



Original Paper

Prospectivity Mapping for Tungsten Polymetallic Mineral Resources, Nanling Metallogenic Belt, South China: Use of Random Forest Algorithm from a Perspective of Data Imbalance

Tongfei Li,^{1,2} Qinglin Xia,^{1,3,6} Mengyang Zhao,⁴ Zhou Gui,⁵ and Shuai Leng¹

Received 31 December 2018; accepted 23 September 2019
Published online: 3 October 2019

Mineral systems are composed of many interacting components that lead to complex, singular and rare properties of geo-data. In mineral prospectivity mapping (MPM), supervised machine learning algorithms, which have advantages in dealing with complex geo-data, usually involve uncertainty resulting from the discretization of continuous evidential maps into arbitrary classes as well as the large data imbalance caused by the rarity of deposit locations. Consequently, the predicted results may be biased. In this paper, a random forest (RF) algorithm based on the bagging technique is used to map the prospectivity of tungsten polymetallic deposits in the Nanling metallogenic belt. Data-driven logistic transformation is employed to obtain continuous evidential maps. Both discretized and continuous evidential maps are used to generate prospectivity models for comparison. To reduce the data imbalance, the under-sampling method and the synthetic minority over-sampling technique (SMOTE) are implemented to generate balanced datasets. The receiver operating characteristic (ROC) curve and improved prediction-area (P-A) plot are applied to evaluate the prospectivity models. The predictive results show that when using the RF algorithm in MPM, the application of continuous evidential maps can improve the performance of prospectivity models and reduce the uncertainty resulting from the discretization of evidential maps. The prospectivity model trained with a balanced SMOTE-generated dataset shows the best overall performance for improving the percentage of deposit locations that are correctly predicted and decreasing the percentage of non-deposit locations that are inaccurately identified as deposit locations to some extent. In addition, the improved P-A plot is superior to the ROC curve because the latter neglects the occupied area, which is critical for mineral exploration and may provide an overly optimistic performance with imbalanced data. However, further testing of the evaluation criteria and the SMOTE approach to reduce data imbalance is warranted to determine fully the universality in MPM from the perspective of data imbalance. Based on prospectivity models, four high-potential areas and five moderate-potential areas are delineated, which indicates good future prospecting for tungsten polymetallic deposits in this region.

KEY WORDS: Random forest, Data imbalance, SMOTE, Improved P-A plot, Mineral prospectivity mapping.

¹School of Earth Resources, China University of Geosciences, Wuhan 430074, China.

²Wuhan Center of Geological Survey CGS, Wuhan 430205, China.

³Cooperative Innovation Center for Scarce Mineral Resources Exploration, China University of Geosciences, Wuhan 430074, China.

⁴Exploration Unit of North China Geological Exploration Bureau, Langfang 065201, China.

⁵School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China.

⁶To whom correspondence should be addressed; e-mail: qlxia@cug.edu.cn

INTRODUCTION

Mineral systems are complex, consisting of many components (e.g., geological, geochemical, and geophysical) that interact with each other as the systems evolve with time (Wyborn et al. 1994; Zhai et al. 1999, 2000, 2002; Zhai 2003a, b; Cheng 2008a). Due to the complexity of mineral systems, geo-data have complex, singular, and rare properties with nonlinear relationships, which lead to challenges in understanding geological processes and uncertainty in predictions (Agterberg 1989; Bárdossy and Fodor 2001, 2004; Yu 2006; Cheng 2008a, b; Zuo and Xia 2008).

Specifically, in mineral exploration, economic geologists are concerned with extracting information associated with mineralization from a very large collection of complex geo-data (e.g., geological, geochemical, geophysical, remote sensing and natural heavy mineral data), which may exhibit the properties of large volume, high dimension, complex distribution, and nonstationary and nonlinear relationships (Viktor and Kenneth 2013; Tan et al. 2017; Zuo 2017; Zhao 2018). Hence, mineral exploration is a kind of decision-making under the condition of uncertainty (Zhao et al. 2013). The basic task of mineral prospectivity mapping (MPM) is to reduce the uncertainty and risk in mineral exploration by narrowing the target area ranging from the regional to the deposit scale (Zhao et al. 2013; Porwal and Carranza 2015). From a mathematical perspective, multisource geo-data can be regarded as the input, while the occurrence of a particular type of mineral deposit can be viewed as the output. The procedure of data fusion can be modeled as a classification function that combines the input and output variables (Hariharan et al. 2017). In MPM, spatially continuous evidential maps (e.g., distance to structures and geochemical signatures) are usually discretized into classes using arbitrary intervals. Weights are then assigned to every class based on the subjective judgment of the analyst or on locations of mineral deposits, or functions are used to calculate the weights of classes of discretized input variables (Luo 1990; Bonham-Carter 1994; Luo and Dimitrakopoulos 2003; Porwal et al. 2003b; Zhang et al. 2014; Cheng 2015). Subsequently, discretized evidential maps, which have been assigned weights by the three approaches mentioned above, are assembled using a mathematical model to generate prospectivity maps. The procedure of MPM to dis-

cretize spatially continuous evidential maps may result in exploration bias and uncertainty resulting from the following three aspects: (1) MPM is sensitive to the choice of the class interval, which may lead to a biased estimation of the weights of classes because of the approximation involved in classification; (2) the assignment of meaningful weights to every discretized evidential map is a highly subjective exercise involving trial and error; and (3) stochastic bias and error can also be induced due to data sufficiency when using the locations of known mineral deposits as training sites to assign weights to evidential maps (Nykänen et al. 2008; Yousefi and Carranza 2015a, b, 2016, 2017a, b). Nykänen et al. (2008) and Yousefi et al. (2012, 2014) assigned weights to continuous evidential maps using specific membership functions (e.g., 'large', 'small', and 'logistic') without discretizing the continuous evidential maps into arbitrary classes and without using locations of known mineral deposits. However, these methods still incurred exploration bias due to the trial-and-error aspect involved [e.g., determining the slope (s) and inflection point (i) in the logistic function]. To overcome this problem, Yousefi and Nykänen (2016) proposed a data-driven method to define s and i .

With respect to data fusion, machine learning algorithms (MLs) have demonstrated great advantages in handling complex input variables compared to traditional data-driven methods in MPM (e.g., weight of evidence and multivariate statistical methods) (Singer and Kouda 1999; Porwal et al. 2003a; Nykänen 2008; Zuo and Carranza 2011; Chen et al. 2014b; Carranza and Laborte 2015a; Harris et al. 2015; McKay and Harris 2016; Chen and Wu 2017; Parsa et al. 2018). In recent years, the random forest (RF) algorithm, which is a kind of supervised ML method, has been widely applied in MPM. This algorithm has shown better performance than other MLs (e.g., neural networks and support vector machines) due to its higher success rate, greater stability, simpler parameter settings and increased resistance to overfitting (Rodríguez-Galiano et al. 2014; Carranza and Laborte 2015a, b, 2016; Gao et al. 2016; Hariharan et al. 2017). Additionally, the RF algorithm can provide the relative importance of the predictive variables, which coincides with well-known geologic expectations (Rodríguez-Galiano et al. 2014). However, in most cases of MPM using the RF algorithm, spatially continuous evidential maps are usually discretized into arbitrary classes, which may lead to exploration bias and uncertainty,

as mentioned above. Roshanravan et al. (2019b) demonstrated the superiority of using continuously weighted spatial evidence values compared to discretely weighted evidence data in MPM using an artificial neural network. Parsa et al. (2018) applied logistic regression and RF models with continuous predictor variables to map skarn-type copper prospectivity to reduce the possible uncertainty resulting from discretized predictor variables. However, little work has been conducted to discuss the performance of prospectivity models generated from discretized and continuous evidential maps using the RF algorithm.

Another problem that occurs with the supervised MLs in MPM is data imbalance. Because mineralization is a singular process, the occurrence of mineral deposits exhibits rarity, resulting in fractal spatial and temporal distributions (Cheng 2006, 2008a, b). Furthermore, limited geological observations and verifications are other factors that cause data imbalance in MPM. Consequently, the number of deposit locations is far smaller than that of non-deposit locations, which results in a large data imbalance in MPM. Therefore, because of data imbalance, supervised MLs tend to ignore the minority class and are biased to the majority class (Chawla et al. 2004; Hariharan et al. 2017). When using MLs in MPM, the output values are a series of floating numbers between 0 (not representing a deposit) and 1 (representing a deposit) denoting the likelihood of mineral deposit occurrence, which can be reclassified using the threshold value 0.5 to map prospective and non-prospective areas when the training dataset is relatively balanced (Carranza and Laborte 2015a, b). In general, there are two types of misclassification errors in the reclassification procedure: (1) false negative (FN) errors that classify a prospective area as a non-prospective area and (2) false-positive (FP) errors that classify a non-prospective area as a prospective area (Zhao et al. 2013; Xiong and Zuo 2017). The costs of these two types of errors are vastly different; the former can result in the loss of important deposits, while the latter can result in the waste of manpower and financial resources (Zhao et al. 2013; Xiong and Zuo 2017). Consequently, if the training dataset is imbalanced, the predictive prospectivity map may then be biased when using 0.5 as the threshold value for reclassification, which may result in exploration risks. Additionally, the predictive accuracy, a kind of performance evaluation index of classifiers, might not be appropriate when the data are imbalanced

and/or the costs of different errors vary markedly (Chawla 2009). To reduce the impact of data imbalance, four strategies have been employed: (1) assign distinct costs to training datasets, such as cost-sensitive neural networks (Pazzani et al. 1994; Xiong and Zuo 2017); (2) use one-class learning algorithms (Chen and Wu 2017; Gonçalves et al. 2018); (3) use unsupervised algorithms, such as deep autoencoder neural networks (Xiong and Zuo 2016; Xiong et al. 2018); and (4) use sampling techniques, such as under-sampling, over-sampling, and other synthetic methods (Chawla et al. 2002; Chawla 2009; He and Garcia 2010; Hoens and Chawla 2013). With respect to the sampling techniques, researchers have noted that the use of an equal number of negative samples (e.g., non-deposit locations) and positive samples (e.g., deposit locations) in a regression is optimal when the latter represents rare events (Breslow and Cain 1988; Schill et al. 1993). King and Zeng (2001) mentioned that the information content of predictors starts to diminish as the number of negative samples exceeds the number of positive samples. Nykänen et al. (2015) suggested using the locations of other deposit types or random locations to represent non-deposit locations. Carranza and Laborte (2015b) summarized three criteria for the selection of target variables: (1) the number of negative samples (or non-deposit locations) should be equal to the number of positive samples (or deposit locations); (2) non-deposit locations should be distal to any deposit location because locations proximal to existing mineral deposits are likely to have multivariate spatial data signatures similar to those of the deposit locations and thus preclude achievement of the desired results; and (3) non-deposit locations must be randomly spatially distributed. To this end, spatial point pattern analysis is recommended to determine the optimal distance to deposit locations and ensure that the non-deposit locations are distributed randomly. Hariharan et al. (2017) selected the non-deposit locations according to the geological conditions and then applied a synthetic minority over-sampling technique (SMOTE) to generate more balanced training datasets to reduce data imbalance.

In this paper, the Nanling metallogenic belt (NMB) in South China is selected as the study area. Discretized and continuous evidential maps are used to obtain prospectivity models using the RF algorithm. The under-sampling method and SMOTE are used to generate training datasets from the perspective of data imbalance. Subsequently, the RF

algorithm is employed to map the prospectivity of the tungsten polymetallic deposits in this region. The receiver operating characteristic (ROC) curve and improved prediction-area (P-A) plot are compared to evaluate the performance of the prospectivity models. This paper has two main purposes: (1) to demonstrate the superiority of using continuous evidential maps over discretized evidential maps in MPM and (2) to examine data imbalance in MPM using the RF algorithm.

METHODOLOGY

Logistic-Based Transformation

Defining a suitable nonlinear transformation into a new space could facilitate the interpretation of a pattern for a set of evidential values in MPM when compared to defining a nonlinear function in the original space (Yousefi et al. 2014; Yousefi and Carranza 2015a). The logistic sigmoid function, which provides an optimal decision boundary for classification, has played an important role in pattern recognition (Bishop 2007; Zhou 2016). The logistic function transforms an individual evidential map into the same space and can distinguish the classification boundary more efficiently (Yousefi and Carranza 2017a). In the general expression of the logistic function, the inflection point (i) and slope (s) are usually determined by trial-and-error methods. Yousefi and Nykänen (2016) proposed a data-driven logistic-based function in which the maximum value of an evidential map is assigned a value of 0.99, while the minimum value of an evidential map is assigned a value of 0.01. By solving the system of equations using the maximum and minimum values of weights, the values of i and s in the logistic function can be obtained. This data-driven logistic-based function can avoid the trial-and-error procedure used for other types of functions (e.g., 'large' and 'small') and can estimate the relative importance of evidential maps for MPM more realistically (Almasi et al. 2017; Yousefi and Carranza 2017a; Yousefi and Nykänen 2017).

Point Pattern Analysis

In a regional-scale prospectivity analysis, it is almost universally accepted that mineral deposits can be adequately represented by spatial points

(Lisitsin 2015). Hence, the distribution of mineral deposits can be investigated by various techniques of point pattern analysis. Fry analysis (Fry 1979), which was originally applied to assess the strain partitioning in rocks, has been widely used to quantify trends in the distributions of mineral deposits (Vearncombe and Vearncombe 1999; Yaghubpur and Hassannejad 2006; Carranza 2009c; Zuo et al. 2009; Najafi et al. 2010). Assume that a sheet of transparent paper with n marked points is placed over the point pattern. Then, the transparent paper is shifted, maintaining its original orientation, so that one of the original points coincides with one of the points in the point pattern; then, the locations of other points in the point pattern can be mapped on the transparent paper. Repeat this process for the remaining $(n - 1)$ points. As a result, $n \times (n - 1)$ Fry points are obtained on the original transparent paper (Fry 1979). A rose diagram is used to portray the orientation frequencies of the vector between any two Fry points. The rose diagram plotted for all Fry points may represent the distribution orientations of mineral deposits at the regional scale, while the rose diagram plotted for Fry points that are located within a specific distance could provide ore control information at the local scale.

Fractal analysis (Mandelbrot 1983) has been widely applied to investigate whether deposits tend to be close or distal, which has major implications for exploration targeting (Carlson 1991; Cheng and Agterberg 1996; Raines 2008; Zuo et al. 2009; Lisitsin 2015; Haddad-Martim et al. 2017; Li et al. 2018; Parsa et al. 2018). Box-counting and radial-density analyses are two common methods to quantify the spatial heterogeneity of mineral deposits. Box-counting analysis involves converting mineral deposits into a series of cells of different sizes. Then, the relationship between cell size and the number of cells containing mineral deposits obeys a power-law relationship (Mandelbrot 1983):

$$N(\varepsilon) = C_1 \times \varepsilon^{-D_b} \quad (1)$$

where $N(\varepsilon)$ is number of cells containing at least one deposit, C_1 is a constant, ε is cell size, and D_b is the box-dimension.

In contrast, radial-density analysis involves calculating the radial density of mineral deposits within circles with different radii. The radial density and the corresponding radius also obey a power-law relationship (Mandelbrot 1983; Carlson 1991; Raines 2008):

$$d = C_2 \times r^{D_r-2} \quad (2)$$

where d is radial density, C_2 is a constant, r is radius and D_r is the radial-density fractal dimension. More detailed information about box-counting and radial-density analysis can be found in related papers (e.g., Cheng and Agterberg 1996; Carranza et al. 2009; Zuo et al. 2009; Li et al. 2018; Parsa et al. 2018).

Sampling Techniques for Imbalanced Datasets

The data imbalance problem, which causes suboptimal classification performance, is one of the challenges that have emerged in the application of ML algorithms (Chawla et al. 2004). Data imbalance occurs when one of the classes in a binary classification dominates in the data. Random over-sampling and under-sampling are non-heuristic but are the most practical methods to address this issue. The former method balances the data through random replication of the minority class, while the latter balances the data through elimination of the majority class. SMOTE (Chawla et al. 2002) is a kind of revised improved over-sampling method in which the minority class is over-sampled by creating synthetic examples in the feature space rather than by over-sampling with replacement in the data space. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. The method takes the difference between the feature vector (sample) under consideration as well as its nearest neighbor and then multiplies the difference by a random number between 0 and 1 and adds the result to the feature vector under consideration. Usually, under-sampling is also performed to reduce the number of majority classes. By applying a combination of under-sampling and over-sampling, the initial bias of the classifier toward the majority class is reversed in favor of the minority class (Chawla et al. 2002).

Random Forest Algorithm

The RF algorithm, which is a kind of ensemble learning method, is a classifier consisting of a collection of independently generated decision trees (Breiman 2001; Liaw and Wiener 2002). For each decision tree, bootstrap sampling with the replace-

ment method called bagging is employed to generate a dataset of which 2/3 is used for training, known as the in-bag data while the remaining 1/3 is used for validation and is known as out-of-bag (OBB) data (Breiman 1996). Afterward, from the root node, the data splitting process in each internal rule node of the tree is repeated until a previously specified stop condition is reached (Rodriguez-Galiano et al. 2015). All decision trees are eventually assembled, and the overall precision depends on the majority vote of the individual trees. The optimal split threshold for a decision tree is determined by the Gini impurity index (I_G) (Breiman et al. 1984), which is defined as:

$$I_G(f) = 1 - \sum_{i=1}^m f_i^2 \quad (3)$$

where f_i is probability of class i at node m and the lowest I_G corresponds to the optimal split threshold. Since the classification of an RF model is determined by the vote of all decision trees, the output of a random forest consisting of k decision trees can be described as (Breiman 2001),

$$P_j = \frac{1}{k} \sum_{j=1}^k y_j^i \quad (4)$$

where P_j is probability of classifying the input into the j th class, j denotes the number of classes (deposit or non-deposit in this case) and y_j^i denotes the predicted result that the input is assigned into the j th class by the i th decision tree.

Model Evaluation

There are many approaches to evaluate prospectivity models, including successive-rate curve (Agterberg and Bonham-Carter 2005), prediction rate curve (Fabbri and Chung 2008), receiver operating characteristic (ROC) curve (Nykänen et al. 2015; Gao et al. 2016; Zhang et al. 2016; Xiong and Zuo 2017), prediction-area (P-A) plot (Yousefi and Carranza 2015b, 2017b) and improved P-A plot (Roshanravan et al. 2019a).

On the ROC curve, the x -axis represents the false-positive rate (i.e., the percentage of non-deposit locations that are falsely predicted), $FPR = FP/(FP + TN)$, while the y -axis represents the true-positive rate (i.e., the percentage of mineral deposits that are truly predicted), $TPR = TP/(TP + FN)$.

Additionally, the area under the curve (AUC) value is an evaluation metric (Bradley 1997). The value of AUC ranges from 0 to 1, and the higher the AUC value, the better the model is.

In the improved P-A plot, the percentage of known deposits anticipated by prospectivity classes, the percentage of non-deposit locations anticipated by prospectivity classes and occupied areas of the corresponding prospectivity classes are employed to evaluate prospectivity models (Roshanravan et al. 2019a). The value on the left y-axis corresponding to the intersection of mineral deposit prediction rate and occupied area curves is similar to TPR in the ROC method, while the value on the left y-axis corresponding to the intersection of the non-deposit location prediction rate curve and occupied area curve is similar to FPR. Because one of the purposes of MPM is to promote the TPR and reduce the FPR at the same time, the index O_e , which is the difference between TPR and FPR, can be used to evaluate the overall performance of a prospectivity model. The value of O_e ranges from -1 to 1 , and the higher the positive O_e value, the better the model is. Detailed information about the advantages and disadvantages of different model evaluation methods can be found in Roshanravan et al. (2019a).

REGIONAL GEOLOGY AND DATASETS

Regional Geology

The NMB, which is part of the Cathaysian block in South China, is one of the largest tungsten polymetallic belts in the world (Fig. 1a) (Mao et al. 2005; Chen et al. 2008; Liu et al. 2010; Shu 2012). As a part of the Cathaysian block, the NMB has experienced multiple stages of tectonic–magmatic activity, forming a series of differently trending folds and faults and large amounts of crustal re-melting granite (S-type) (Shu et al. 2004; Shu and Wang 2006; Shu 2012). The exposed strata in this region can be divided into three groups: (1) Precambrian to Silurian strata composed of slate, sandstone and limestone; (2) Devonian to Triassic strata consisting of carbonate rocks and marlstone with interbedded clastic deposits; and (3) Jurassic to Cretaceous strata composed of volcanic rocks and red beds (Mao et al. 2009; Hua et al. 2013) (Fig. 1b). Most of the strata are enriched in mineralization elements to some extent, such as W, Sn, and Bi (Yu et al. 1987). Since the Mesozoic, this region has been influenced by the

transformation from the Tethys tectonic system to the Pacific tectonic system and has experienced lithospheric delamination and thinning, resulting in the formation of large amounts of S-type granite accompanied by large-scale tungsten polymetallic mineralization (Hua et al. 2003; Mao et al. 2007; Shu 2012). Chronologic studies show that the timing of tungsten polymetallic mineralization extended from 170 to 90 Ma, with peak mineralization ranging from 170 to 150 Ma (Mao et al. 2004, 2005, 2007; Zhou 2007). This S-type granite related to tungsten polymetallic mineralization shows highly differential geochemical characteristics of enrichments in Y and Rb and has a high Rb/Sr ratio, while it is depleted in Eu, Ba + Sr, and TiO_2 and has a low LREE/HREE ratio (Hua et al. 2003; Zhou 2007; Chen et al. 2008; Hu and Zhou 2012). There are three main types of tungsten polymetallic deposits in this region: quartz vein-, skarn-, and greisen-type deposits. The type of tungsten polymetallic deposit is largely dependent on the wall rock. Quartz vein-type tungsten polymetallic deposits often occur when the wall rock is shallow metamorphic sandstone or clastic rock, while skarn-type and greisen-type tungsten polymetallic deposits often occur when the wall rock is carbonate (Mao et al. 2008). The metallogenic conditions in this region are superior, and many large tungsten polymetallic deposits have been discovered, including Shizhuyuan, Piaotang, Xihuashan, Pangushan, Huangsha, and Dajishan.

Spatial Datasets

The 1:200,000 scale Bouguer gravity data and 1:200,000 scale geological map showing strata, magmatic rocks, faults, and tungsten polymetallic deposits originate from the China Geological Survey (CGS). The 1:200,000 scale geochemical data with 39 geochemical elements come from the Regional Geochemistry National Reconnaissance (RGNR) Project (Xie et al. 1997). Detailed information about the geochemical data used in this paper can be found in Liu et al. (2016).

Granite

Previous studies on S, Pb, and Hf isotopes indicate that the ore-forming materials and ore-bearing granite have a homologous relationship and exhibit characteristics of an upper crustal source

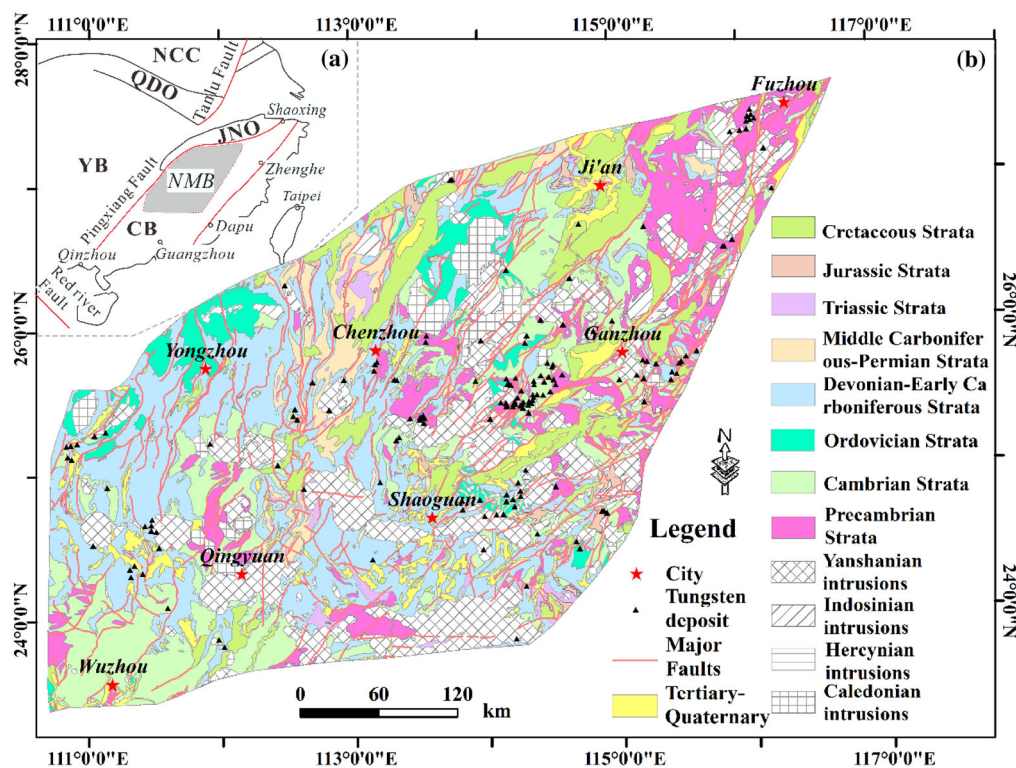


Figure 1. Simplified geological maps of the NMB: (a) tectonic sketch map of the NMB (NCC denotes the North China Craton; YB denotes the Yangtze Block; QDO denotes the Qinling-Dabie Orogen; JNO denotes the Jiangnan Orogen; CB denotes the Cathaysian Block) (revised from Zhou 2007). (b) Simplified 1: 2,000,000 scale geological map of the NMB (revised from Liu et al. 2014).

(Zhao et al. 2010; Chen et al. 2013; Xu and Wang 2014). Moreover, studies on H and O isotopes show that the ore-forming fluid was dominated by magmatic water, with a mixture of construction water and meteoric water (Mu et al. 1981). Hence, S-type granite provides important materials and fluid sources for tungsten mineralization (Wang et al. 2008, 2010; Song et al. 2011; Wei et al. 2011; Xu and Wang 2014; Zhu et al. 2014; Huang et al. 2015; Wu et al. 2016). In addition, the granite provides the necessary energy for the extraction, migration, and precipitation of ore-forming materials (Barnes 2000). Hence, the inference of concealed granite is of great significance to MPM. According to the physical parameters of the rocks in this region, the density of granite is approximately $2.60 \times 10^3 \text{ kg/m}^3$, which is generally lower than that of the wall rock, while the magnetic susceptibility of granite is not significantly different from that of the wall rock

(Rao et al. 2006). The Bouguer gravity anomaly can be employed to infer the concealed granite in this region. Chen et al. (2014a) applied the singularity mapping technique based on the density/concentration-area power-law model, which can act as a high-pass filter for extracting gravity anomalies regardless of the background value, to detect the edges of the gravity sources in the Nanling region. Since the regular singularity analysis method cannot process maps with negative values, a modified algorithm to calculate the singularity index was proposed (Wang and Zuo 2015; Zuo et al. 2015). In this paper, we adopt the modified algorithm to infer concealed granite in the NMB (Fig. 2). The singularity index map of the Bouguer gravity anomaly is classified into 10 categories to generate a discretized map (Fig. 2a). In contrast, a continuous evidential map of the singularity index is also generated via logistic transformation (Fig. 2b).

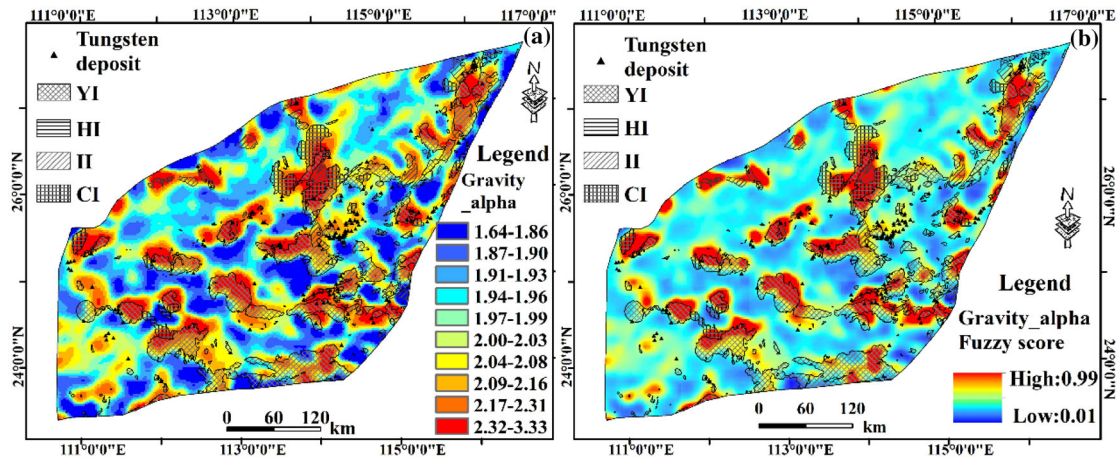


Figure 2. Evidential maps of derived from Bouguer gravity anomaly (YI denotes the Yanshanian intrusive rocks; HI denotes the Hercynian intrusive rocks; II denotes the Indosinian intrusive rocks; CI denotes the Caledonian intrusive rocks): (a) singularity index map of Bouguer gravity anomaly. (b) Fuzzy score of the singularity index of Bouguer gravity anomaly.

Faults

The NMB has experienced tectonic movement from the Caledonian to the Himalayan, forming a series of complex fold-fault structures (Shu et al. 2004; Shu 2012). Among the structures, faults cutting different circles (e.g., the NE-, NW-, and EW-striking deep faults) not only provide important channels for magma migration but also control the distribution of granite (Zhou 2007). In addition, the secondary faults of different strikes provide important channels for the migration of ore-forming fluids (Wei et al. 2004). Hence, fault systems are the major ore- and rock controlling structures in this region, forming a mineralization network (Zhai et al. 2002; Zhai 2003a, 2003b; Zhou 2007). Pei et al. (1999) proposed a “line-row-cluster” ore-controlling model consisting of EW- or NS-striking rows, NE- or NW-striking lines and intersection points of the lines and rows, which coincide with deep tectonic processes in this region. In this paper, faults of different strikes are selected as evidential data representing pathways for ore-forming materials and fluids. A multi-ring buffer with an interval of 2 km is employed to distinguish faults with different strikes. Then, discretized evidential maps of the distances to faults with different strikes are generated (Fig. 3a, b, c, and d). In contrast, a data-driven logistic-based transformation is conducted on the distance to faults with different strikes to generate continuous evidential maps. In this transformation approach, the minimum value in each map is assigned the maxi-

imum weight (e.g., 0.99), while the maximum value in each map is assigned the minimum weight (e.g., 0.01) (Fig. 3e, f, g, and h).

Geochemical Anomalies

In fact, most of the strata in the NMB are enriched in ore-forming elements such as tungsten and tin, which has laid a good foundation for tungsten polymetallic mineralization in the region (Yu et al. 1987). Identifying the geochemical anomalies associated with tungsten mineralization is critical for mineral exploration. Factor analysis is a widely used technique to explain the variation in a multivariate geochemical dataset by a few factors containing crucial information regarding the geochemical processes (Tripathi 1979; Reimann et al. 2002). Since geochemical data are compositional data that may produce data closure problems when using traditional statistical methods, several transformations (e.g., additive log-ratio, centered log-ratio, and isometric log-ratio) are proposed to preprocess the data prior to data analysis (Aitchison 1986; Egozcue et al. 2003). Many studies that discuss the closure problem of stream sediment geochemical data in this region have been performed (Liu et al. 2016). In contrast to previous works, all 39 geochemical elements used in this paper were selected to avoid possible information losses (Zuo and Xiong 2018). Subsequently, the centered log-ratio (CLR) transformation is described by Eq. (5), which involves the transforma-

tion from the simplex sample space to the D -dimensional real space, was performed to preprocess the geochemical data.

$$y = (y_1, y_2, \dots, y_D) \\ = \left(\ln \frac{x_1}{\sqrt[n]{\prod_{i=1}^D x_i}}, \ln \frac{x_2}{\sqrt[n]{\prod_{i=1}^D x_i}}, \dots, \ln \frac{x_D}{\sqrt[n]{\prod_{i=1}^D x_i}} \right) \quad (5)$$

where $x = (x_1, x_2, \dots, x_D)^T$ is a compositional vector.

Filzmoser et al. (2009) proposed a robust factor analysis method for compositional data, which not only overcomes the singularity problem in factor analysis when using a centered log-ratio transformation but also provides a meaningful biplot for interpretation of the results. Robust factor analysis of the compositional data is applied in this paper to obtain the geochemical relationship associated with tungsten polymetallic mineralization (Fig. 4a and b).

It is essential to recognize the geochemical association anomalies related to mineralization to diminish the impact of the high background values. With respect to anomaly recognition, fractal models based on the nonlinear theory, such as the concentration-area (C-A) model (Cheng et al. 1994), spectrum-area (S-A) model (Cheng 1999) and local singularity analysis (LSA) (Cheng 2007a, b), have shown advantages over traditional methods in separating anomalies from the background in many practical cases (Carranza 2009b; Zuo et al. 2012; Zuo and Wang 2016). Because anomalous areas delineated by the S-A model may have no direct correspondence to potential sources (Zuo et al. 2016), LSA is employed to detect geochemical association anomalies. These anomalies are then classified into 10 classes to obtain a discretized evidential map (Fig. 4c), and a logistic transformation is employed to generate a continuous evidential map (Fig. 4d).

Ore-Host Structures

According to the tungsten polymetallic model, the three major types of deposits in this region occur near the contact zones between S-type granite and wall rock. Hence, the contact zones are favorable ore-host structures, and multi-ring buffers of the contact zones between granite and the wall rock are selected as a predictive map indicating the favorable

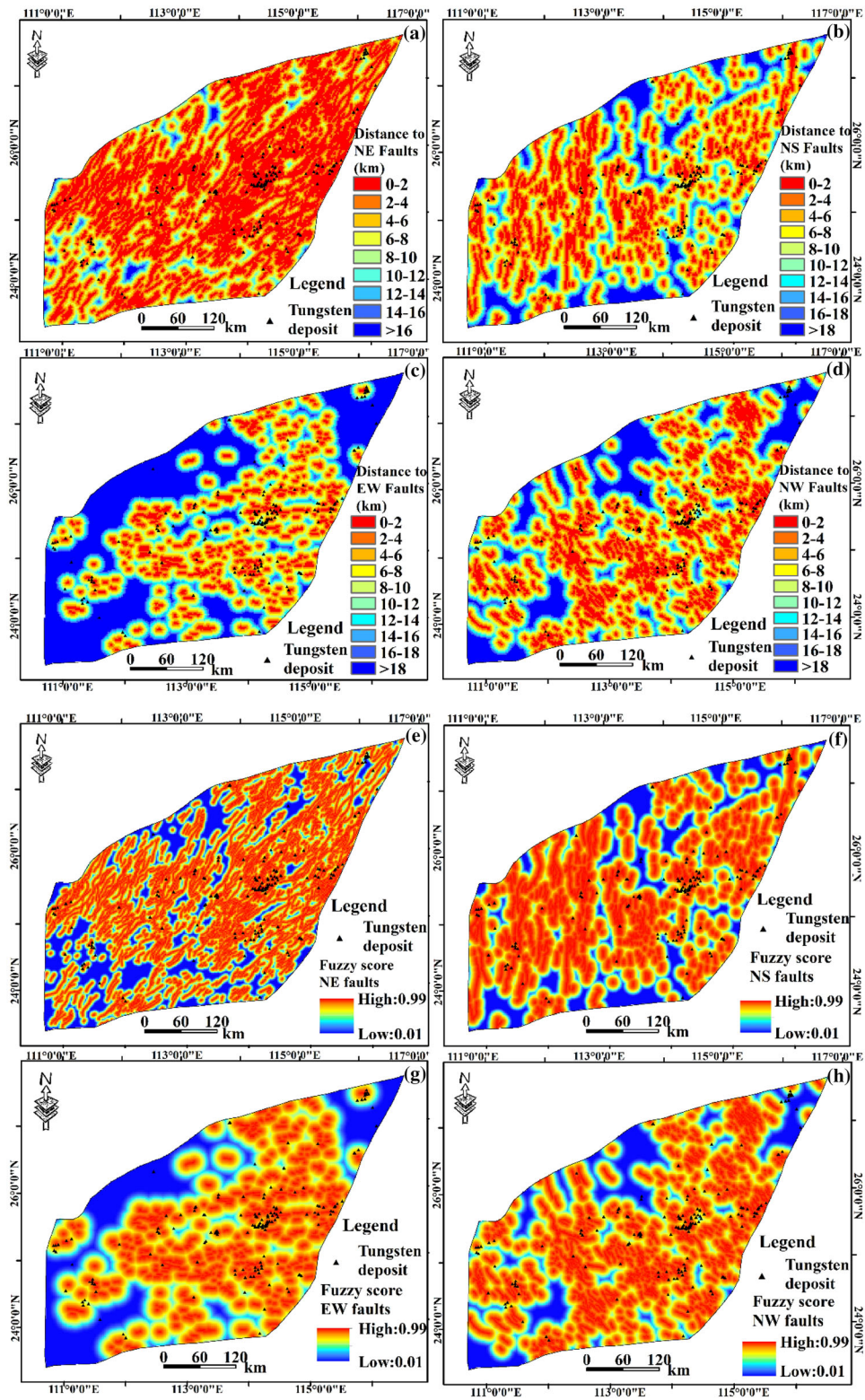
locations for mineral deposits. A multi-ring buffer with an interval of 2 km is employed for contact. Then, a discretized evidential map of the distance to the contact is generated (Fig. 5a). In contrast, logistic transformation is applied to the distance to the contact to generate a continuous evidential map (Fig. 5b). In this transformation, the minimum value is assigned the maximum weight (e.g., 0.99), while the maximum value is assigned the minimum weight (e.g., 0.01).

RESULTS AND DISCUSSION

Spatial Distribution of Deposits

Most applications of Fry analysis are based on regularly shaped study areas (Carranza 2009a; Carranza et al. 2009; Salati et al. 2013; Haddad-Martim et al. 2017; Parsa et al. 2018). Little work has discussed the shape of the study area, which may influence the Fry analysis results for the subjective determination of the scope of the study area. Hence, in this paper, the applicability of Fry analysis to an irregularly shaped study area is discussed at the outset. A series of randomly distributed points (S1) is generated in a regularly shaped study area (Fig. 6a). For comparison, an irregular shape is also generated inside the regular shape, and the points located in the irregular shape comprise S2 (Fig. 6d). Next, Fry analysis is performed for the S1 points, and the S2 points are analyzed separately, while rose diagrams of the different series of points are also generated. The distribution of the S1 points shows no dominant directions at the local scale, resulting in a random distribution (Fig. 6b), but there is an EW-striking distribution at the regional scale, which is consistent with the shape of the study area (Fig. 6c). In contrast to the Fry analysis of the regularly shaped study area, the Fry analysis of the S2 points indicates no notable dominant directions at the local scale (Fig. 6e), but there is a NE-striking distribution at the regional scale. These results indicate that the shape of the study area should not be ignored in the application of Fry analysis since the shape may influence the interpretation of the distribution of mineral deposits at the regional scale.

On this basis, all of the tungsten polymetallic deposits in the NMB are subjected to Fry analysis, and rose diagrams are plotted (Fig. 7b). The spatial distribution of the tungsten polymetallic deposits



◀ **Figure 3.** Evidential maps of distance to faults of different strikes: (a) distance to NE-striking faults; (b) distance to NS-striking faults; (c) distance to EW-striking faults; (d) distance to NW-striking faults; (e) fuzzy score of distance to NE-striking faults; (f) fuzzy score of distance to NS-striking faults; (g) fuzzy score of distance to EW-striking faults; (h) fuzzy score of distance to NW-striking faults.

clearly shows that the distribution of mineral deposits is controlled by faults with different strikes (Fig. 7c), and the EW-striking and NE-striking faults dominate the distribution of tungsten deposits at a regional scale (Fig. 7d). The results of the Fry analysis are consistent with the tectonic stress of this region since the Yanshanian period. Since the Yanshanian, influenced by the subduction of the Pacific Plate, this region has formed a series of NNE-striking faults with sinistral strike-slip characteristics. Additionally, a set of NEE-striking faults with dextral strike-slip characteristics and NW-

striking faults with sinistral strike-slip characteristics was formed at the same time (Liang et al. 2016) (Fig. 7a). Among these results, the NNE-striking and NEE-striking faults control the distribution of the Yanshanian granite. Although the NE-striking distribution of mineral deposits in the NMB at the regional scale may be due to the shape of the study area, the outline of the study area is based on deep faults. Therefore, the NE-striking distribution of the mineral deposits still reflects the ore-controlling behavior of the NE-striking faults. At the local scale, the faults of different strikes provide abundant migration channels and convergence zones for the ore-forming fluids. Hence, the mineral deposits in this region exhibit multidirectional behaviors at the local scale.

With the aid of GIS, mineral deposits can be converted into a series of grids with different cell sizes. The box-counting analysis is implemented, and a log-log plot of the cell size vs. the number of cells

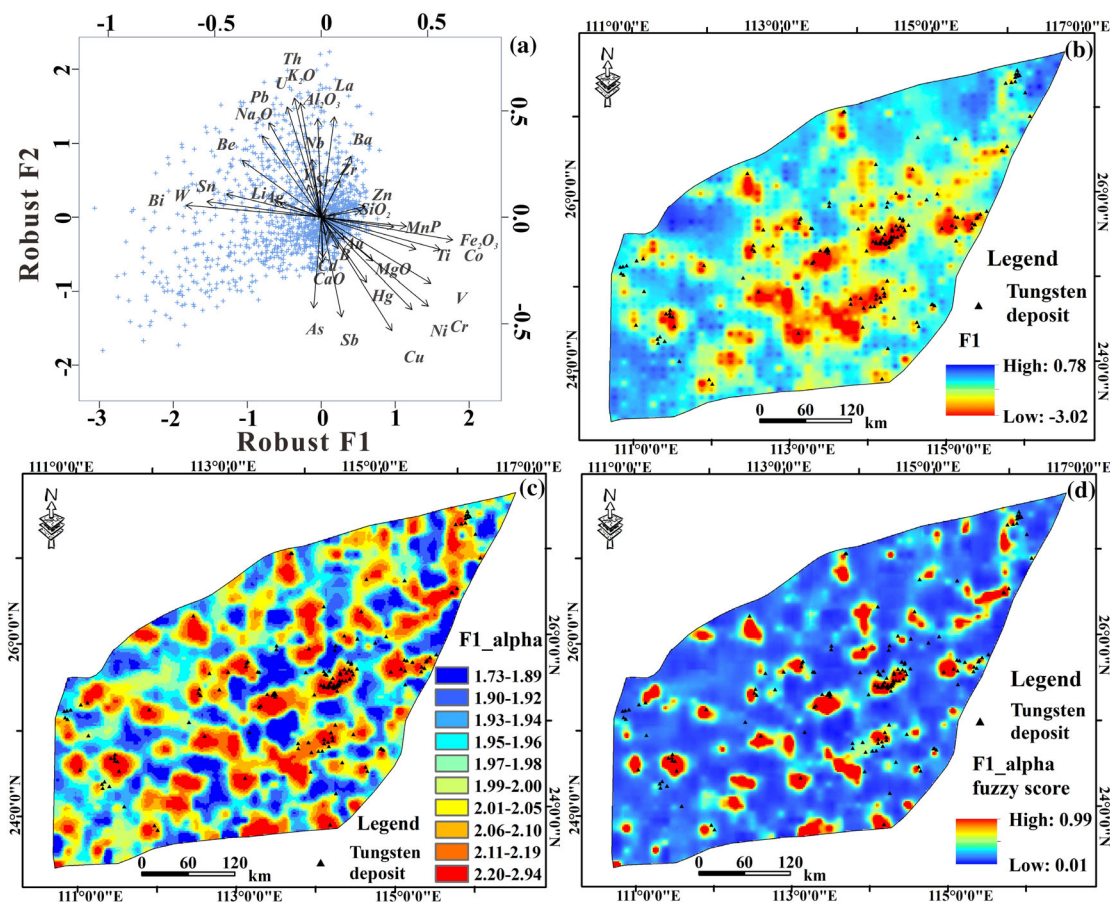


Figure 4. Geochemical association related to tungsten polymetallic mineralization: (a) biplot of 39 geochemical elements; (b) Geochemical association related to tungsten polymetallic deposits (F1); (c) singularity index map of F1 (F1_alpha); (d) fuzzy score of singularity map of F1 (F1_alpha).

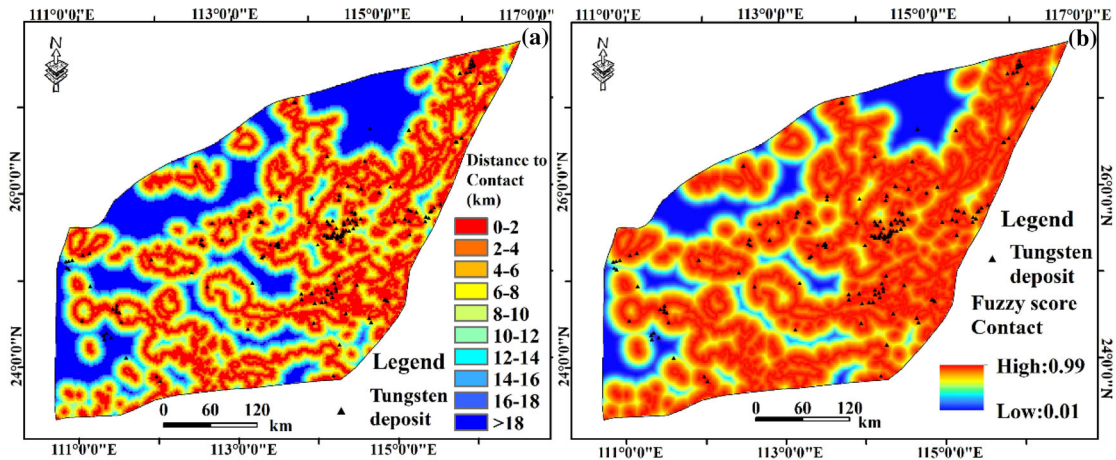


Figure 5. Evidential maps of distance to contact: (a) distance to contact; (b) fuzzy score of distance to contact.

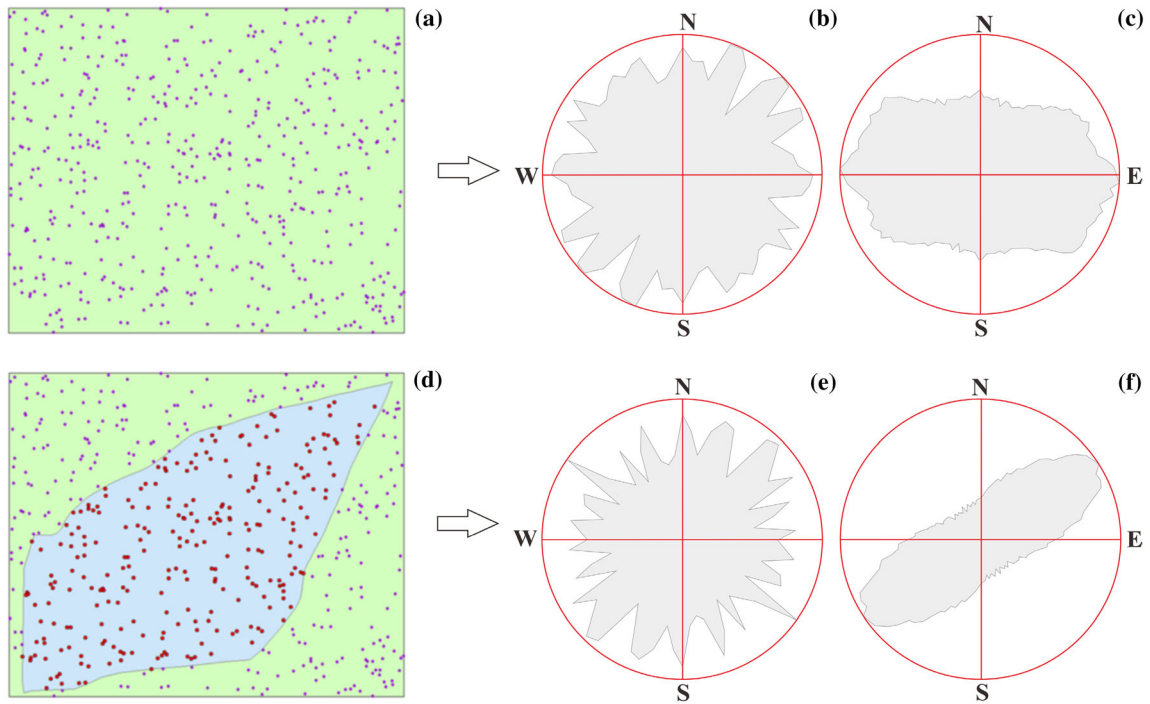


Figure 6. Fry analysis to differently shaped areas: (a) randomly distributed points S1 in a regularly shaped area; (b) Fry analysis of points (S1) at the local scale; (c) Fry analysis of points (S1) at the regional scale; (d) randomly distributed points (S2) in an irregularly shaped area; (e) Fry analysis of points (S2) at the local scale; (f) Fry analysis of points (S2) at the regional scale.

containing a deposit is generated. It is clear that the scatter in the log-log plot can be fitted by two straight lines: a left straight line with a box-dimension of 0.44 and a right straight line with a box-dimension of 1.02 (Fig. 8a). The threshold value is approximately 35 km; i.e., the tungsten mineraliza-

tion may be controlled by local geological factors at scales smaller than 35 km, while the scales greater than 35 km may reflect regional geological processes.

The radial density is calculated based on various circles with different radii centered at the min-

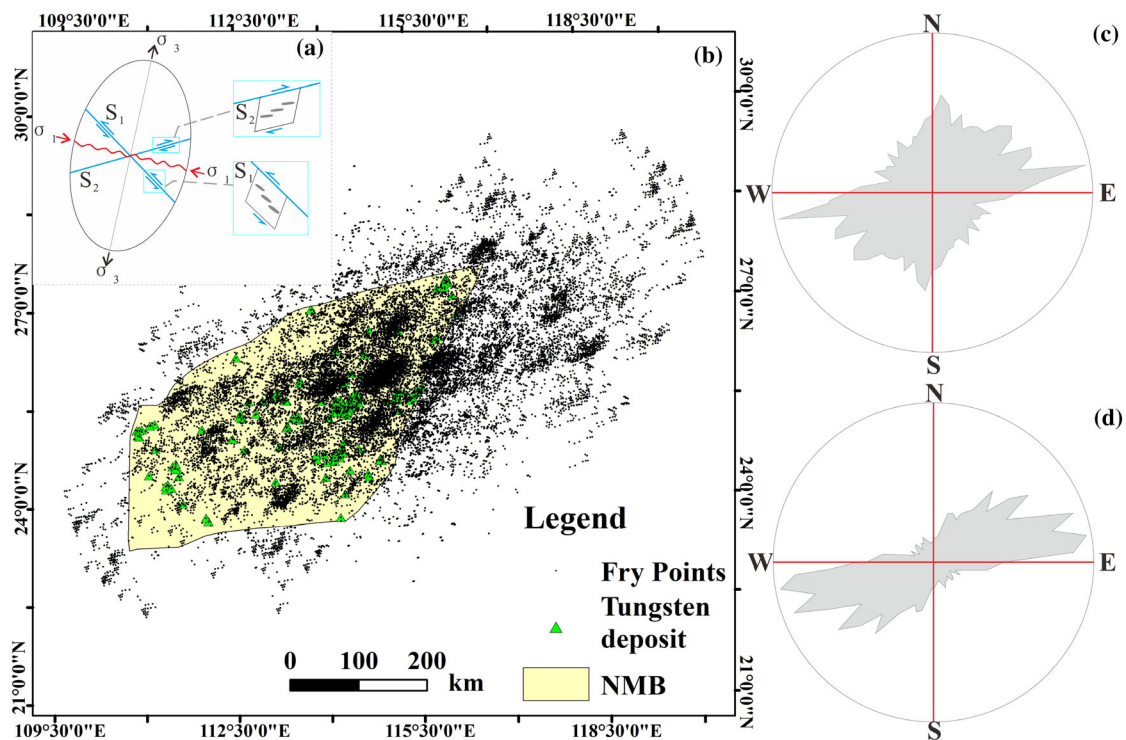


Figure 7. Fry analysis of tungsten polymetallic deposits in the NMB: (a) regional stress in the NMB since the Yanshanian (revised from Liang et al. 2016); (b) Fry points generated from Fry analysis; (c) Rose diagram for dominant directions of tungsten polymetallic deposits at the local scale; (d) Rose diagram for dominant directions of tungsten polymetallic deposits at the regional scale.

eral deposits. Then, a log–log plot of radius vs. the corresponding radial density is implemented (Fig. 8b). It is obvious that the scatter in the log–log plot can be fitted by two straight lines: a left straight line with a fractal dimension of 0.65 and a right straight line with a fractal dimension of 1.53. The threshold value is approximately 35 km, which is consistent with the results of box-counting fractal analysis.

For comparison, the nearest-neighbor distances of every two deposits in this region are calculated. More than 90% of the nearest-neighbor distances are smaller than 30 km. Moreover, the 95th percentile is a commonly used threshold to determine the lower limit of geochemical anomalies (Reimann et al. 2008; Moeini and Torab 2017; Filzmoser et al. 2018). In this case, the 95th percentile of the nearest neighbor distances is approximately 35 km. That is, the probability of finding a deposit 35 km away from a known deposit is low. Hence, the threshold value in the log–log plot of the cell size vs. the number of cells containing deposits can be regarded as the

optimal distance when selecting the non-deposit locations.

In addition, according to previous studies, most of the tungsten polymetallic deposits are within 1 km of the contact zone between S-type granite and wall-rock (Liu et al. 2014, 2015). Therefore, the non-deposit locations in this paper are defined by considering the following criteria: (1) the non-deposit locations should be at least 35 km away from known deposits; (2) the non-deposit locations should be at least 1 km away from the contact between S-type granite and wall rock; and (3) the non-deposit locations should be distributed randomly.

Mineral Prospectivity Mapping

The under-sampling method, which is widely used in MPM, is applied to generate a training dataset with 154 deposit locations and 154 non-deposit locations. There are two parameters in the RF algorithm: the number of predictors (m) randomly sampled at each split and the number of trees (k) to

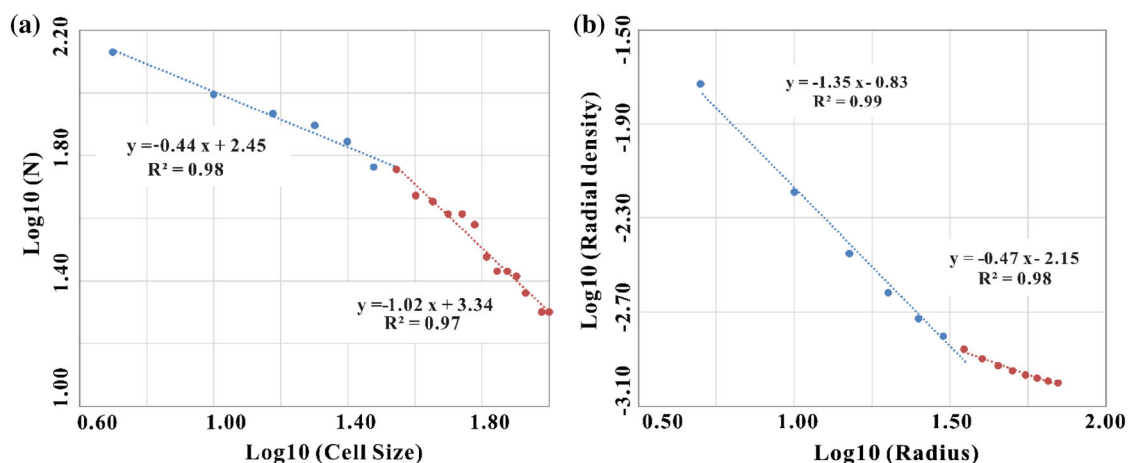


Figure 8. Fractal analysis of spatial distribution of mineral deposits in the NMB: (a) log–log plot of cell size vs. the number of cells containing deposits; (b) log–log plot of radial density vs. the radius to the known deposit.

be used. The value m , a fraction of the total number of predictors, is determined using the *tuneRF* function in the CRAN package *randomForest*, which calculates the optimum value of m that minimizes the OBB error (Liaw and Wiener 2002). There is no fixed optimal value of parameter k . A practical way to determine the optimal value is by using the value of k when the classification error tends to become stable (Carranza and Laborte 2015b). In this study, five under-sampled discretized datasets (DUS, DUS1, DUS2, DUS3, and DUS4) and five continuous weighted datasets (CUS, CUS1, CUS2, CUS3, and CUS4) are generated. According to the plot of the OBB error vs. the number of trees, the OBB errors of both the discretized training datasets and continuous weighted training datasets tend to become stable when the number of trees is greater than 500 (Fig. 9a and b). Hence, the optimal value of k is 500.

In contrast, 1540 randomly distributed non-deposit locations are chosen based on the selection criteria for non-deposit locations. Hence, the discretized training dataset (DIM) and the continuous weighted training dataset (CIM) compose the imbalanced training dataset. In addition, the ratio of deposit locations to non-deposit locations in both imbalanced datasets is 1:10.

Subsequently, the SMOTE approach is adopted to generate balanced datasets from the imbalanced datasets DIM and CIM. In this procedure, the deposit locations are over-sampled in the feature space. At the same time, the non-deposit locations

are under-sampled by randomly removing samples until the deposit locations become closer in number to the non-deposit locations. By iteratively under-sampling and over-sampling the non-deposit locations and deposit locations, the balanced datasets are used to train the RF model and generate classification errors. The fraction of over-sampling locations with stable classification errors generated from the corresponding training datasets is regarded as the optimal value (Hariharan et al. 2017). The classification error obtained by RF models with discretized training datasets is the smallest when the over-sampling ratio reaches 1000% (Fig. 10a), while the classification error obtained by RF models with continuous weighted training datasets is the smallest when the over-sampling ratio is 900% (Fig. 10b). In this case, a balanced discretized training dataset with 1694 deposit locations and 1694 non-deposit locations is generated (DSMOTE10), and balanced continuous weighted training datasets with 1540 deposit locations and 1538 non-deposit locations are generated (CSMOTE9).

When compared to other ML algorithms (e.g., neural networks), the RF algorithm can determine the relative importance of predictive variables, which provides insights into the ore-controlling factors as well as guidance for mineral exploration. In this paper, the RF models are trained using different training datasets, and the relative importance of the predictive variables of different prospectivity models is obtained (Fig. 11a and b). It is clear that the relative importance of the predictive variables is inde-

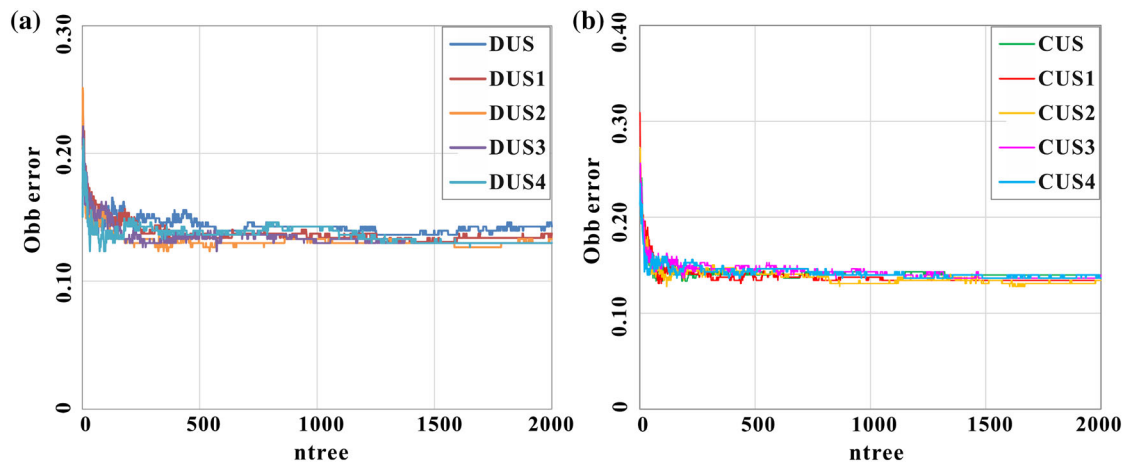


Figure 9. Plot of OBB error vs. number of trees used in the RF algorithm: (a) OBB error generated from different discretized training datasets; (b) OBB error generated from different continuous training datasets.

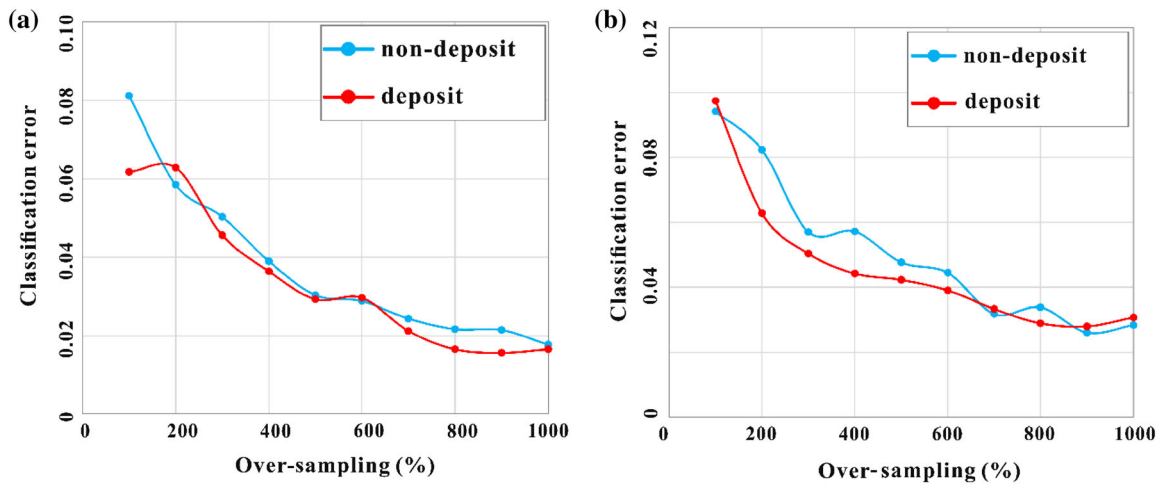


Figure 10. Plots of different over-sampling rates vs. classification errors: (a) classification error of different RF models generated from discretized evidential maps; (b) classification error of different RF models generated from continuous evidential maps.

pendent of whether the predictive variables are discretized or continuous. However, the relative importance depends on how the training dataset is generated. Because a certain number of positive samples (e.g., deposit locations) is synthesized according to the features of the predictive variables in SMOTE, the importance of predictive variables may be difficult to explain. The number of samples synthesized by SMOTE may also result in uncertainty in the relative importance of the predictive variables. However, the relative importance of predictive variables in a prospectivity model trained by DUS (or CUS) data is similar to the relative

importance of the predictive variables in a prospectivity model trained by DIM (or CIM) data. It is understood that the geochemical association anomalies related to tungsten polymetallic deposits ($F1_{\alpha}$) are of vital importance to the mineralization in this region. Furthermore, the areas with high anomalies are spatially correlated with the areas with high prospectivity scores. This is consistent with the geological settings with high degrees of enrichment in W, Sn, and other elements, which establishes an important foundation for tungsten mineralization (Fig. 12). In addition, the study area is a hilly area covered by vegetation, with an ele-

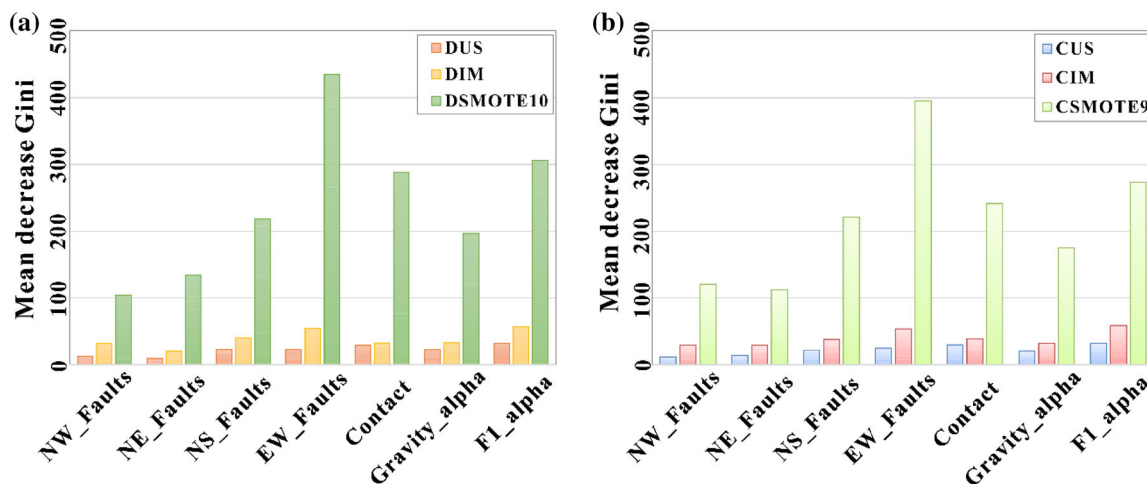


Figure 11. Relative importance of predictor variables obtained from different prospectivity models: (a) relative importance of discretized predictor variables using different prospectivity models; (b) relative importance of continuous predictor variables using different prospectivity models.

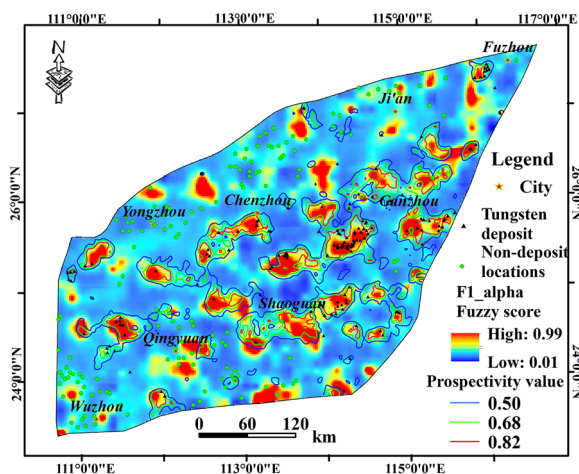


Figure 12. Prospectivity model with continuous evidential map F1_alpha (singularity index of F1) using the continuous balanced dataset CSMOTE9.

vation of more than 300 m. The terrain is deeply eroded, and the water system is well developed. Therefore, geochemical exploration may be a prior method for mineral exploration in this region.

Model Performance Evaluation

Based on the training datasets generated above, six prospectivity maps of the tungsten polymetallic deposits are generated using the RF algorithm. Subsequently, each prospectivity map is classified by the C-A model (Figs. 13a, c, and e, 14a, c, and e). It

is clear that most of the tungsten polymetallic deposits are located in areas with high prospectivity values (Figs. 13b, d, and f, 14b, d, and f).

The ROC curve and improved P-A plot are employed to evaluate the performance of the prospectivity models. The performance of the RF models trained by DSMOTE10 and DIM rank first with the largest AUC value of 0.9789, while the performance of the RF model trained by DUS ranks last with the smallest AUC value of 0.9452 (Fig. 15a).

The improved P-A plot is also employed to evaluate the performance of the prospectivity model. Every prospectivity map is classified by the C-A model to generate the improved P-A plots. From the improved P-A plots, the prediction rates of prospectivity models generated from both DSMOTE10 and DIM are 93%, while the prediction rate of the prospectivity model generated from DUS is 88%. At the same time, the prediction rates for the non-deposit locations of the prospectivity models generated from DUS, DIM, and DSMOTE8 are 36%, 41%, and 36%, respectively. Hence, the overall performance values (O_e) of the prospectivity models generated from DUS, DIM, and DSMOTE10 are 0.52, 0.52, and 0.57, respectively (Fig. 15b, c, d). These results indicate that the overall performance of the prospectivity model generated from DSMOTE10 is the best.

With respect to MPM using the RF algorithm with continuous evidential maps, the AUC value of the prospectivity model generated from CIM is 0.9891 and ranks first. The AUC value of the

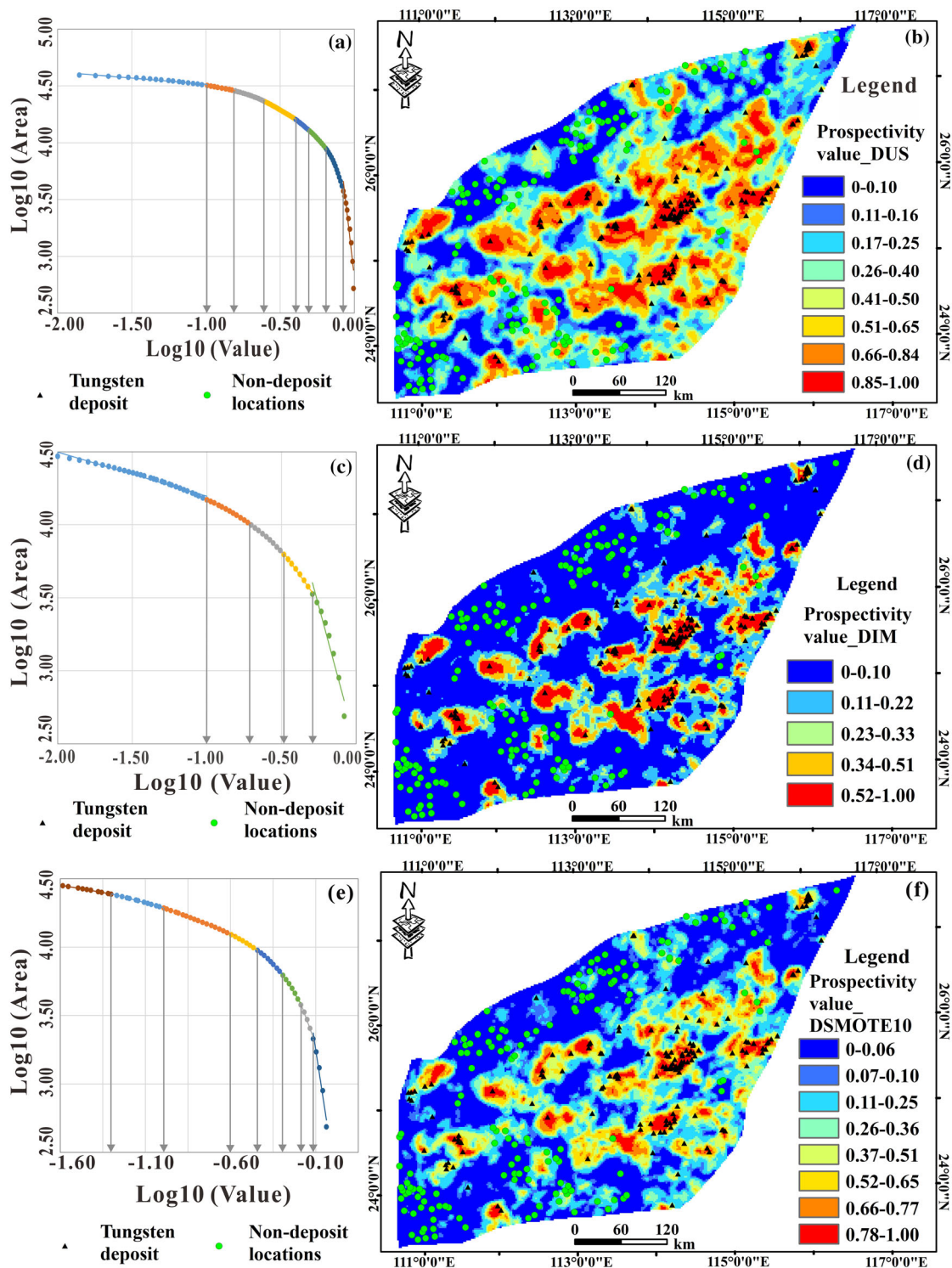


Figure 13. Prospectivity maps generated from different discretized evidential maps: (a) C-A model of prospectivity map generated from dataset DUS; (b) prospectivity map generated from dataset DUS; (c) C-A model of prospectivity map generated from dataset DIM; (d) Prospectivity map generated from dataset DIM; (e) C-A model of prospectivity map generated from dataset DSMOTE10; (f) Prospectivity map generated from dataset DSMOTE10.

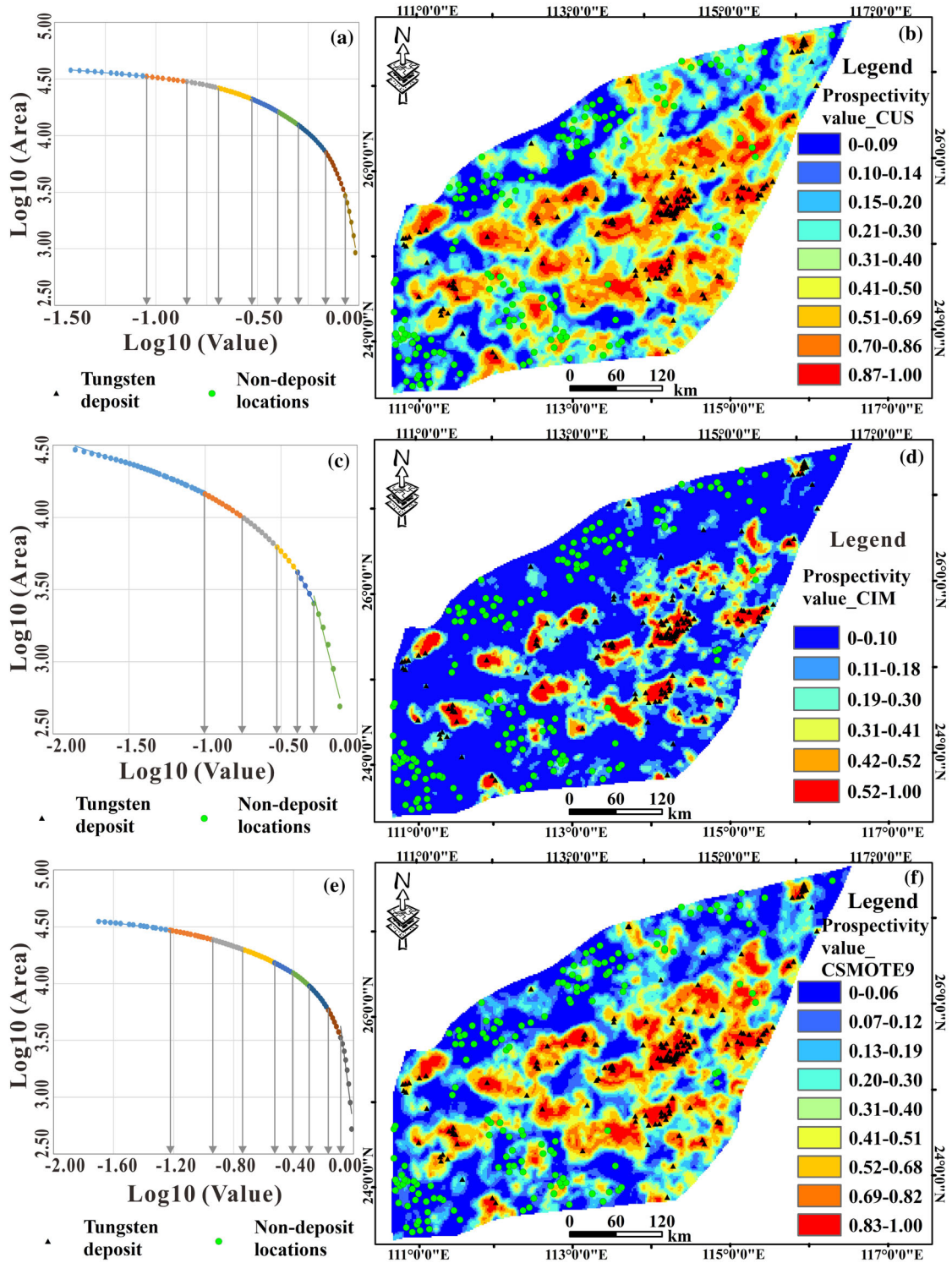


Figure 14. Prospectivity maps generated from different continuous evidential maps: (a) C-A model of prospectivity map generated from dataset CUS; (b) prospectivity map generated from dataset CUS; (c) C-A model of prospectivity map generated from dataset CIM; (d) prospectivity map generated from dataset CIM; (e) C-A model of prospectivity map generated from dataset CSMOTE9; (f) Prospectivity map generated from dataset CSMOTE9.

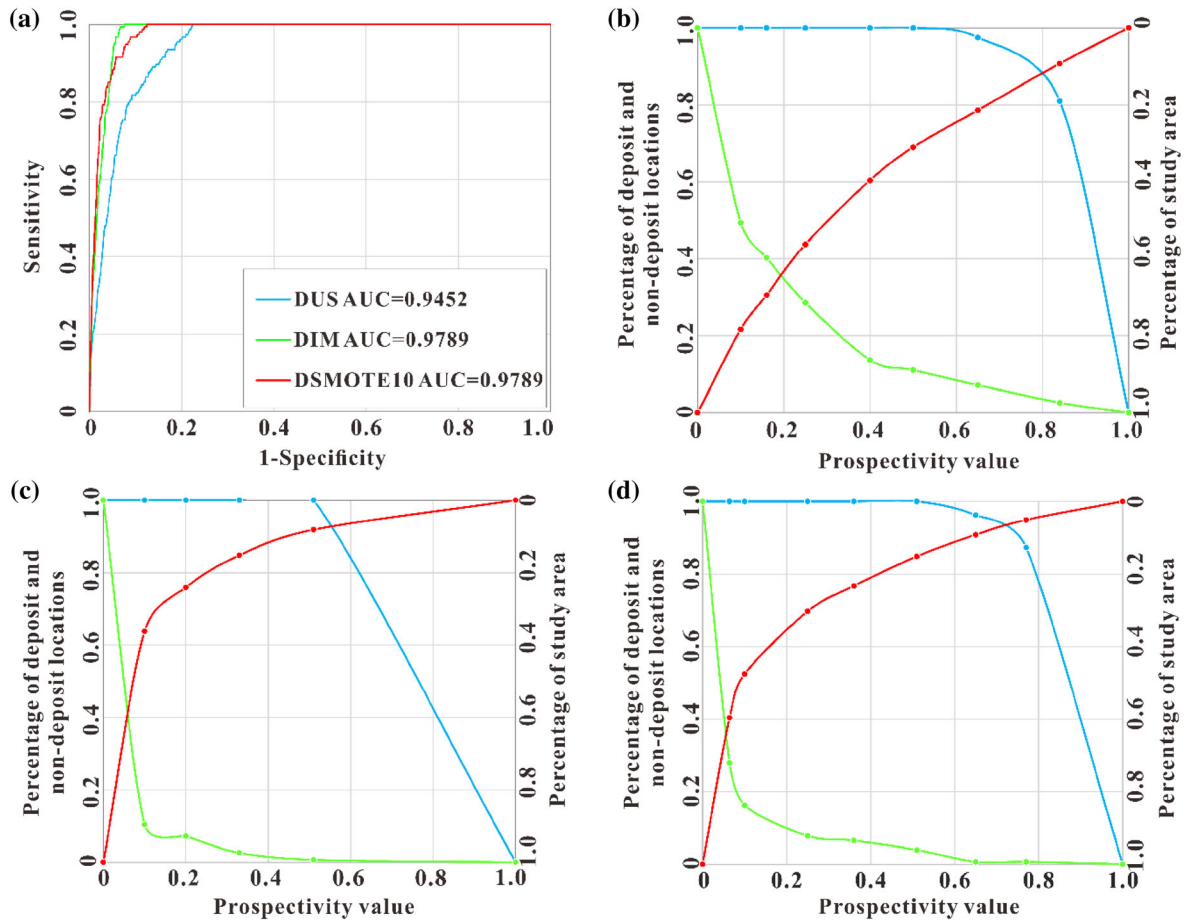


Figure 15. Performance of prospectivity models generated from discretized evidential maps: (a) ROC curves of prospectivity models generated from different datasets (DUS, DIM, and DSMOTE10); (b) improved P-A plot of prospectivity model generated from DUS; (c) improved P-A plot of prospectivity model generated from DIM; (d) improved P-A plot of prospectivity model generated from DSMOTE10. The blue, green and red lines in (b), (c) and (d) denote prediction rate of deposits, prediction rate of non-deposit locations and percentages of the areas occupied by the corresponding prospectivity classes, respectively.

prospectivity model generated from CSMOTE9 is 0.9862, followed by the prospectivity model generated from CUS with an AUC value of 0.9546 (Fig. 16a). When using the improved P-A plot to evaluate the performance, the prospectivity model generated from CIM with a prediction rate of 94% ranks first, followed by the prospectivity model generated from CSMOTE9 with a prediction rate of 93%. The prospectivity model generated from CUS has a prediction rate of 90% for the deposit locations. The prediction rates for the non-deposit locations of the prospectivity models generated from CUS, CIM, and CSMOTE9 are 35%, 41%, and 34%, respectively. The overall performance values (O_e) of the prospectivity models generated from CUS, CIM, and CSMOTE9 are 0.55, 0.53, and 0.59, respectively (Fig. 16b, c, and d). These results indi-

cate that the overall performance of the prospectivity model generated from CSMOTE9 is the best, followed by the overall performance of the prospectivity model generated from CUS.

Additionally, the performance of the prospectivity models obtained from continuous evidential maps is obviously superior to the performance of the prospectivity models obtained from discretized evidential maps. One possible reason is that mapping the mineral resource prospectivity using continuous evidential maps can avoid the uncertainty and bias resulting from the discretization of the evidential maps. As a result, the prediction rate for the deposit locations is improved and the prediction rate for the non-deposit locations is reduced. It can also be observed that the AUC value of the prospectivity model generated from DIM (or CIM) is greater than

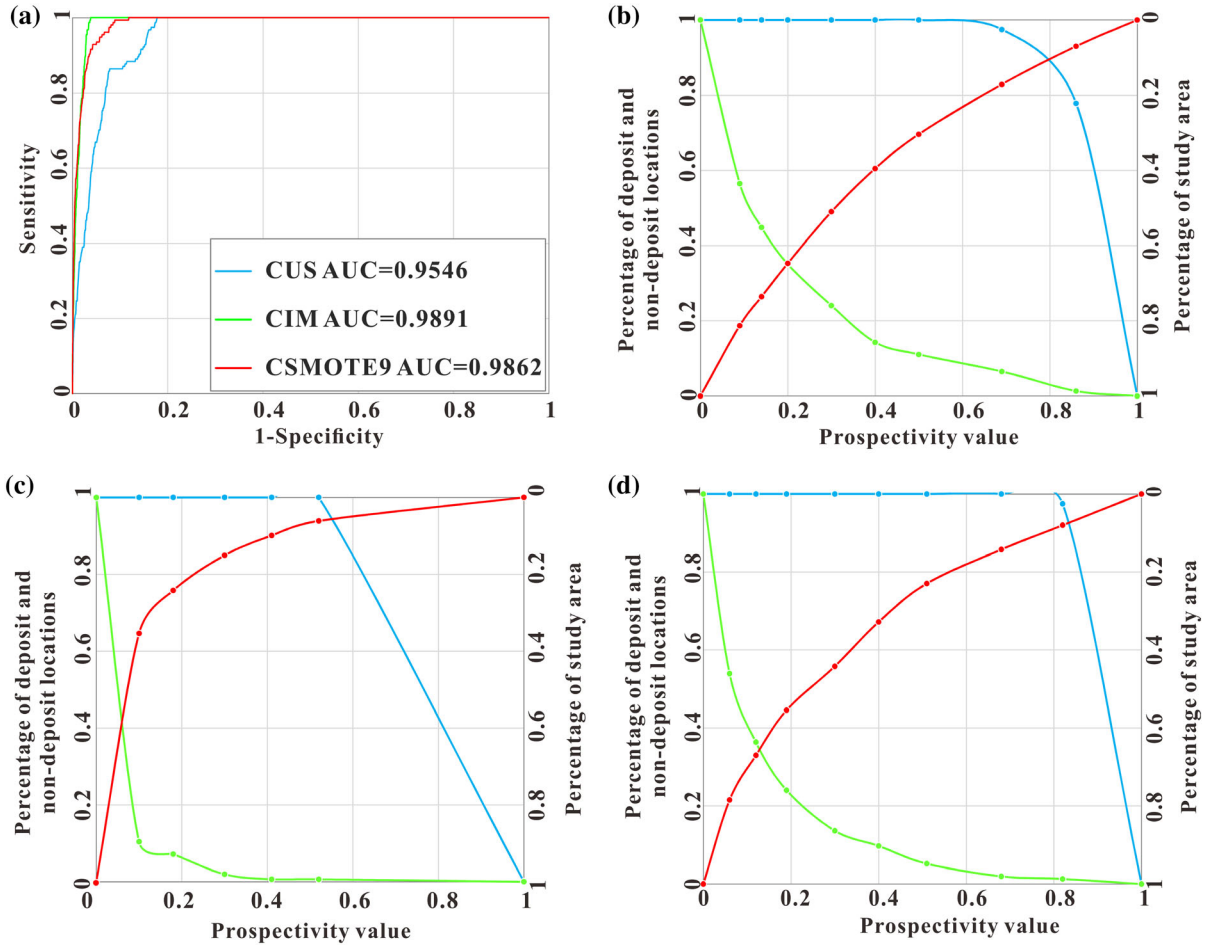


Figure 16. Performance of prospectivity models generated from continuous evidential maps: (a) ROC curves of prospectivity models generated from different datasets (CUS, CIM and CSMOTE9); (b) improved P-A plot of prospectivity model generated from CUS; (c) improved P-A plot of prospectivity model generated from CIM; (d) improved P-A plot of prospectivity model generated from CSMOTE9. The blue, green and red lines in (b), (c) and (d) denote prediction rate of deposits, prediction rate of non-deposit locations and percentages of the areas occupied by the corresponding prospectivity classes, respectively.

the AUC value of the prospectivity model generated from DUS (or CUS). Because the number of non-deposit locations is much larger than that of deposit locations, a large change in the number of FP (i.e., non-deposit locations that are falsely predicted as deposit locations) can lead to a small change in the FPR in the ROC curve of the prospectivity model trained by an imbalanced dataset when compared to the prospectivity model trained by a balanced dataset generated from under-sampling. This result indicates that the performance of a prospectivity model may be optimistically evaluated when using the ROC curve (Davis and Goadrich 2006; He and Garcia 2010). Additionally, the improved P-A plot may be more reliable in evaluating the performance of prospectivity models than the ROC curve. The

improved P-A plot is also superior to the original P-A plot for its ability to evaluate the correlation between the prospectivity models and non-deposit locations. Furthermore, the performance of the prospectivity models trained by datasets that are generated by SMOTE is the best, indicating that SMOTE can be a useful tool to reduce the data imbalance and improve the performance of prospectivity models.

According to the discussion above, the performance of the prospectivity model trained by CSMOTE9 is the best. From this predictive result (Fig. 17), four potential areas with high exploration degrees are delineated, including Qitianling–Qianlishan (A1), Dayu–Chongyi–Shangyou (A2), Yudu–Huichang (A3) and Shixing–Quannan (A4). Several

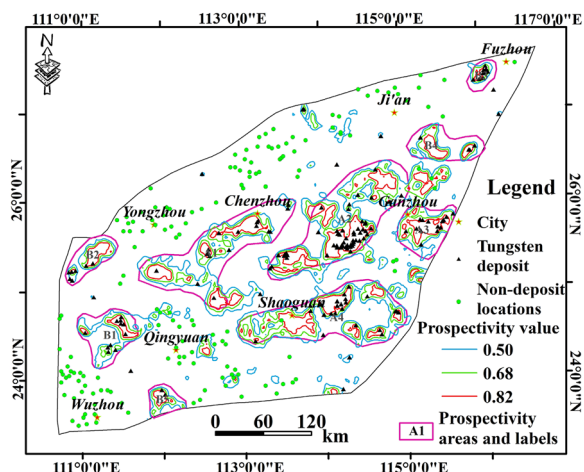


Figure 17. Prospectivity map for tungsten polymetallic deposits in the NMB.

large tungsten polymetallic deposits are located in the four potential areas, including Shizhuyuan, Yaogangxian, Xintianling, Xihuashan, Piaotang, Pangushan, Huangsha, Dajishan, and Kuimeishan. The four potential areas have a large number of geochemical anomalies and are located at the confluence of deep multidirectional faults. Moreover, large areas of granite are exposed in these areas. Therefore, the metallogenic geological conditions in these four potential areas are superior. In addition, although there is little granite exposed in some areas, including Zhongshan–Hexian (B1), Jiangyong–Guanxian (B2), Le’an–Yihuang (B3), Ningdu–Xingguo (B4), and Fengkai–Huaiji (B5), most of these areas are located in regions with prolific geochemical anomalies, in which a certain number of tungsten polymetallic deposits occur. Hence, these areas also have good prospecting potential.

CONCLUSIONS

The following conclusions can be drawn from this paper:

1. The shape of the study area should not be ignored when studying the spatial distribution of mineral deposits using Fry analysis. Fry analysis indicates that the tungsten polymetallic deposits in this region display EW-striking and NE-striking distributions at a regional scale, which is consistent with the

regional tectonic stress since the Yanshanian. At the local scale, the distribution of tungsten polymetallic deposits is mainly controlled by structures striking in multiple directions.

2. Fractal analysis shows that the tungsten polymetallic deposits in the NMB satisfy a multifractal distribution. The intersection point in the log–log plot can be a potential measure to determine the optimal distance from known deposits, which may be useful for the selection of non-deposit locations in MPM.
3. The application of data-driven logistic transformation to generate continuous evidential maps in MPM using the RF algorithm can avoid the uncertainty and information loss resulting from the discretization of evidential maps, which helps to improve the performance of prospectivity models.
4. When evaluating prospectivity models, the improved P-A plot is superior to the ROC curve because the ROC curve neglects the occupied area, which is critical for mineral exploration and may provide an overly optimistic performance with imbalanced data. The SMOTE approach can reduce the data imbalance and improve the performance of prospectivity models. However, in the SMOTE approach, several deposit locations are synthesized according to the features of the predictors, which may lead to uncertainty in the explanation of the predictor relative importance. Further testing of the improved P-A plot and SMOTE approach is warranted in MPM.
5. The predictor importance shows that geochemical association anomalies contribute the most to mineralization and that areas with many geochemical anomalies are highly correlated spatially with areas having high prospectivity values. This result indicates that geochemical exploration may be a prior method for tungsten deposit exploration in this region.
6. According to the predictive results, four areas with high exploration potential and five moderate-potential areas are delineated. These results indicate good future prospecting for tungsten polymetallic deposits in this region.

ACKNOWLEDGMENTS

The authors are grateful to Prof. John Carranza and Prof. Renguang Zuo for handling this paper. We also thank the two anonymous reviewers for their comments to improve this manuscript. This research benefited from joint financial support from the National Natural Science Foundation of China (No. 41672328) and the Geological Survey Projects of the China Geological Survey (No. DD20179385).

REFERENCES

- Agterberg, F. P. (1989). Systematic approach to dealing with uncertainty of geoscience information in mineral exploration. In *Proceedings of the 21st APCOM symposium* (pp. 165–178).
- Agterberg, F. P., & Bonham-Carter, G. F. (2005). Measuring the performance of mineral-potential maps. *Natural Resources Research*, 14(1), 1–17.
- Aitchison, J. (1986). The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, 44(2), 139–177.
- Almasi, A., Yousefi, M., & Carranza, E. J. M. (2017). Prospectivity analysis of orogenic gold deposits in Saqez-Sardasht Goldfield, Zagros Orogen, Iran. *Ore Geology Reviews*, 91, 1066–1080.
- Bárdossy, G., & Fodor, J. (2001). Traditional and new ways to handle uncertainty in geology. *Natural Resources Research*, 10(3), 179–187.
- Bárdossy, G., & Fodor, J. (2004). *Evaluation of uncertainties and risks in geology*. Berlin: Springer.
- Barnes, H. L. (2000). Energetics of hydrothermal ore deposition. *International Geology Review*, 42(3), 224–231.
- Bishop, C. M. (2007). *Machine learning and pattern recognition*. New York: Springer.
- Bonham-Carter, G. F. (1994). *Geographic information systems for geoscientists: Modelling with GIS*. Oxford: Pergamon Press.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. London: Chapman & Hall.
- Breslow, N. E., & Cain, K. C. (1988). Logistic regression for two-stage case-control data. *Biometrika*, 75(1), 11–20.
- Carlson, C. A. (1991). Spatial distribution of ore deposits. *Geology*, 19(2), 111–114.
- Carranza, E. J. M. (2009a). Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features. *Ore Geology Reviews*, 35(3–4), 383–400.
- Carranza, E. J. M. (2009b). Geochemical anomaly and mineral prospectivity mapping in GIS. In M. Hale (Ed.), *Handbook of exploration and environmental geochemistry* (Vol. 11, pp. 3–351). Amsterdam: Elsevier.
- Carranza, E. J. M. (2009c). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. *Computers & Geosciences*, 35(10), 2032–2046.
- Carranza, E. J. M., & Laborte, A. G. (2015a). Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of random forests algorithm. *Ore Geology Reviews*, 71, 777–787.
- Carranza, E. J. M., & Laborte, A. G. (2015b). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences*, 74, 60–70.
- Carranza, E. J. M., & Laborte, A. G. (2016). Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes island (Philippines). *Natural Resources Research*, 25(1), 35–50.
- Carranza, E. J. M., Owusu, E. A., & Hale, M. (2009). Mapping of prospectivity and estimation of number of undiscovered prospects for lode gold, southwestern Ashanti Belt, Ghana. *Mineralium Deposita*, 44(8), 915–938.
- Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook*. Boston, MA: Springer.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chawla, N. V., Japkowicz, N., & El-Duek, P. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1), 1–6.
- Chen, G., Cheng, Q., Zuo, R., Liu, T., & Xi, Y. (2014a). Identifying gravity anomalies caused by granitic intrusions in Nanling mineral district, China: a multifractal perspective. *Geophysical Prospecting*, 63(1), 1–15.
- Chen, J., Lu, J., Chen, W., Wang, R., Ma, D., Zhu, J., et al. (2008). W–Sn–Nb–Ta-bearing granites in the Nanling range and their relationship to metallogenesis. *Geological Journal of China Universities*, 14(4), 459–473.
- Chen, Y., Lu, L., & Li, X. (2014b). Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *Journal of Geochemical Exploration*, 140, 56–63.
- Chen, J., Wang, R., Zhu, J., Lu, J., & Ma, D. (2013). Multiple-aged granitoids and related tungsten-tin mineralization in the Nanling Range, South China. *Science China Earth Sciences*, 56(12), 2045–2055.
- Chen, Y., & Wu, W. (2017). Mapping mineral prospectivity by using one-class support vector machine to identify multivariate geological anomalies from digital geological survey data. *Australian Journal of Earth Sciences*, 64(5), 639–651.
- Cheng, Q. (1999). Spatial and scaling modelling for geochemical anomaly separation. *Journal of Geochemical Exploration*, 65(3), 175–194.
- Cheng, Q. (2006). Singularity-generalized self-similarity-fractal spectrum (3S) models. *Journal of Earth Science*, 31(3), 337–348.
- Cheng, Q. (2007a). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geology Reviews*, 32(1), 314–324.
- Cheng, Q. (2007b). Singular mineralization processes and mineral resources quantitative prediction: New theories and methods. *Earth Science Frontiers*, 14(5), 44–55.
- Cheng, Q. (2008a). Non-linear theory and power-law models for information integration and mineral resources quantitative assessments. *Mathematical Geosciences*, 40(5), 503–532.
- Cheng, Q. (2008b). Singularity of mineralization and multifractal distribution of mineral deposits. *Bulletin of Mineralogy, Petrology and Geochemistry*, 27(3), 298–305.
- Cheng, Q. (2015). BoostWofE: A new sequential weights of evidence model reducing the effect of conditional dependency. *Mathematical Geosciences*, 47(5), 591–621.
- Cheng, Q., & Agterberg, F. P. (1996). Multifractal modeling and spatial statistics. *Mathematical Geosciences*, 28(1), 1–16.

- Cheng, Q., Agterberg, F. P., & Ballantyne, S. B. (1994). The separation of geochemical anomalies from background by fractal methods. *Journal of Geochemical Exploration*, 51(2), 109–130.
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd international conference on machine learning (ICML 2006)* (pp. 233–240).
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279–300.
- Fabbri, A. G., & Chung, C. J. (2008). On blind tests and spatial prediction models. *Natural Resources Research*, 17(2), 107–118.
- Filzmoser, P., Hron, K., Reimann, C., & Garrett, R. (2009). Robust factor analysis for compositional data. *Computers & Geosciences*, 35(9), 1854–1861.
- Filzmoser, P., Hron, K., & Templ, M. (2018). *Applied compositional data analysis (with worked example in R)*. Berlin: Springer.
- Fry, N. (1979). Random point distributions and strain measurement in rocks. *Tectonophysics*, 60(1), 89–105.
- Gao, Y., Zhang, Z., Xiong, Y., & Zuo, R. (2016). Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. *Ore Geology Reviews*, 75, 16–28.
- Gonçalves, M. A., Mateus, A., Pinto, F., & Vieira, R. (2018). Using multifractal modelling, singularity mapping, and geochemical indexes for targeting buried mineralization: Application to the W–Sn Panasqueira ore-system, Portugal. *Journal of Geochemical Exploration*, 189, 42–53.
- Haddad-Martim, P. M., de Souza Filho, C. R., & Carranza, E. J. M. (2017). Spatial analysis of mineral deposit distribution: A review of methods and implications for structural controls on iron oxide–copper–gold mineralization in Carajás, Brazil. *Ore Geology Reviews*, 81, 230–244.
- Hariharan, S., Tirodkar, S., Porwal, A., Bhattacharya, A., & Joly, A. (2017). Random forest-based prospectivity modelling of Greenfield Terrains using sparse deposit data: An example from the Tanami Region, Western Australia. *Natural Resources Research*, 26(4), 489–507.
- Harris, J. R., Grunsky, E., Behnia, P., & Corrigan, D. (2015). Data- and knowledge-driven mineral prospectivity maps for Canada's North. *Ore Geology Reviews*, 71, 788–803.
- He, H., & Garcia, E. A. (2010). Learning from imbalanced data sets. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1264.
- Hoens, T. R., & Chawla, N. V. (2013). Imbalanced datasets: From sampling to classifiers. In H. He & Y. Ma (Eds.), *Imbalanced learning: Foundations, algorithms, and applications* (pp. 43–59). Hoboken, NJ: Wiley.
- Hu, R., & Zhou, M. (2012). Multiple Mesozoic mineralization events in South China—An introduction to the thematic issue. *Mineralium Deposita*, 47(6), 579–588.
- Hua, R., Chen, P., Zhang, W., Liu, X., Lu, J., Lin, J., et al. (2003). Metallogenic systems related to Mesozoic and Cenozoic granitoids in South China. *Science China Earth Sciences*, 46(8), 816–829.
- Hua, R., Zhang, W., Chen, P., Zhai, W., & Li, G. (2013). Relationship between Caledonian granitoids and large-scale mineralization in South China. *Geological Journal of China Universities*, 19(1), 1–11.
- Huang, H., Chang, H., Tan, J., Li, F., Zhang, C., & Zhou, Y. (2015). Contrasting infrared microthermometry study of fluid inclusions in coexisting quartz, wolframite and other minerals: A case study of Xihuashan quartz-vein tungsten deposit, China. *Acta Petrologica Sinica*, 31(4), 925–940.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137–163.
- Li, T., Xia, Q., Chang, L., Wang, X., Liu, Z., & Wang, S. (2018). Deposit density of tungsten polymetallic deposits in the eastern Nanling metallogenic belt, China. *Ore Geology Reviews*, 94, 73–92.
- Liang, L., Liu, Z., Liu, S., & Zhang, S. (2016). Mineralization fracture characteristics and causes for Southern Jiangxi's Shilei tungsten–tin deposit. *China Tungsten Industry*, 31(1), 27–34.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Lisitsin, V. (2015). Spatial data analysis of mineral deposit point patterns: Applications to exploration targeting. *Ore Geology Reviews*, 71, 861–881.
- Liu, B., Chen, Y., Fan, S., Xu, J., Qu, W., & Ying, L. (2010). The second ore-prospecting space in the eastern and central parts of the Nanling metallogenic belt: Evidence from isotopic chronology. *Geology in China*, 37(4), 1034–1049.
- Liu, Y., Cheng, Q., Xia, Q., & Wang, X. (2014). Mineral potential mapping for tungsten polymetallic deposits in the Nanling metallogenic belt, South China. *Journal of Earth Science*, 25(4), 689–700.
- Liu, Y., Cheng, Q., Xia, Q., & Wang, X. (2015). The use of evidential belief functions for mineral potential mapping in the Nanling belt, South China. *Frontiers of Earth Science*, 9(2), 342–354.
- Liu, Y., Cheng, Q., Zhou, K., Xia, Q., & Wang, X. (2016). Multivariate analysis for geochemical process identification using stream sediment geochemical data: A perspective from compositional data. *Geochemical Journal*, 50(4), 293–314.
- Luo, J. (1990). Statistical mineral prediction without defining a training area. *Mathematical Geology*, 22(3), 253–260.
- Luo, X., & Dimitrakopoulos, R. (2003). Data-driven fuzzy analysis in quantitative mineral resource assessment. *Computers & Geosciences*, 29(1), 3–13.
- Mandelbrot, B. B. (1983). *The fractal geometry of nature (updated and augmented edition)*. New York: W. H. Freeman & Company.
- Mao, J., Xie, G., Cheng, Y., & Chen, Y. (2009). Mineral deposit models of Mesozoic ore deposits in South China. *Geological Review*, 55(3), 347–354.
- Mao, J., Xie, G., Guo, C., & Chen, Y. (2007). Large-scale tungsten-tin mineralization in the Nanling region, South China: Metallogenic ages and corresponding geodynamic processes. *Acta Petrologica Sinica*, 23(10), 2329–2338.
- Mao, J., Xie, G., Guo, C., Yuan, S., & Cheng, Y. (2008). Spatial-temporal distribution of Mesozoic ore deposits in South China and their metallogenic settings. *Geological Journal of China Universities*, 14(4), 510–526.
- Mao, J., Xie, G., Li, X., Zhang, C., & Mei, Y. (2004). Mesozoic large scale mineralization and multiple lithospheric extension in South China. *Earth Science Frontiers*, 11(1), 45–55.
- Mao, J., Xie, G., Li, X., Zhang, Z., Wang, Y., Wang, Z., et al. (2005). Geodynamic process and metallogeny: History and present research trend, with a special discussion on continental accretion and related metallogeny throughout geological history in South China. *Mineral Deposits*, 24(3), 193–205.
- McKay, G., & Harris, J. R. (2016). Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, 25(2), 125–143.
- Moeini, H., & Torab, F. M. (2017). Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran. *Journal of Geochemical Exploration*, 180, 15–23.
- Mu, Z., Huang, F., Chen, C., Zheng, S., Fan, S., Liu, D., et al. (1981). C–H–O stable isotope study on Piaotang–Xihuashan quartz-vein type tungsten deposits. In: H. Yu (Ed.), *Tungsten*

- deposit geological conference. Beijing: Geological Publishing House.
- Najafi, A., Abdi, M., Rahimi, B., & Motevali, K. (2010). Spatial integration of Fry and fractal analyses in regional exploration: A case study from Bafq-Posht-e-Badam, Iran. *Geologia Colombiana*, 35(0072–0992), 113–130.
- Nykänen, V. (2008). Radial basis functional link nets used as a prospectivity mapping tool for orogenic gold deposits within the central Lapland greenstone belt, Northern Fennoscandian Shield. *Natural Resources Research*, 17(1), 29–48.
- Nykänen, V., Groves, D. I., Ojala, V. J., Eilu, P., & Gardoll, S. J. (2008). Reconnaissance-scale conceptual fuzzy-logic prospectivity modelling for iron oxide copper-gold deposits in the northern Fennoscandian Shield, Finland. *Australian Journal of Earth Sciences*, 55(1), 25–38.
- Nykänen, V., Lahti, I., Niiranen, T., & Korhonen, K. (2015). Receiver operating characteristic (ROC) as validation tool for prospectivity models—A magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geology Reviews*, 71, 853–860.
- Parsa, M., Maghsoudi, A., & Yousefi, M. (2018). Spatial analyses of exploration evidence data to model skarn-type copper prospectivity in the Varzaghan district, NW Iran. *Ore Geology Reviews*, 92, 97–112.
- Pazzani, M., Merz, C., Murphy, P., Ali, K., Hume, T., & Brunk, C. (1994). Reducing misclassification costs. In *Proceedings of the 17th international conference on machine learning*. New Brunswick, NJ: Morgan Kaufmann Publishers Inc.
- Pei, R., Peng, C., & Xiong, Q. (1999). Deep tectonic processes and super accumulation of metals related to granitoid in the Nanling metallogenic province, China. *Acta Geologica Sinica*, 73(2), 191.
- Porwal, A., & Carranza, E. J. M. (2015). Introduction to the special issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geology Reviews*, 71, 477–483.
- Porwal, A., Carranza, E. J. M., & Hale, M. (2003a). Artificial neural networks for mineral potential mapping: A case study from Aravalli Province, Western India. *Natural Resources Research*, 12(3), 155–171.
- Porwal, A., Carranza, E. J. M., & Hale, M. (2003b). Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping. *Natural Resources Research*, 12(1), 1–25.
- Raines, G. L. (2008). Are fractal dimensions of the spatial distribution of mineral deposits meaningful? *Natural Resources Research*, 17(2), 87–97.
- Rao, J. R., Jin, X. Y., & Zeng, C. F. (2006). Deep tectonic-magmatic (rock) ore-controlling regularity and prospecting direction in the northern margin of middle Nanling range. *Land and Resources Herald*, 3(3), 31–36.
- Reimann, C., Filzmoser, P., & Garrett, R. G. (2002). Factor analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry*, 17(3), 185–206.
- Reimann, C., Filzmoser, P., Garrett, R., & Dutter, R. (2008). *Statistical data analysis explained: Applied environmental statistics with R*. Chichester: Wiley.
- Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*, 28(7), 1336–1354.
- Rodriguez-Galiano, V. F., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804–818.
- Roshanravan, B., Aghajani, H., Yousefi, M., & Kreuzer, O. (2019a). An improved prediction-area plot for prospectivity analysis of mineral deposits. *Natural Resources Research*, 28(3), 1089–1105.
- Roshanravan, B., Aghajani, H., Yousefi, M., & Kreuzer, O. (2019b). Particle swarm optimization algorithm for neuro-fuzzy prospectivity analysis using continuously weighted spatial exploration data. *Natural Resources Research*, 28(2), 309–325.
- Salati, S., van Ruitenbeek, F. J. A., Carranza, E. J. M., van der Meer, F. D., & Tangestani, M. H. (2013). Conceptual modeling of onshore hydrocarbon seep occurrence in the Dezful Embayment, SW Iran. *Marine and Petroleum Geology*, 43, 102–120.
- Schill, W., Jockel, K. H., Drescher, K., & Timm, J. (1993). Logistic analysis in case-control studies under validation sampling. *Biometrika*, 80, 339–352.
- Shu, L. (2012). An analysis of principal features of tectonic evolution in South China Block. *Geological Bulletin of China*, 31(7), 1035–1053.
- Shu, L., & Wang, D. (2006). A comparison study of basin and range tectonics in the western North America and southeastern China. *Geological Journal of China Universities*, 12(1), 1–13.
- Shu, L., Zhou, X., Deng, P., Yu, X., Wang, B., & Zu, F. (2004). Geological features and tectonic evolution of Meso-Cenozoic basins in southeastern China. *Geological Bulletin of China*, 23(9–10), 876–884.
- Singer, D. A., & Kouda, R. (1999). A comparison of the weights-of-evidence method and probabilistic neural networks. *Natural Resources Research*, 8(4), 287–298.
- Song, S., Hu, R., Bi, X., Wei, W., & Shi, S. (2011). Hydrogen, oxygen and sulfur isotope geochemical characteristics of Taoxikeng tungsten deposit in Chongyi County, Southern Jiangxi Province. *Mineral Deposits*, 30(1), 1–10.
- Tan, Y., Wen, M., Zhu, Y., & Qu, H. (2017). Research on the big data characteristics of geological data. *China Mining Magazine*, 26(9), 67–84.
- Tripathi, V. S. (1979). Factor analysis in geochemical exploration. *Journal of Geochemical Exploration*, 11(3), 263–275.
- Vearncombe, J. R., & Vearncombe, S. (1999). The spatial distribution of mineralization: Applications of Fry analysis. *Economic Geology*, 94(4), 475–486.
- Viktor, M., & Kenneth, C. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt Publishing Company.
- Wang, X., Ni, P., Jiang, S., Huang, J., & Sun, L. (2008). Fluid inclusion study on the Piaotang tungsten deposit, southern Jiangxi province, China. *Acta Petrologica Sinica*, 24(9), 2163–2170.
- Wang, X., Ni, P., Jiang, S., Zhao, K., & Wang, T. (2010). Origin of ore-forming fluid in the Piaotang tungsten deposit in Jiangxi Province: Evidence from helium and argon isotopes. *Chinese Science Bulletin*, 55(7), 628–634.
- Wang, J., & Zuo, R. (2015). A MATLAB-based program for processing geochemical data using fractal/multifractal modeling. *Earth Science Informatics*, 8(4), 937–947.
- Wei, C., Cai, M., Cai, J., Wang, X., Che, Q., & Du, H. (2004). Characteristics of structural control of ore deposition in South China in the Mesozoic. *Journal of Geomechanics*, 10(2), 113–121.
- Wei, W., Hu, R., Peng, J., Bi, X., Song, S., & Shi, S. (2011). Fluid mixing in Xihuashan tungsten deposit, Southern Jiangxi Province: Hydrogen and oxygen isotope simulation analysis. *Geochemica*, 40(1), 45–55.
- Wu, S., Dai, P., & Wang, X. (2016). C, H, O, Pb isotopic geochemistry of W polymetallic skarn-greisen and Pb-Zn-Ag veins in Shizhuyuan orefield, Hunan Province. *Mineral Deposits*, 35(3), 633–647.
- Wyborn, L. A. I., Heinrich, C. A., & Jaques, A. L. (1994). Australian Proterozoic mineral systems: Essential ingredients and

- mappable criteria. In P. C. Hallenstein (Ed.), *Australian mining looks north—The challenges and choices: Australian Institute of Mining and Metallurgy Publication Series* (Vol. 5, pp. 109–115).
- Xie, X., Mu, X., & Ren, T. (1997). Geochemical mapping in China. *Journal of Geochemical Exploration*, 60(1), 99–113.
- Xiong, Y., & Zuo, R. (2016). Recognition of geochemical anomalies using a deep autoencoder network. *Computers & Geosciences*, 86, 75–82.
- Xiong, Y., & Zuo, R. (2017). Effects of misclassification costs on mapping mineral prospectivity. *Ore Geology Reviews*, 82, 1–9.
- Xiong, Y., Zuo, R., & Carranza, E. J. M. (2018). Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geology Reviews*, 102, 811–817.
- Xu, T., & Wang, Y. (2014). Sulfur and lead isotope composition on tracing ore-forming materials of the Xihuashan tungsten deposit in Southern Jiangxi. *Bulletin of Mineralogy, Petrology and Geochemistry*, 33(3), 342–347.
- Yaghubpur, A., & Hassannejad, A. A. (2006). The spatial distribution of some chromite deposits in Iran, using Fry analysis. *Journal of Sciences*, 17(2), 147–152.
- Yousefi, M., & Carranza, E. J. M. (2015a). Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. *Computers & Geosciences*, 74, 97–109.
- Yousefi, M., & Carranza, E. J. M. (2015b). Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Computers & Geosciences*, 79, 69–81.
- Yousefi, M., & Carranza, E. J. M. (2016). Data-driven index overlay and Boolean logic mineral prospectivity modeling in Greenfields exploration. *Natural Resources Research*, 25(1), 3–18.
- Yousefi, M., & Carranza, E. J. M. (2017a). The efficiency of logistic function and prediction-area plot in prospectivity analysis of mineral deposits. In *Mineral prospectivity, current approaches and future innovations, Orléans, France* (pp. 68–69).
- Yousefi, M., & Carranza, E. J. M. (2017b). Union score and fuzzy logic mineral prospectivity mapping using discretized and continuous spatial evidence values. *Journal of African Earth Sciences*, 128, 47–60.
- Yousefi, M., Kamkar-Rouhani, A., & Carranza, E. J. M. (2012). Geochemical mineralization probability index (GMPI): A new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping. *Journal of Geochemical Exploration*, 115, 24–35.
- Yousefi, M., Kamkar-Rouhani, A., & Carranza, E. J. M. (2014). Application of staged factor analysis and logistic function to create a fuzzy stream sediment geochemical evidence layer for mineral prospectivity mapping. *Geochemistry: Exploration, Environment, Analysis*, 14(1), 45–58.
- Yousefi, M., & Nykänen, V. (2016). Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping. *Journal of Geochemical Exploration*, 164, 94–106.
- Yousefi, M., & Nykänen, V. (2017). Introduction to the special issue: GIS-based mineral potential targeting. *Journal of African Earth Sciences*, 128, 1–4.
- Yu, C. (2006). *Fractal growth of mineral deposits at the edge of chaos* (Vol. 1). Hefei: Anhui Education Publishing House.
- Yu, C., Luo, T., Bao, Z., & Hu, Y. (1987). *Regional geochemistry of the Nanling district*. Beijing: Geological Publishing House.
- Zhai, Y. (2003a). Metallogenic system and its evolution: From primary practice to theoretical consideration. *Journal of Earth Science*, 25(4), 333–339.
- Zhai, Y. (2003b). Research on metallogenic system. *Geological Survey and Research*, 26(2), 129–135.
- Zhai, Y., Deng, J., Cui, B., Ding, S., Peng, R., Wang, J., et al. (1999). Ore-forming system and comprehensive geo-anomaly. *Journal of Graduate School (China University of Geosciences)*, 13(1), 99–104.
- Zhai, Y., Deng, J., & Peng, R. (2000). Research contents and methods for post-ore changes, modifications and preservation. *Journal of Earth Science*, 25(4), 340–345.
- Zhai, Y., Wang, J., Deng, J., & Peng, R. (2002). Metallogenic system and mineralization network. *Mineral Deposits*, 21(2), 106–112.
- Zhang, D., Agterberg, F. P., Cheng, Q., & Zuo, R. (2014). A comparison of modified fuzzy weights of evidence, fuzzy weights of evidence, and logistic regression for mapping mineral prospectivity. *Mathematical Geosciences*, 46(7), 869–885.
- Zhang, Z., Zuo, R., & Xiong, Y. (2016). A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, 59(3), 556–572.
- Zhao, P. (2018). Characteristics and rational utilization of geological big data. *Earth Science Frontiers*, 25, 1–5.
- Zhao, P., Chi, S., Li, D., & Cao, X. (2013). *Theory and methods for mineral exploration* (6th ed.). Wuhan: China University of Geosciences Press.
- Zhao, K., Jiang, S., Zhu, J., Li, L., Dai, B., Jiang, Y., et al. (2010). Hf isotopic composition of zircons from the Huashan-Guoshan intrusive complex and their mafic enclaves in north-eastern Guangxi: Implication for petrogenesis. *Chinese Science Bulletin*, 55(6), 509–519.
- Zhou, X. (2007). *Late Mesozoic granite genesis and lithospheric dynamics evolution in the Nanling region* (1st ed.). Beijing: Science Press.
- Zhou, Z. (2016). *Machine learning* (1st ed.). Beijing: Peking University Press.
- Zhu, X., Wang, J., Wang, Y., Cheng, X., Fu, Q., & Yu, Z. (2014). The characteristics of the S, Pb, O, H isotope of Yaogangxian tungsten deposits, Hunan Province. *Geology and Exploration*, 50(5), 947–960.
- Zuo, R. (2017). Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural Resources Research*, 26(4), 457–464.
- Zuo, R., Agterberg, F. P., Cheng, Q., & Yao, L. (2009). Fractal characterization of the spatial distribution of geological point processes. *International Journal of Applied Earth Observation and Geoinformation*, 11(6), 394–402.
- Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers & Geosciences*, 37(12), 1967–1975.
- Zuo, R., Carranza, E. J. M., & Cheng, Q. (2012). Fractal/multifractal modelling of geochemical exploration data. *Journal of Geochemical Exploration*, 122, 1–3.
- Zuo, R., Carranza, E. J. M., & Wang, J. (2016). Spatial analysis and visualization of exploration geochemical data. *Earth-Science Reviews*, 158, 9–18.
- Zuo, R., & Wang, J. (2016). Fractal/multifractal modeling of geochemical data: A review. *Journal of Geochemical Exploration*, 164, 33–41.
- Zuo, R., Wang, J., Chen, G., & Yang, M. (2015). Identification of weak anomalies: A multifractal perspective. *Journal of Geochemical Exploration*, 154, 200–212.
- Zuo, R., & Xia, Q. (2008). The uncertainty propagation model in mineral resources prediction. *Progress in Geophysics*, 23(4), 1282–1285.
- Zuo, R., & Xiong, Y. (2018). Big data analytics of identifying geochemical anomalies supported by machine learning methods. *Natural Resources Research*, 27(1), 5–13.