can be identified.

(2) The MaxEnt model used in this study is simple, accurate, and easy to operate, and can fuse multivariate data quickly and effectively. The AUC, kappa, and TSS values illustrate the ability of the model to correctly classify copper deposits.

(3) The spatial association of individual ore-controlling variables with occurrences of copper deposits was investigated by response curves, and the relative importance of ore-controlling variables was examined by jackknife analysis in the MaxEnt model, indicating
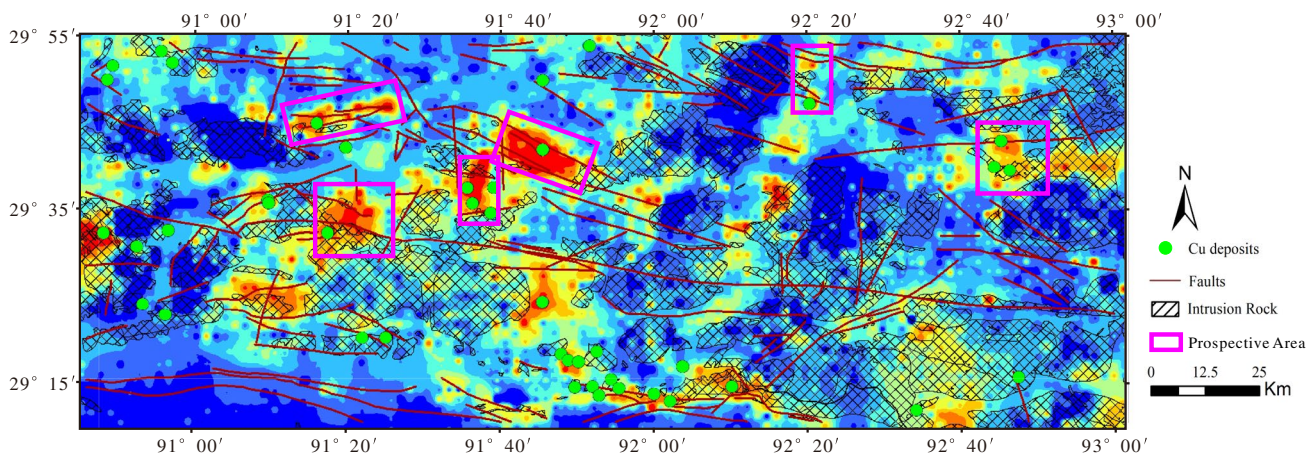
Fig. 9 The targets area for copper deposits

that the second factor score was the most important variable, followed by the fifth factor score.

(4) The MaxEnt model can automatically extract and analyze multivariate geochemical anomaly information without relying on expert experience. It is universal and efficient, and can adapt to processing in the big data environment. However, this method also has the problem that the particle size of anomaly extraction is not fine enough. Further narrowing the scope of anomaly extraction to delineate the prospecting target area more accurately is necessary.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aghelpour P, Mohammadi B, Biazar SM, Kisi O, Sourmirinezhad Z (2020) A theoretical approach for forecasting different types of drought simultaneously, using entropy theory and machine-learning methods. ISPRS Int J Geo-Inf 9:1–26

Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, London

Aitchison JA, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2000) Logratio analysis and compositional distance. [J]. Mathe Geol 32(3):271–275

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). [J]. J Appl Ecol 43(6):1223–1232

Araújo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. [J]. Glob Change Biol 11(9):1504–1513

Azareh A, Rahmati O, Rafiei-Sardooi E, Sankey JB, Lee S, Shahabi H et al (2019) Modelling gully-erosion susceptibility in a semi-arid region, Iran: investigation of applicability of certainty factor and maximum entropy models. Sci Total Environ 655:684–696

Biazar SM, Fard AF, Singh VP, Dinpashoh Y, Majnooni-Heris A (2020) Estimation of evaporation from saline-water with more efficient input variables. Pure Appl Geophys 177:5599–5619

Biazar SM, Ferdosi FB (2020) An investigation on spatial and temporal trends in frost indices in Northern Iran. Theor Appl Climatol 141(3):907–920

Buccianti A, Grunsky E (2014) Compositional data analysis in geochemistry: are we sure to see what really occurs during natural processes? J Geochem Exploration 141:1–5

Carranza EJM (2011) Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values [J]. J Geochem Explor 110(2):167–185

Chen X, Xu R, Zheng Y, Jiang X, Du W (2018) Identifying potential Au-Pb-Ag mineralization in SE Shuangkoushan, North Qaidam, Western China: combined log-ratio approach and singularity mapping. [J]. J Geochem Explor 189:109–121

Chen X, Zheng Y, Xu R, Wang H, Jiang X (2016) Application of classical statistics and multifractals to delineate Au mineralization-related geochemical anomalies from stream sediment data: a case study in Xinghai-Zeku, Qinghai, China. [J]. Geochemistry: Explor Environ Anal 16(3-4):253-264

Chen Y, and Wu, W (2017) Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. [J]. Geochemistry: Explor Environ Anal 17(3):231–238

Cheng Q (2012) Ideas and methods for mineral resources integrated prediction in covered areas. [J]. Earth Sci 37(6):1109–1125

Cheng Q (2012) Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. [J]. J Geochem Explor 122(11):55–70

Cheng QM, Chen YQ, Zuo RG (2021) Preface to the special issue on digital geosciences and quantitative exploration of mineral resources. [editorial material]. J Earth Sci 32(2):267–268. https://doi.org/10.1007/s12583-021-1460-9

Coetzee BW, Robertson MP, Erasmus BF, Van Rensburg BJ, Thuiller W (2009) Ensemble models predict important bird areas in southern Africa will become less effective for conserving endemic birds under climate change. [J]. Glob Ecol Biogeogr 18(6):701–710

Cohen J (1960) A coefficient of agreement for nominal scales. [J]. Educ Psychol Meas 20(1):37–46

Cong Y, Chen J, Xiao K, Dong Q (2012) The extraction of regional geochemical element composite anomalies in northern Sanjiang region and its prospecting significance. Geol Bull China 31(7):1164–1169

Dong Y, Hinton GE, Morgan N, Chien JT, Sagayama S (2012) Introduction to the special section on deep learning for speech and language processing. [J]. IEEE Trans Audio Speech Lang Process 20(1):4–6

Dudík M, Phillips SJ, Schapire RE (2004) Performance guarantees for regularized maximum entropy density estimation. In Proceedings of the 17th Annual Conference on Computational Learning Theory, New York, (pp. 655-662): ACM Press

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric log ratio transformations for compositional data analysis. [J]. Mathe Geol 35(3):279–300

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. [J]. Divers Distrib 17(1):43–57

Filzmoser P, Hron K (2009) Correlation analysis for compositional data. [J]. Mathe Geosci 41(8):905

Filzmoser P, Hron K, Reimann C (2009) Principal component analysis for compositional data with outliers. [J]. Environmetrics 20(6):621–632

Filzmoser P, Hron K, Reimann C (2009) Univariate statistical analysis of environmental (compositional) data: problems and possibilities. [J]. Sci Total Environ 407(23):6100–6108

Filzmoser P, Hron K, Reimann C, Garrett R (2009) Robust factor analysis for compositional data [J]. Comput Geoences 35(9):1854–1861

Gao Y (2019) Mineral prospecting information mining and mapping mineral prospectivity for copper polymetallic mineralization in southwest Fujian province, China. China University of Geosciences, China

Hong S, Zuo RG, Huang XW, Xiong YH (2021) Distinguishing IOCG and IOA deposits via random forest algorithm based on magnetite composition. [Article]. J Geochem Explor 230:9 https://doi.org/10.1016/j.gexplo.2021.106859

Hoseinpoor MK, Aryafar A (2016) Using robust staged R-mode factor analysis and logistic function to identify probable Cu-mineralization zones in Khusf 1:100,000 sheets, east of Iran. [J]. Arab J Geosci 9(2):1–11

Hou Z, Qu X, Yang Z, Meng X, Li Z, Yang Z, et al. (2006) Metallogenesis in Tibetan collisional orogenic belt: III. Mineralization in post collisional extension setting. Mineral Depos 25(6):629–651

Jahangir MS, Biazar SM, Hah D, Quilty J, Isazadeh M (2021) Investigating the impact of input variable selection on daily solar radiation prediction accuracy using data-driven models: a case study in northern Iran. Stoch Environ Res Risk Assess 35:1–25

Jaynes ET (1957) Information theory and statistical mechanics. [J]. Phys Rev 106(4):620

Johnson R, DW W (2002) Applied multivariate statistical analysis. Prentice Hall, Upper Saddle River

Jolliffe IT (2002) Principal component analysis. [J]. J Mark Res 87(4):513

Kong W, Li X, Zou H (2019) Optimizing MaxEnt model in species distribution predication. Chin J Appl Ecol 30(6):2116–2128

Kramer-Schadt S, Niedballa J, Pilgrim JD, Schrder B, Lindenborn J, Reinfelder V et al (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. Divers Distrib 19(11):1366–1379

Lang XH, Tang JX, Chen YC, Li ZJ, Huang Y, Wang CH et al (2012) Neo-Tethys mineralization on the southern margin of the Gangdise Metallogenic Belt, Tibet, China: evidence from Re-Os ages of Xiongcun orebody No.I. Earth Sci 37:515–525

Li B, Liu B, Guo K, Li C, Wang B (2019) Application of a maximum entropy model for mineral prospectivity maps. [J]. Minerals 9(9):1–23

Li B, Liu B, Wang G, Chen L, Guo K (2021) Using geostatistics and maximum entropy model to identify geochemical anomalies: a case study in Mila Mountain region, southern Tibet. [J]. Appl Geochem 124:104843

Li G, Rui Z (2004) Diagenetic and mineralization ages for the porphyry copper deposits in the Gangdise metallogenic belt, southern Xizang. Geotecton Metallog 28(2):165–170

Li H (2012) Statistical methods. Tsinghua University, Beijing

Lin B, Chen Y, Tang J, Wang Q, Song Y, Yang C et al (2017) 40Ar/39Ar and Rb-Sr ages of the Tiegelongnan Porphyry Cu-(Au) deposit in the Bangong Co-Nujiang Metallogenic Belt of Tibet, China: implication for generation of super-large deposit. Acta Geol Sin (English Edition) 91(2):602–616

Lin B, Tang J, Chen Y, Baker M, Song Y, Yang H et al (2019) Geology and geochronology of Naruo large porphyry-breccia Cu deposit in the Duolong district, Tibet. Gondwana Res 66:168–182

Lin B, Tang JX, Chen YC, Song Y, Hall G, Wang Q et al (2017) Geochronology and genesis of the Tiegelongnan Porphyry Cu (Au) deposit in Tibet: evidence from U-Pb, Re–Os Dating and Hf, S, and H–O isotopes. Resour Geol 67(1):1–21

Liu B, Guo K, Li C, Zhou J, Liu X, Wang X et al (2020) Copper prospectivity in Tibet, China: based on the identification of geochemical anomalies. Ore Geol Rev 120:1–8

Liu Y, Zhou K, Zhang N, Wang J (2018) Maximum entropy modeling for orogenic gold prospectivity mapping in the Tangbale-Hatu belt, western Junggar, China. [J]. Ore Geol Rev 100:133–147

Meigoony MS, Afzal P, Gholinejad M, Yasrebi AB, Sadeghi B (2014) Delineation of geochemical anomalies using factor analysis and multifractal modeling based on stream sediments data in Sarajeh 1:100,000 sheet, Central Iran. [J]. Arab J Geoences 7(12):5333–5343

Mejía-Herrera P, Royer J-J, Caumon G, Cheilletz A (2015) Curvature attribute from surface-restoration as predictor variable in Kupferschiefer copper potentials. [J]. Nat Resour Res 24(3):275–290

Moeini H, Torab FM (2017) Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran. [J]. J Geochem Explor 180:15–23

Muller J, Kylander M, Martinez-Cortizas A, Wüst RAJ, Weiss D, Blake K et al (2008) The use of principle component analyses in characterising trace and major elemental distribution in a 55 kyr peat deposit in tropical Australia: implications to paleoclimate. [J]. Geochim Cosmochim Acta 72:449–463

Phillips SJ (2005) A brief tutorial on Maxent. AT&T Res 190(4):231–259

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. [J]. Ecol Model 190(3–4):231–259

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. Paper presented at the Proceedings of the twenty-first international conference on Machine learning, New York.

Phillips SJ, Jane E (2013) On estimating probability of presence from use-availability or presence-background data. [J]. Ecol 94(6):1409–1419

Radosavljevic A, Anderson RP (2014) Making better MaxEnt models of species distributions: complexity, overfitting and evaluation. [J]. J Biogeogr 41(4):629–643

Rahimi H, Abedi M, Yousefi M, Bahroudi A, Elyasi G-R (2021) Supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites thereof. Appl Geochem 128:104940

Ramezanali AK, Feizi F, Jafarirad A, Lotfi M (2019) Geochemical anomaly and mineral prospectivity mapping for vein-type copper mineralization, Kuhsiah-e-Urmak area, Iran: application of sequential Gaussian simulation and multivariate regression analysis. Nat Resour Res 1–30

Ratnaparkhi A (2016) A simple introduction to maximum entropy models for natural language processing. Encyclopedia of Machine Learning and Data Mining (pp. 1-6). New York

Reimann C, Filzmoser P, Garrett RG (2005) Background and threshold: critical comparison of methods of determination. [J]. Sci Total Environ 346(1–3):1–16

Song Y, Yang C, Wei S, Yang H, Fang X, Lu H (2018) Tectonic control, reconstruction and preservation of the Tiegelongnan porphyry and epithermal overprinting Cu (Au) deposit, central Tibet, China. Minerals 8(9):398

Sun T, Chen F, Zhong L, Liu W, Wang Y (2019) GIS-based mineral prospectivity mapping using machine learning methods: a case study from Tongling ore district, eastern China. [J]. Ore Geol Rev 109:26–49

Swets JA (1988) Measuring the accuracy of diagnostic systems. [J]. Science 240:1285–1293

Wang B, Xu Y, Ran J (2017) Predicting suitable habitat of the Chinese monal (Lophophorus lhuysii) using ecological niche modeling in the Qionglai Mountains, China. [J]. Peerj 5(7):e3477

Wang J (2018) Identification of geochemical anomalies based on geo-statistical simulation. China University of Geosciences, China

Wang J, Zuo RG (2020) Assessing geochemical anomalies using geographically weighted lasso. [Article]. Appl Geochem 119:15. https://doi.org/10.1016/j.apgeochem.2020.104668

Wang X, Xie X, Zhang B, Hou Q (2011) Geochemical probe into China's continental crust. Acta Geosci Sin 32(2011):65–83

Wang Y, Gong P, Gong M, Ma Z (2010) Geological subdivisions in metallogenic belt with the 1:200000 stream sediments and its geological significance: a case study in Gangdise copper polymetallic metallogenic belt. Geoscience 24(4):801–806

Wu R, Chen J, Zhao J, Chen J, Chen S (2020) Identifying geochemical anomalies associated with gold mineralization using factor analysis and spectrum-area multifractal model in Laowan District, Qinling-Dabie Metallogenic Belt, Central China. [J]. Minerals 10(3):1–18

Xie X (2008) Global geochemical mapping — historical development and recommendations for future work. Geol China 35(3):357–374

Xie X, Mu X, Ren T (1997) Geochemical mapping in China. [J]. J Geochem Explor 60(1):99–113

Xie X, Wang X, Zhang Q, Zhou G, Cheng H, Liu D, et al. (2008) Multi-scale geochemical mapping in China. [Article]. Geochemistry: Explor Environ Anal 8(1):333–341

Xiong Y, Zuo R, Wang K, Wang J (2018) Identification of geochemical anomalies via local RX anomaly detector. [J]. J Geochem Explor 189:64–71

Xu Y, Wu Z, Long J, Song X (2014) A maximum entropy method for a robust portfolio problem. [J]. Entropy 16(6):3401–3415

Yang B, Li D, Yuan S, Jin L (2021) Role of biochar from corn straw in influencing crack propagation and evaporation in sodic soils. Catena 204:105457

Yang B, Xu K, Zhang Z (2020) Mitigating evaporation and desiccation cracks in soil with the sustainable material biochar. Soil Sci Soc Am J 84(2):461–471

Yang Z, Hou Z (2009) Genesis of giant porphyry Cu deposit at Qulong, Tibet: constraints from fluid inclusions and H-O isotopes. Acta Geol Sin (English Edition) 83:1838–1859

Yin A, Harrison TM (2000) Geologic evolution of the Himalayan-Tibetan Orogen. Ann Rev Earth Planet Sci 28(1):211–284

Yousefi M, Kamkar-Rouhani A, Carranza EJM (2014) Application of staged factor analysis and logistic function to create a fuzzy stream sediment geochemical evidence layer for mineral prospectivity mapping. [J]. Geochem-Explor Environ Anal 14(1):45–58

Zhang B, Fan J, Luo A, Yu Y, Hao Y (2019) Characteristics and tectonic significance of the Miocene Strata in the Milashan Area, Eastern Lhasa Terrane. Earth Sci 44(7):2392–2407

Zhang S, Carranza EJM, Wei H, Xiao K, Yang F, Xiang J et al (2021) Data-driven mineral prospectivity mapping by joint application of unsupervised convolutional auto-encoder network and supervised convolutional neural network. Nat Resour Res 30(2):1011–1031

Zhang S, Xiao K, Carranza EJM, Yang F (2019) Maximum entropy and random forest modeling of mineral potential: analysis of Gold prospectivity in the Hezuo-Meiwu District, West Qinling Orogen, China. [J]. Nat Resour Res 28(3):645–664

Zheng W, Liu B, M MJ, R CM, Wang L (2021) Geology and geochemistry-based metallogenic exploration model for the eastern Tethyan Himalayan metallogenic belt, Tibet. [J]. J Geochem Explor 224:106743

Zheng W, Tang J, Zhong K, Ying L, Leng Q, Ding S et al (2016) Geology of the Jiama porphyry copper–polymetallic system, Lhasa Region, China. Ore Geol Rev 74:151–169

Zuo R (2011) Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal modeling in the Gangdese Belt, Tibet (China) [J]. J Geochem Explor 111(1):13–22

Zuo R (2014) Identification of weak geochemical anomalies using robust neighborhood statistics coupled with GIS in covered areas. [J]. J Geochem Explor 136:93–101

Zuo R, Cheng Q, Agterberg FP, Xia Q (2009) Application of singularity mapping technique to identify local anomalies using stream sediment geochemical data, a case study from Gangdese, Tibet, western China [J]. J Geochem Explor 101(3):225–235

Zuo R, Kreuzer OP, Wang J, Xiong Y, Zhang Z, Wang Z (2021) Uncertainties in GIS-based mineral prospectivity mapping: key types, potential impacts and possible solutions. Nat Resour Res 30:3059–3079

Zuo R, Xia Q, Wang H (2013) Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. [J]. Appl Geochem 28:202–211

Zuo R, Xiong Y, Wang J, Carranza EJM (2019) Deep learning and its application in geochemical mapping. [J]. Earth-Sci Rev 192:1–14