Ore Geology Reviews 74 (2016) 26–38

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeorev

A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve analysis

Yongliang Chen [a,*], Wei Wu [b]

[a] Institute of Mineral Resources Prognosis on Synthetic Information, Jilin University, Changchun, Jilin Province 130026, China
[b] Changchun Institute of Urban Planning and Design, Changchun, Jilin Province 130033, China

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 31 October 2015
Accepted 7 November 2015
Available online 10 November 2015

Keywords:
ROC curve analysis
Youden index
Likelihood ratio
Lift index
Mineral potential mapping
Model performance assessment

ABSTRACT

A prospecting cost-benefit strategy is developed by quantitatively defining the prospecting cost and benefit in mineral potential mapping. Suppose that some mineral deposits have been discovered in a study area of a set of grid cells, the prospecting cost and benefit of a "unique" condition can be defined as the percentage of non-deposit-bearing and deposit-bearing cells within the "unique" condition, respectively. By replacing the false positive and true positive rates in the receiver operating characteristic (ROC) curve analysis with the prospecting cost and benefit, the Youden index, likelihood ratio, and lift index can be computed and used to express the mineral potential of the "unique" condition. Thus, the mineral potential mapping in a study area can be implemented by identifying all the possible "unique" conditions and then computing their mineral potential indicators such as the Youden index, likelihood ratio, and lift index. By integrating an automatic "unique" condition searching algorithm with the techniques for computing the mineral potential indicators for each "unique" condition, the following prospecting cost-benefit strategy is developed for mineral potential mapping: (a) select map patterns closely associated with the discovered mineral deposits using their mineral potential indicators, (b) automatically search for all the possible "unique" conditions, (c) evaluate the mineral potential of each "unique" condition using its mineral potential indicators, and (d) assess mineral potential mapping performance using the mineral potential indicator diagrams. For demonstration purposes, the Baishan district in Southern Jilin Province in China, which has a complex geological setting, is chosen as a case study area. The weights of evidence (WofE) modeling posterior probability, Youden index, likelihood ratio, and lift index are applied in the mineral potential mapping and their performance are assessed using their ROC curves, cumulative lift charts, and Youden and likelihood ratio diagrams. The results show that (a) the likelihood ratio and lift index perform similarly well and (b) the posterior probability performs a little bit worse than the likelihood ratio and lift index while a little bit better than the Youden index. Therefore, the prospecting cost-benefit strategy provides a common paradigm for both mineral potential mapping and the performance assessment.

1. Introduction

Mineral potential mapping can be a key procedure in mineral exploration. Its fundamental purpose is to minimize prospecting cost while maximizing prospecting benefit of a mineral exploration program. In the past decades, several dozens of mineral potential mapping approaches have been developed. They can be classified into data-driven, knowledge-driven, and hybrids of the two methods. Data-driven methods make use of the relationship between discovered mineral deposits and their surrounding map patterns to set up a mineral potential mapping model. These methods include weights of evidence or WofE (Agterberg, 1990, 1992; Agterberg et al., 1990; Bonham-Carter et al., 1988, 1989; Brown et al., 2000; Carranza and Hale, 2002a; Nykänen et al., 2008; Tangestani and Moore, 2001), extended weights of evidence (Mansour et al., 2009; Pan, 1996), fuzzy weights of evidence (Cheng et al., 2007; Porwal et al., 2006a), logistic regression (Agterberg, 1974, 1989; Agterberg and Bonham-Carter, 1999; Carranza and Hale, 2001b; Chen et al., 2011; Nykänen et al., 2008), feed-forward neural networks (Brown et al., 2000; Skabar, 2003), multilayer perceptrons (Skabar, 2007), Bayesian networks (Porwal et al., 2006b), radial basis functional link net (Leite and de Souza Filho, 2009a; Nykänen, 2008; Porwal et al., 2003), probabilistic neural networks (Leite and de Souza Filho, 2009b), certainty factor (Chen, 2003), evidence belief functions (An and Moon, 1993; Carranza and Hale, 2003; Carranza et al., 2005; Chen, 2004; Moon, 1989, 1990, 1993; Moon and So, 1995), multifractal singularity (Cheng, 2006; Cheng et al., 2009a, 2009b), and support vector machines (Abedi et al., 2012a; Zuo and Carranza, 2011). Knowledge-driven methods, however, apply empirical metallogenic and mineral-system knowledge to establish a mineral potential mapping model. They include Boolean logic (Bonham-Carter et al., 1989), index overlay

* Corresponding author at: Institute of Mineral Resources Prognosis on Synthetic Information, Jilin University, 938 Ximinzhu Street, Changchun, Jilin 130026, China.
E-mail address: chenyongliang2009@hotmail.com (Y. Chen).

http://dx.doi.org/10.1016/j.oregeorev.2015.11.011
0169-1368/

(Bonham-Carter et al., 1989; Carranza et al., 1999), fuzzy logic (An et al., 1991; Brown et al., 2000; Carranza and Hale, 2001a; Chung and Moon, 1990; D'Ercole et al., 2000; Karimi et al., 2008; Knox-Robinson, 2000; Molan and Behnia, 2013; Nykänen et al., 2008), wildcat mapping (Carranza, 2010; Carranza and Hale, 2002b), multiple criteria decision-making (Abedi et al., 2012b, 2012c, 2013a), interval valued fuzzy sets TOPSIS (Rad and Busch, 2011), outranking method (Abedi et al., 2013b), and restricted Boltzmann machine (Chen, 2014).

In this paper, by quantitatively defining prospecting cost and benefit, the fundamental purpose of mineral potential mapping is transformed into a prospecting cost-benefit strategy. Suppose that some mineral deposits have been discovered in a study area of a set of grid cells, then the prospecting cost and benefit of a "unique" condition can be defined as the percentage of non-deposit-bearing and deposit-bearing cells within the "unique" condition, respectively. The assumption is made that the non-deposit bearing cells have been adequately sterilized. This definition just coincides with the false positive and true positive rates in the receiver operating characteristic (ROC) curve analysis (Barreno et al., 2008). By replacing the false positive and true positive rates with the prospecting cost and benefit, the Youden index (Chen, 2014), likelihood ratio (Chen et al., 2014), and lift index (Anjum, 2014) can be computed and used as mineral potential indicators to express the mineral potential of the "unique" condition. The Youden index is negatively related to the prospecting cost while positively related to the prospecting benefit, and the likelihood ratio and lift index are inversely proportional to the prospecting cost but proportional to the prospecting benefit. Thus, maximizing these three mineral potential indicators is equivalent to minimizing the prospecting cost while maximizing the prospecting benefit of the mineral exploration program.

In mineral exploration, each "unique" condition is one class of mineral potential targets in a study area. Thus, the mineral potential mapping in a study area needs only to identify all the possible "unique" conditions and then compute their mineral potential indicators such as the Youden index, likelihood ratio, and lift index. The authors have developed an algorithm for automatically searching for all the possible "unique" conditions in a study area. By integrating this automatic searching algorithm with the techniques for computing the mineral potential indicators for each "unique" condition, the authors set up the following prospecting cost-benefit strategy for mineral potential mapping: (a) select map patterns closely associated with the discovered mineral deposits using their mineral potential indicators, (b) automatically search for all the possible "unique" conditions, (c) evaluate the mineral potential of each "unique" condition using its mineral potential indicators, and (d) assess mineral potential mapping performance using the mineral potential indicator diagrams. This strategy provides a common paradigm for both mineral potential mapping and the mineral potential mapping performance assessment of various models.

The Baishan district in Southern Jilin Province in China, which has a complex geological setting, is chosen as a case study area. The WofE modeling posterior probability, Youden index, likelihood ratio, and lift index are applied to map the mineral potential of the study area and their performance is assessed using their ROC curves, cumulative lift charts, and Youden and likelihood ratio diagrams. The results show that (a) the likelihood ratio and lift index perform similarly well and (b) the posterior probability performs a little bit worse than the likelihood ratio and lift index while a little bit better than the Youden index. The ROC curve analysis is overviewed in Section 2, the prospecting cost-benefit strategy are discussed in Section 3. A case study follows in Section 4 and finally the conclusion and discussion.

2. Overview on ROC curve analysis

In dealing with a binary classification problem, one class can be labeled as a positive and the other one as a negative class. Assume that the training sample set consists of p positive and n negative samples. A classifier assigns a class label to each of them, but some of the assignments may of course be wrong. To assess the classification results, the number of true positive (tp), true negative (tn), false positive (fp) (actually negative, but classified as positive) and false negative (fn) (actually positive, but classified as negative) samples can be estimated using a 2×2 contingence table. They satisfy

$$tp + fn = p; tn + fp = n. \tag{1}$$

The classifier assigned $tp + fp$ samples to the positive class and $tn + fn$ samples to the negative class. The following measures can be defined:

$$\text{fprate} = \frac{fn}{n} = 1 - \frac{tn}{n} = 1 - \text{specificity} \tag{2}$$

$$\text{tprate} = \frac{tp}{p} = \text{sensitivity} = \text{recall} \tag{3}$$

$$\text{yrate} = \frac{tp + fp}{p + n} \tag{4}$$

$$\text{Youden index} = \text{tprate} - \text{fprate} \tag{5}$$

$$\text{likelihood ratio} = \text{tprate}/\text{fprate} \tag{6}$$

$$\text{lift} = \text{tprate}/\text{yrate}. \tag{7}$$

The tprate, sensitivity, or recall measures the fraction of positive samples correctly classified, i.e., the classification accuracy of positive samples. The fprate or 1- specificity measures the fraction of negative samples that are misclassified as positive ones, i.e., the classification error of negative samples. The yrate measures the fraction of samples that are classified as positive ones, i.e., the fraction of targeted population. The Youden index and likelihood ratio represent respectively the difference and ratio between probability of a sample predicted as positive when it is truly positive, and probability of the sample predicted as positive when actually it is not positive. The Youden index is a comprehensive classification accuracy, a higher Youden index indicates better ability to avoid failure in binary classification. The likelihood ratio is a comprehensive performance measure in binary classification, a higher likelihood ratio means better classification performance on positive class. The lift represents the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. It is the ratio between the classification accuracy of positive samples to the fraction of targeted population. A lift value tells how much better a classifier predicts compared to a random guess.

A classification model is a function $f: X \rightarrow [0,1]$ that maps each sample x to a real number $f(x)$. Usually, a threshold t is chosen for which the samples where $f(x) \geq t$ are considered positive samples and the others are considered negative ones. This implies that each pair of a classifier and threshold t defines a binary classifier. Thus, a number of different binary classifiers, that is to say a binary classifier system, can be obtained by varying the threshold t.

A ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the tprate against the fprate at various threshold settings. The steeper the curve is toward the upper left corner, the better is the ability of the classifier to discriminate between positive and negative classes. The ROC curve was first developed by electrical and radar engineers during World War II for detecting enemy objects in battlefields and was soon introduced to psychology to account for perceptual detection of stimuli (Swets, 1996). The ROC curve analysis since then has been used in medicine, radiology, biometrics, and other areas for many decades (Zou et al., 2007) and is increasingly used in machine learning and data mining research (Hernandez-Orallo, 2013).

The ROC curve analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from the class

population distribution. Several ROC curves can be plotted in the same space to compare the results of different models, in the simplest case, a curve dominates others and modeling associated with the dominant curve is considered the most effective. This represents a visually simple instead of a qualitative interpretation of the ROC curve based on its shape. However, in more complex cases, which are more frequent, several curves may intersect, and it is more difficult to identify the top model.

Various ideas were proposed to solve the ROC curve intersect situation. One feasible way is to compute the area under the ROC curve (*AUC*). The *AUC* value is a summary indicator of ROC curve performance that can summarize the performance of a classifier system into a single metric. Unlike difficulties encountered in the comparison of different ROC curves especially in cases where they intersect, the *AUC* value can sort models by overall performance, as a result, the *AUC* value is considered more often in models assessment. The *AUC* value can vary between 0.5 and 1 in practice. If the *AUC* value equals 0.5, the classification performance is equivalent to a complete random guess; while if the *AUC* value equals 1, the classification performance is perfect, i.e., the classifier can correctly classify all samples. The *AUC* value of a classifier usually falls in somewhere between 0.5 and 1. The *AUC* value of a classifier system has an important statistical property as it is equivalent to the probability that the classifier system will rank a randomly chosen positive sample higher than a randomly chosen negative one.

The *AUC* value is estimated through various techniques, and the Wilcoxon Mann–Whitney test (Bergmann et al., 2000) is an often used nonparametric method. According to this method, the *AUC* value estimation is equivalent to the Wilcoxon test of ranks. Let x_i ($i = 1, 2, \ldots, p$) represent the predicted value of the ith positive sample and y_j ($j = 1, 2, \ldots, n$) represent the predicted value of the jth negative sample. Then, the *AUC* value can be estimated by

$$AUC = \frac{1}{p \times n} \sum_{i=1}^{p} \sum_{j=1}^{n} \varphi\left(x_i, y_j\right) \tag{8}$$

$$\varphi\left(x_i, y_j\right) = \begin{cases} 1 x_i > y_j \\ 0.5 x_i = y_j \\ 0 x_i < y_j \end{cases} \tag{9}$$

The standard deviation of *AUC* can be written as

$$S_{AUC} = \sqrt{\frac{AUC(1-AUC) + (p-1)\left(Q_1 - AUC^2\right) + (n-1)\left(Q_2 - AUC^2\right)}{p \times n}}. \tag{10}$$

Q_1 and Q_2 in Eq. (10) can be expressed by

$$Q_1 = \frac{AUC}{2-AUC}, Q_2 = \frac{2AUC^2}{1+AUC}. \tag{11}$$

Test of hypotheses can be implemented to test whether the *AUC* value is significantly different from $AUC = 0.5$. The following random variable can be estimated:

$$Z_{AUC} = \frac{AUC - 0.5}{S_{AUC}}. \tag{12}$$

Random variable Z_{AUC} satisfies the standard normal distribution. The value of Z_{AUC} can be used to determine the probability of a Type I error by finding the area in the tail of the normal distribution using the normal distribution table. For example, if $Z_{AUC} \pm 1.96$, the probability of a Type I error is 0.05.

3. Prospecting cost-benefit strategy

In mineral exploration, the aforementioned prospecting cost and benefit can be estimated and used to compute the Youden index, likelihood ratio, and lift index which can be further applied to select map patterns, map mineral potentials, and assess mineral potential mapping performance.

3.1. Prospecting cost and benefit estimation

In mineral potential mapping, it is very often the case that some mineral deposits have been discovered in a study area of a set of grid cells. Under this circumstance, deposit-bearing and non-deposit-bearing cells can be identified by superimposing the map of discovered mineral deposits on the map of grid cells. It is quite reasonable to view a mineral potential mapping procedure as a binary classification process that classifies grid cells into deposit-bearing and non-deposit-bearing cells. Then the prospecting cost and benefit defined in Section 1 can be estimated on the basis of the identified and predicted deposit-bearing and non-deposit-bearing cells by referring to the estimation method of the false positive and true positive rates in the ROC curve analysis.

Similar to binary classification, each pair of a mineral potential mapping model and a threshold can define a binary classifier which predicts grid cells as mineral potential targets (deposit-bearing cells) and non-mineral potential targets (non-deposit-bearing cells). Based on the identified and predicted results, for each threshold, the number of correctly predicted deposit-bearing cells, the number of incorrectly predicted deposit-bearing cells, the number of correctly predicted non-deposit-bearing cells, and the number of incorrectly predicted non-deposit-bearing cells, as generated under the assumptions adopted in the model, can be counted. These four indexes sequentially correspond to the number of true positive (*tp*), the number of false negative (*fn*), the number of true negative (*tn*), and the number of false positive (*fp*) in a binary classification. Consequently, the prospecting cost and benefit, with respect to the false positive rate (fprate) and true positive rate (tprate) in a binary classification, can be straightforwardly computed using Eqs. (2) and (3), respectively.

Prospecting cost and benefit may range from 0 to 1 which is same as the range of the fprate and tprate in a binary classification. A prospecting cost = 0 implies that all the non-deposit-bearing cells are correctly predicted; a prospecting cost = 1 indicates that all the non-deposit-bearing cells are incorrectly predicted; a prospecting benefit = 0 presents that all the deposit-bearing cells are incorrectly predicted; and a prospecting benefit = 1 denotes that all the deposit-bearing cells are correctly predicted.

3.2. Mineral potential indicator calculation

In mineral potential mapping, the mineral potential indicators such as the Youden index, likelihood ratio, and lift index can be easily computed on the basis of the estimated prospecting cost and benefit.

The Youden index can be computed using Eq. (5). It has been successfully applied in diagnostic test (Ruopp et al., 2008). Chen (2014) tentatively applied the Youden index to determine the optimal threshold for geochemical anomaly identification. In mineral potential mapping, the Youden index is defined as the difference between the fraction of the correctly predicted deposit-bearing cells and the fraction of incorrectly predicted non-deposit-bearing cells. The range of Youden index is between −1 and 1. A value 1 indicates a perfect prediction, a value 0 indicates that a prediction is the same as a random guess, and a negative value indicates that a prediction is worse than a random guess.

The likelihood ratio can be computed using Eq. (6). It has been widely applied in machine learning (Johnson, 2004). Chen et al. (2014) preliminarily applied the likelihood ratio to assess the performance of geochemical anomaly recognition. In mineral potential

mapping, the likelihood ratio is defined as the ratio between probability of a cell predicted as a deposit-bearing cell when it is truly a deposit-bearing cell and probability of the cell predicted as a deposit-bearing cell when actually it is a non-deposit-bearing cell. The range of the likelihood ratio is from zero to positive infinity. A value 1 indicates that the prediction is same as a random guess, a value toward infinity indicates a perfect prediction, and a value less than one implies that the prediction is worse than a random guess.

The lift index can be computed using Eq. (7). It has been widely applied in custom churn prediction (Anjum, 2014). In this paper, the lift index is preliminarily applied in mineral potential mapping, it is used to represent how much a mineral potential mapping model is more effective than a random guess. The range of the lift index is from zero to positive infinity. A value 1 indicates that the prediction is same as a random guess, a value toward positive infinity indicates a perfect prediction, and a value less than one means that the prediction is worse than a random guess.

The Youden index and likelihood ratio are independent of the relative sample size of deposit-bearing and non-deposit-bearing cells as well as the absolute sample size of total grid cells. The lift index is not independent of the relative sample size of deposit-bearing and non-deposit-bearing cells but independent of the absolute sample size of total grid cells. In mineral potential mapping, these three mineral potential indicators can be used in binary map pattern selection, continuous map pattern optimal segmentation, mineral potential evaluation, and model performance assessment.

3.3. Binary map pattern selection

A binary map pattern used for mineral potential mapping must be closely associated with the discovered mineral deposits in a study area. The Youden index, likelihood ratio, and lift index can be used to measure the association between a map pattern and the discovered mineral deposits. Suppose that a map pattern separates the study area into two classes of subareas, one where the map pattern exists and one where the map pattern does not exist. Each class of subareas can be viewed as mineral exploration targets, so the fprates, tprates, and yrates can be estimated using Eqs. (2) through (4), respectively. Then the Youden index, likelihood ratio, and lift index can be computed using Eqs. (5) through (7) sequentially. The map pattern is regarded to be closely associated with the discovered mineral deposits if the Youden index for one class of subareas is much more than zero, or the likelihood ratio or lift index for one class of subareas is much more than one.

3.4. Continuous map pattern selection

A continuous map pattern used for mineral potential mapping must be positively correlated to the discovered mineral deposits in a study area. The higher the positive correlation between a continuous map pattern and the discovered mineral deposits, the higher is the percentage of the discovered mineral deposits within the positive anomalies of the continuous map pattern, and consequently the better the continuous map pattern performs in the mineral potential mapping.

The ROC curve of a continuous map pattern can graphically reflect the correlation between the continuous map pattern and the discovered mineral deposits. The steeper the curve is toward the upper left corner, the higher is the correlation between the continuous map pattern and the discovered mineral deposits. Thus, the continuous map patterns which are highly correlated to the discovered mineral deposits can be recognized using their ROC curves.

The *AUC* of a continuous map pattern can quantitatively express the correlation between the continuous map pattern and the discovered mineral deposits. It can be calculated by Eqs. (8) and (9) and its standard deviation can be computed by Eqs. (10) and (11). An *AUC*-dependent statistics Z_{AUC} can be computed using Eq. (12) and used to test whether the continuous map pattern is significantly correlated to the discovered mineral deposits. The continuous map pattern is regarded to be significantly correlated to the discovered mineral deposits at the level $\alpha = 0.05$ if the absolute value of Z_{AUC} is more than 1.96.

3.5. Continuous map pattern optimal segmentation

The selected continuous map patterns must be transformed into binary map patterns so that they can be further applied in mineral potential mapping. An anomaly identification method can transform continuous map patterns (like geochemical element concentration values) into binary map patterns (such as geochemical anomalies). There is a variety of anomaly identification methods, for example, the iterative mean $\pm 2\sigma$ statistical methods (Galuszka, 2007; Hawkes and Webb, 1962), box plot (Tukey, 1997), and fractal and multifractal methods (Cheng, 1995, 2000, 2006, 2007, 2008; Zuo et al., 2009), to name but a few. However, these methods do not take into account the relationship between a continuous map pattern and the discovered mineral deposits in the continuous map pattern segmentation.

In this paper, the authors provide an optimal threshold method which can approximately maximize the association between the transformed binary map pattern (i.e., the identified anomalies) and the discovered mineral deposits in the continuous map pattern segmentation. First, the continuous distribution of threshold values is discretized by dividing the difference between the maximum and minimum into a number of equal intervals, which are progressively cumulated. Then the Youden index, likelihood ratio, and lift index, with respect to each of the discretized thresholds, can be computed and the threshold with respect to the largest Youden index or likelihood ratio or lift index is selected as the optimal threshold for the continuous map pattern segmentation.

3.6. Search for "unique" conditions

Suppose that all the binary map patterns which are closely associated with the discovered mineral deposits have been selected and all the selected continuous map patterns have been optimally segmented. Then the cells with the same map pattern composition display "unique" conditions and represent a class of mineral exploration targets. The main task of mineral potential mapping is to estimate the mineral potential of each class of mineral exploration targets, thus all the possible "unique" conditions must be identified first so that their mineral potentials can be estimated further.

Let n and p denote the total numbers of grid cells and binary map patterns, respectively. Two-dimensional array "data" with size of $n \times (p + 1)$ is used to preserve the input data as follows: item datum $[i, 0] = 1$ denotes that a mineral deposit exists in grid cell i while item datum $[i, 0] = 0$ denotes that no mineral deposit exists in grid cell i; and item datum $[i, j] = 1$ $(1 \le j \le p)$ denotes that map pattern j exists in grid cell i while item datum $[i, j] = 0$ denotes that map pattern j does not exist in grid cell i. One-dimensional array "d" with size of n is used to indicate whether a grid cell is located in a blank area. Item d $[i] = 0$ denotes that grid cell i is located in a blank area while item d $[i] = 1$ denotes that grid cell i contains a possible "unique" condition. One-dimensional array "pattern" with size of n is used to preserve which "unique" condition occurs in a grid cell. Item pattern $[i] = 0$ denotes that grid cell i is located in a blank area and item pattern $[i] = kk$ $(kk \ge 1)$ denotes that "unique" condition kk occurs in grid cell i. Table 1 lists a Python 2.x module which can be applied to search for each of all the possible "unique" conditions in a study area of n grid cells.

3.7. Mineral potential estimation

Suppose that all the possible "unique" conditions have been identified in a study area. Then for each "unique" condition, the following three steps can be implemented: (a) count the number of deposit-bearing and non-deposit-bearing cells that belong and do not belong

Table 1
Python 2.x module for "unique" condition search.

```
import numpy as np

def pattern_search (n, p, d, data):
    pattern = np.zeros (n, int)
    kk = 0
    for i in range(n):
        if d[i] == 1:
            if pattern[i] ! = 0:
                continue
            kk += 1
            pattern[i] = kk
            for j in range(i + 1, n):
                if d[i] == 1:
                    if pattern[j] ! = 0:
                        continue
                    num = 0
                    for k in range(1, p + 1):
                        if data[i, k] ! = data[j, k]:
                            break
                        num += 1
                    if num == p:
                        pattern[j] = pattern[i]
    return pattern
```

to the "unique" condition, these four statistics just correspond to *tp*, *fp*, *fn*, and *tn*, respectively; (b) the fprate (prospecting cost) and tprate (prospecting benefit) as well as the yrate of the "unique" condition are then computed using Eqs. (2) through (4) sequentially; and (c) the Youden index, likelihood ratio, and lift index of the "unique" condition are computed on the basis of fprate, tprate, and yrate using Eqs. (5) through (7), respectively. After the Youden indexes, likelihood ratios, and lift indexes of all the possible "unique" conditions have been computed, the mineral potential maps can be plotted on the basis of each of the three mineral potential indicators.

In mineral potential mapping, each grid cell is mapped to a real number used for expressing its mineral potential. Then a threshold value can be chosen for which the cells the mineral potentials of which are more than the threshold are considered mineral potential targets while the others are considered non-mineral potential targets. By varying threshold values step by step, different classes of mineral potential targets can be sequentially obtained.

3.8. Model performance assessment

The performance of different mineral potential mapping models can be evaluated using the similar performance assessing method in binary classification. First, predefine a set of thresholds which distribute evenly between the minimum and maximum values of a mineral potential mapping result. Then with respect to each threshold, the prospecting cost and benefit for the mineral potential mapping result are computed and a ROC curve is drawn by plotting the prospecting benefit against the prospecting cost at various threshold settings.

The ROC curves for different mineral potential mapping results can be plotted in the same space and the performance of these mineral potential mapping results can be graphically compared to one another. The mineral potential mapping result the ROC curve of which is nearer to the up-left corner has comparatively better performance. Under the situation where the ROC curves of different mineral potential mapping results cross, *AUC* values can be computed and used to assess the overall performance of different mineral potential mapping results (Chen, 2014). An additional statistics Z_{AUC} can be further computed on the basis of the *AUC* values and used to test whether a mineral potential mapping result is significantly different from the result obtained by a random guess.

Besides the ROC curves, cumulative gains and lift charts (Anjum, 2014) can be drawn by respectively plotting prospecting benefit and lift index against yrate at various threshold settings. Similar to the ROC curves, the cumulative gain curves of different mineral potential mapping results can be plotted in the same ROC space. The mineral potential mapping result the cumulative gain curve of which is nearer to the up-left corner has comparatively better performance. For a cumulative lift chart, the further it is away from the baseline of lift = 1.0, the better is the performance of the mineral potential mapping result. In the situations where different cumulative gain curves cross, the area under the cumulative gain curve or *AUL* (Bekkar et al., 2013) can be computed on the basis of the corresponding *AUC* and used to measure the overall performance of different mineral potential mapping results. The relation between the *AUL* and *AUC* can be expressed by

$$AUL = \frac{p}{2(p+n)} + \left(1 - \frac{p}{p+n}\right) \times AUC \tag{13}$$

where: p and n are the number of deposit-bearing and non-deposit-bearing cells, respectively.

In this paper, the authors provide two new curves that can be used to compare the performance of different mineral potential mapping results. They are the Youden and likelihood ratio diagrams obtained by respectively plotting the Youden index and likelihood ratio against the prospecting cost at various threshold settings. The performance of different mineral potential mapping results can be assessed using these two diagrams as follows. For the Youden diagram, the further the curve is away from the axis of prospecting cost, the better is the performance of the mineral potential mapping result. While for the likelihood ratio diagram, the further the curve is away from the baseline of likelihood ratio = 1.0, the better is the performance of the mineral potential mapping result.

4. Case study

The Baishan district in Southern Jilin Province in China, which has a complex geological setting, is chosen as the case study area. The WofE model and the prospecting cost-benefit strategy are applied in mapping the mineral potential of the area. The posterior probability, Youden index, likelihood ratio, and lift index are applied to express mineral potential of each grid cell. The ROC curves, cumulative lift charts, Youden and likelihood ratio diagrams, and *AUC*s (Chen, 2014) and *AUL*s are applied to measure the performance of these four mineral potential indicators.

4.1. Geological settings and mineral deposits

The study area is located in the eastern section of the northern margin of the North China Platform. It has experienced four geotectonic evolution stages which include the formation of the Archean-Proterozoic basement, rifting in the middle to late Proterozoic, "soft" orogenesis in the Paleozoic, extensional regime in the Mesozoic, and evolution of basin-ridge geotectonic regime in the Cenozoic (Liu et al., 2000). This complicated geotectonic evolution resulted in the distribution of the Archean granite–greenstone and Paleoproterozoic rift formation, and late Proterozoic–Paleozoic and Mesozoic sedimentation in the study area (Fig. 1). The late Mesozoic was, in addition, a period of vigorous volcanic eruptions and large-volume granitoid intrusion (Wu et al., 2005). These volcanic eruptions and granitoid intrusion directly resulted in the widely distributed volcanic rocks and a number of outcrops of the Mesozoic granites and granite porphyries in the study area (Fig. 1). The complicated geotectonic evolution also led to the appearance of middle Proterozoic and Mesozoic metallogenic epochs in the study area (Liu et al., 2000). Dahenglu copper–cobalt deposit, Huanggoushan and Banmiaozi gold deposits, and other 27 mineral deposits and mineral occurrences have been discovered in the study area.

Dahenglu copper–cobalt deposit is genetically a deposit of submarine hot-water deposition superimposed by late thermal solutions. It was formed in the formation of phyllite with the intercalated beds of quartzite (siliceous rocks) (Yang et al., 2001). The phyllite formation is rich in boron and carbon and belongs to early Proterozoic Laoling Group (Yang et al., 2001). Ore-bodies occurred in the form of beds and

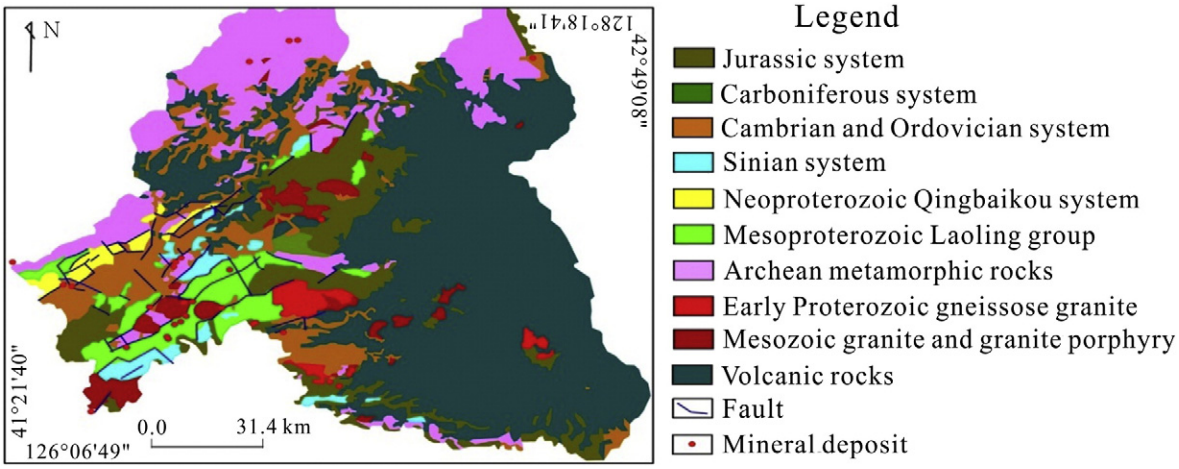

Fig. 1. Simplified geologic map superimposed with the discovered mineral deposits.

saddles of carbon-bearing sericite phyllite, tourmaline-bearing sericite phyllite, quartzite as well as tourmaline quartzite and their occurrences are concordant with the bedding of the wall rocks (Yang et al., 2001).

The Huanggoushan and Banmiaozi gold deposits are genetically a submarine hot-water deposition superimposed by late metamorphic-hydrothermal solutions and karst water. The deposits occurred in the accretionary terrain of early Proterozoic Laoling-Jilin rift zone and the ore-bearing rocks are magnesium-rich carbonate rocks of Zhenzhumen Formation of early Proterozoic Laoling Group (Yang et al., 1999). Archean supracrustal rocks and the Late Proterozoic Laoling Group comprise the main source beds and the northeastern deep faults are the main ore-controlling tectonics (Li, 2009; Li et al., 2010). The gold deposits were located at the intersection of NE-trending fractural belt and angular unconformity between marble of the early Proterozoic Zhenzhumen Formation and quartzite of Diaoyutai Formation of Neoproterozoic Qingbaikou System. Gold localization was controlled by both the geological boundaries and fractural belts (Su and Zang, 2010).

The gold and polymetallic mineralization in the study area was closely related to the four-stage tectonic evolution (Liu et al., 2000). Proterozoic and Mesozoic are the two main metallogenic epochs. The Dahenglu copper-cobalt deposit formed mainly in the Proterozoic while the Huanggoushan and Banmiaozi gold deposits formed mainly in the Mesozoic. However, these deposits are controlled by similar regional geological factors, such as the northeastern deep faults, Archean supracrustal rocks, and Proterozoic formations, Mesozoic volcano-magmatic and hydrothermal activities are in these cases the principal regional metallogenic controlling factors. Therefore, a combination of the discovered mineral deposits was used to predict mineral potential targets in this case study.

4.2. Geological map pattern selection

Based on the previous works and results discussed in Section 4.1, the following geologic entities were preliminarily selected as the evidence map patterns for mineral potential mapping: (a) Archean metamorphic rocks, (b) Proterozoic gneissose granitites, (c) Mesozoic granitite and granitite porphyry, (d) Laoling Group rocks, (e) Neoproterozoic Qingbaikou System rocks, (f) Sinian System rocks, (g) Cambrian-Ordovician System rocks, (h) Carboniferous System rocks, (i) Jurassic System rocks, (j) Jurassic volcanic rocks, and (k) northeastern faults.

Before quantitatively selecting the binary map patterns, the study area was first divided into 15,000 grid cells (100 by 150) and each cell is 1.2167 km high and 1.6365 km wide. A total of 5932 cells out of the original 15,000 were categorized as the blank area. The linear structures were then optimally buffered. In order to determine the optimal buffering width of the linear structures, the following buffering widths are predefined: 0.1 km, 0.2 km, 0.3 km, 0.4 km, 0.5 km, 0.6 km, 0.7 km, 0.8 km, 0.9 km, and 1.0 km. The optimal buffering width should maximize the association between the buffered linear structures and the discovered mineral deposits. The lift index was applied to express the association between the buffered linear structures and the discovered mineral deposits. Except for cells in the blank area, 9068 cells were used to estimate the lift indexes of the linear structures buffered at different widths. Fig. 2 shows that the lift index reaches its maximum value at 0.4 km buffer. Thus, 0.4 km was chosen as the optimal width for buffering the linear structures in the study area.

Among binary map patterns, only those which are closely associated with the discovered mineral deposits can contribute to mineral potential mapping. The lift indexes were applied to measure the relationship between a binary map pattern and the discovered mineral deposits. The 9068 cells were used to estimate lift index of each binary map pattern. The estimated lift indexes for the 11 geological binary map patterns are listed in Table 2. These estimated lift indexes vary from 0.0 to 11.312. Among them, 6 lift indexes are more than 2.0 and 5 lift indexes equal to 0.0. Thus, the following 6 geological map patterns are selected for mineral potential mapping: (a) Archean metamorphic rocks, (b) Mesozoic granitite

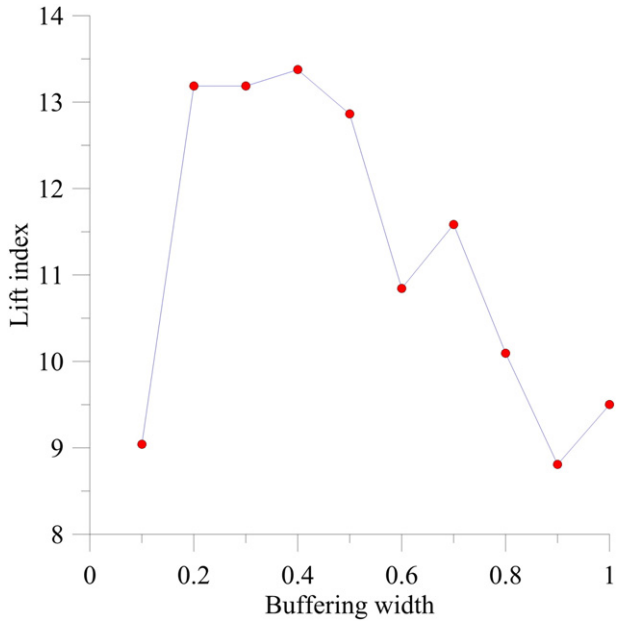

Fig. 2. Diagram of lift indexes varying with buffering widths.

Table 2
Lift indexes for the 11 geological map patterns.

Map pattern	Lift index	Map pattern	Lift index
Archean metamorphic rocks	2.072	Cambrian–Ordovician System rocks	2.121
Proterozoic gneissose granitites	0.0	Carboniferous System rocks	0.0
Mesozoic granitite and granitite porphyry	4.559	Jurassic System rocks	2.298
Laoling Group rocks	10.707	Jurassic volcanic rocks	0.0
Neoproterozoic Qingbaikou System rocks	0.0	Northeastern faults	11.312
Sinian System rocks	0.0		

and granitite porphyry, (c) Jurassic System rocks, (d) Cambrian–Ordovician System rocks, (e) Laoling Group rocks, and (f) buffered northeastern faults.

The above 6 geological map patterns approximately coincide with the principal metallogenic controlling factors in the study area. The Archean metamorphic rocks provided the ore-forming material sources for the formation of the Laoling Group rocks which are the main source beds of polymetallic mineralization. The magmatic and hydrothermal activities, which are genetically related to the Mesozoic granitite and granitite porphyries, provide the heat source and magmatic hydrothermal solutions for the polymetallic mineralization. The volcanic activities genetically related to the Jurassic System rocks, which are volcanic-sedimentary formations, provide the heat source and volcanic hydrothermal solutions for the polymetallic mineralization. The northeastern faults provide the most favorable host space for the enrichment of ore-forming materials. The Cambrian–Ordovician System rocks are mainly carbonate and clastic rock formations which are not genetically related to the polymetallic mineralization but spatially associated with the discovered mineral deposits.

4.3. Geochemical map pattern selection

A geochemical stream sediment survey has been conducted in the study area. Stream sediment samples were collected at a 1 km × 1 km grid within drainage basins in the 4 adjacent geological maps of scale 1:200,000 (Fig. 3). X-ray fluorescence spectrometry was used to measure the concentration values of geochemical indicators. The concentration values of 29 elements and 6 oxides in each of the sediment samples have been analyzed in different concentration units: parts per billion (ppb) for gold and silver; parts per million (ppm) for the other 27 elements; and weight/weight percent for the 6 oxides.

In order to express the spatial correlation between the 35 geochemical variables and the discovered mineral deposits in the study area, the *AUC* value and *AUC*-dependent statistics Z_{AUC} for each geochemical variable were calculated and listed in Table 3. From Table 3, it can be seen that there are 17 geochemical variables which are significantly and positively correlated to the discovered mineral deposits because their Z_{AUC}s are more than the critical value 1.96 at the level $\alpha = 0.05$. Thus, they are selected for mineral potential mapping.

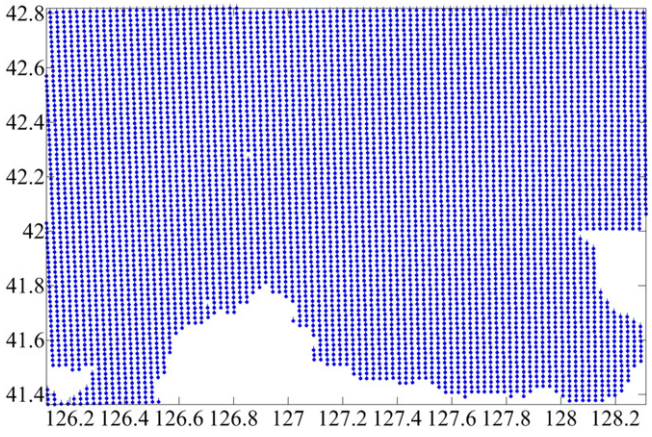

Fig. 3. Map of geochemical sampling positions.

The lift index is applied to optimal segmentation of the above 17 selected geochemical variables. In order to determine the optimal threshold for identifying geochemical anomalies, 1000 thresholds were predefined to distribute uniformly between the minimum and maximum concentration values of each geochemical variable. The lift index, with respect to each threshold, is computed and the threshold with respect to the maximized lift index is chosen as the optimal threshold. Table 4 lists the maximized lift indexes with respect to the optimal thresholds for the 17 selected geochemical variables. The extracted geochemical anomalies then serve as binary geochemical map patterns.

4.4. Conditional independence test and map pattern combination

The WofE model can be used to estimate the metallogenic posterior probabilities if the conditionally independent map patterns given the discovered mineral deposits are used in the mineral potential mapping in a study area. An unfailing WofE modeling result requires the map patterns to satisfy both pairwise and overall conditional independence assumption. However, the different map patterns used for mineral potential mapping are always somewhat mutually conditionally dependent in practical applications. Thus, the conditional independence tests were conducted for the WofE modeling in our case study.

The contingency table and new conditional independence tests (Agterberg and Cheng, 2002) were applied to test the 23 selected binary map patterns (i.e., the 6 selected geological map patterns and the 17 selected and discretized geochemical map patterns). The pairwise test statistic defined by Agterberg and Cheng (2002) is distributed as χ^2 with a single degree of freedom if the two binary map patterns are conditionally independent. The critical value of χ^2-statistic with freedom one is 6.635 at the level $\alpha = 0.01$. The results of pairwise tests show that the following map patterns are not conditionally independent: (a) Ag is conditionally dependent with Au, Cao, Cd, Hg, MgO, Pb, V, and Zn; (b) As is conditionally dependent with B, Sb, and Sn; (c) Au is conditionally dependent with CaO, MgO, and V; (d) B is conditionally dependent with Sb and Sn; (e) Bi is conditionally dependent with P; (f) CaO is conditionally dependent with Hg, MgO, and V; (g) Cd is conditionally dependent with Hg and Pb; (h) Hg is conditionally dependent with MgO, Pb, V, and Zn; (i) MgO is conditionally dependent with V; (j) P is conditionally dependent with Mesozoic granitite and granitite porphyry; (k) Pb is conditionally dependent with Zn; and (l) Sb is conditionally

Table 3
Estimated *AUC*s and Z_{AUC}s of 35 geochemical variables.

Variable	*AUC*	Z_{AUC}	Variable	*AUC*	Z_{AUC}	Variable	*AUC*	Z_{AUC}
Ag	0.741	4.616	Cu	0.801	6.190	Pb	0.718	4.095
Al_2O_3	0.567	1.231	Fe_2O_3	0.558	1.073	Sb	0.720	4.141
As	0.717	4.074	Hg	0.682	3.365	SiO_2	0.385	−2.443
Au	0.773	5.412	La	0.484	−0.306	Sn	0.667	3.071
B	0.626	2.301	Li	0.590	1.651	Sr	0.555	1.019
Ba	0.491	−0.179	MgO	0.817	6.692	Ti	0.427	−1.477
Be	0.312	−4.560	Mn	0.501	0.011	V	0.683	3.373
Bi	0.742	4.629	Mo	0.498	−0.037	W	0.675	3.222
CaO	0.785	5.735	Na_2O	0.407	−1.935	Y	0.386	−2.426
Cd	0.798	6.109	Nb	0.298	−5.035	Zn	0.670	3.120
Co	0.527	0.497	Ni	0.605	1.911	Zr	0.239	−7.630
Cr	0.556	1.029	P	0.697	3.663			

Table 4
The maximized lift indexes and optimal thresholds for 17 geochemical variables.

Geochemical variable	Maximized lift	Optimal threshold
Ag	29.412	312.953
As	38.462	57.348
Au	25.862	6.886
B	100.0	208.192
Bi	90.909	1.778
CaO	71.429	10.281
Cd	18.519	1053.746
Cu	10.638	55.243
Hg	55.556	325.979
MgO	71.429	12.895
P	45.455	2432.925
Pb	22.727	91.561
Sb	41.667	2.291
Sn	13.514	10.713
V	11.628	184.664
W	8.929	4.269
Zn	55.556	504.984

dependent with Sn. The new conditional independence test show that $ET - n = 57.205–30 = 27.205 > 1.645 * s(T) = 1.645 \times 8.836 = 14.535$. Therefore, the 23 selected map patterns satisfy neither the pairwise nor the overall conditional independence assumption.

In order to make the above 23 map patterns satisfy the conditional independence assumptions in the WofE modeling. The two mutually dependent map patterns A and B are combined into a comprehensive map pattern AB using the set operation: AB = A∪B, and then replace B with AB and delete A. This map pattern combination procedure can be automatically implemented in a software program. Through this processing, the 23 selected map patterns are combined into 11 comprehensive map patterns. Both the pairwise and overall conditional independence assumptions of the 11 comprehensive map patterns were tested using Agterberg and Cheng's methods (2002). The results show that the 11 comprehensive map patterns violate the overall conditional independence assumption due to the inequality that $ET - n = 61.537–30 = 31.537 > 1.645 * s(T) = 1.645 \times 13.512 = 22.227$. This overall conditional dependence is caused by the two pairs of pairwise conditional dependence that exist among the 11 comprehensive map patterns. So, the map pattern combination procedure was further applied to the 11 comprehensive map patterns and finally 9 comprehensive map patterns were constructed. The pairwise conditional independence test reveals that the 9 comprehensive map patterns satisfy pairwise conditional independence assumption (Table 5). The overall conditional independence test shows that $ET - n = 33.698–30 = 3.698 < 1.645 * s(T) = 1.645 \times 7.391 = 12.158$. Thus, the 9 comprehensive map patterns satisfy overall conditional independence assumption.

4.5. Mineral potential mapping with the WofE model

The 9 comprehensive map patterns constructed in Section 4.4 were applied to the WofE modeling. For each comprehensive map pattern, two weights and the weight variances and the variance due to missing data were estimated on the basis of the binary attributive data of the 9068 grid cells and listed in Table 6. The metallogenic posterior probability and its deviation for each cell are computed by the WofE modeling procedure. Fig.4 shows the posterior probability and posterior probability deviation maps.

4.6. Mineral potential mapping with the prospecting cost-benefit strategy

The 9 comprehensive map patterns used in the WofE modeling were also used in the prospecting cost-benefit strategy in mineral potential mapping. First, the "unique" condition searching algorithm discussed in Section 3.5 were applied to identify all the possible "unique" conditions in the study area. Then for each "unique" condition, the number of both non-deposit-bearing and deposit-bearing cells within subareas where the "unique" condition appears are counted and used to compute the prospecting cost and benefit defined in Section 3.1. Finally, the Youden index, likelihood ratio, and lift index are calculated on the basis of the prospecting cost and benefit of the "unique" condition. The estimated three mineral potential indicators of all the possible "unique" conditions in the study area are used to draw mineral potential maps. Fig. 5 shows the maps of the Youden indexes, likelihood ratios, and lift indexes.

4.7. Performance evaluation

For the four mineral potential mapping results obtained in Sections 4.5 and 4.6, 1000 thresholds were predefined to distribute evenly between the minimum and maximum values of each mineral potential mapping result. Then, the prospecting cost and benefit with respect to each threshold are computed and 1000 pairs of the prospecting cost and benefits were finally used to draw the ROC curve. Fig. 6a shows the ROC curves of the four mineral potential mapping results. From Fig. 6a, the following conclusion can be drawn: (a) the ROC curves of likelihood ratio and lift index almost coincide and they are located nearer to the upper left corner compared to the other two, thus the likelihood ratio and lift index perform similarly well and a little bit better than the WofE and Youden index; and (b) the ROC curves of the WofE model and Youden index cross, so the performance of these two mineral potential indicators is difficult to differentiate using their ROC curves.

Fig. 6b through d sequentially shows the Youden and likelihood ratio diagrams and cumulative lift charts of the four mineral potential mapping results. The Youden diagrams shown in Fig. 6b are similar to the ROC curves shown in Fig. 6a. The Youden diagrams of the WofE and Youden index cross. Thus the Youden diagrams cannot discriminate the performance of the WofE and Youden index. However, the likelihood ratio diagrams shown in Fig. 6c and the cumulative lift charts shown in Fig. 6d can easily differentiate the performance of the four mineral potential indicators. According to Fig. 6c and d, it can be concluded that: (a) the likelihood ratio and lift index perform similarly well and (b) the WofE model performs a little bit worse than the likelihood ratio and lift index but a little bit better than the Youden index.

In order to measure the overall performance of the four mineral potential mapping results, the *AUC*, *AUL* and Z_{AUC} values are computed and listed in Table 7. According to Table 7, the following conclusion can be

Table 5
Pairwise conditional independence test for the 9 comprehensive map patterns.

	a	b	c	d	e	f	g	h
b	0.321							
c	0.231	0.321						
d	1.214	1.214	1.214					
e	0.806	0.806	0.806	0.032				
f	1.001	1.001	3.650	0.202	0.021			
g	0.806	0.806	0.806	0.032	0.100	0.0023		
h	0.626	0.626	0.419	2.204	0.831	1.47	0.831	
i	1.001	3.650	1.001	0.202	0.0023	0.048	2.827	1.153

Note: a through i are the code number of the 9 comprehensive map patterns.

Table 6
Weights, weight variances, and the variances due to missing data of the 9 comprehensive map patterns.

Map patterns	W^+	$\sigma(W^+)$	W^-	$\sigma(W^-)$	σ(missing)
a	1.879	1.022	−0.0288	0.0346	1.682e−6
b	1.614	1.017	−0.0272	0.0346	1.159e−6
c	1.701	1.018	−0.0278	0.0346	1.315e−6
d	0.226	0.167	−0.0493	0.0418	1.326e−7
e	1.259	0.253	−0.105	0.0386	2.702e−6
f	0.330	0.201	−0.0547	0.0401	2.259e−7
g	0.249	0.251	−0.0334	0.0386	1.011e−7
h	1.649	0.102	−0.339	0.0501	1.298e−5
i	1.942	0.205	−0.158	0.0401	9.320e−6

Note: a through j denote the code number of the 9 comprehensive map patterns.

drawn: (a) the overall performance of the likelihood ratio and lift index is similarly good; and (b) the overall performance of the WofE model is a little bit worse than that of the likelihood ratio and lift index but a little bit better than that of the Youden index.

5. Conclusion and discussion

Mineral potential mapping is a key procedure for integrating multi-source geological map patterns to delineate mineral potential targets for minimizing the prospecting cost while maximizing the prospecting benefit in mineral exploration. In this study, prospecting cost and benefit are quantitatively defined and used to construct the Youden index, likelihood ratio, and lift index which can serve as mineral potential indicators to select map patterns, estimate mineral potentials, and assess mineral potential mapping performance. Based on these three mineral potential indicators, a four-step prospecting cost-benefit strategy for mineral potential mapping is developed and applied in the mineral potential mapping in a study area. The performance of different mineral potential indicators was assessed using both the ROC curve analysis and the diagrams of the three mineral potential indicators. The results show that (a) the likelihood ratio and lift index perform similarly well and (b) the WofE model performs a little bit worse than the likelihood ratio and lift index while a little bit better than the Youden index. Therefore, the prospecting cost-benefit strategy is a feasible paradigm for both mineral potential mapping and the performance assessment.

A useful way of assessing the performance of different mineral potential mapping models is to plot both the ROC curves and the mineral potential indicator diagrams. Different graphs can mutually complement in the model performance assessment. The case study illustrates that the ROC curves and the Youden diagrams cannot properly distinguish the performance of the WofE model and the Youden index but the cumulative lift charts and the likelihood ratio diagrams however can easily differentiate the performance of the two models.

In the case study, the 9 comprehensive map patterns which satisfy both pairwise and overall conditional independence assumption were used in the mineral potential mapping with the prospecting cost-

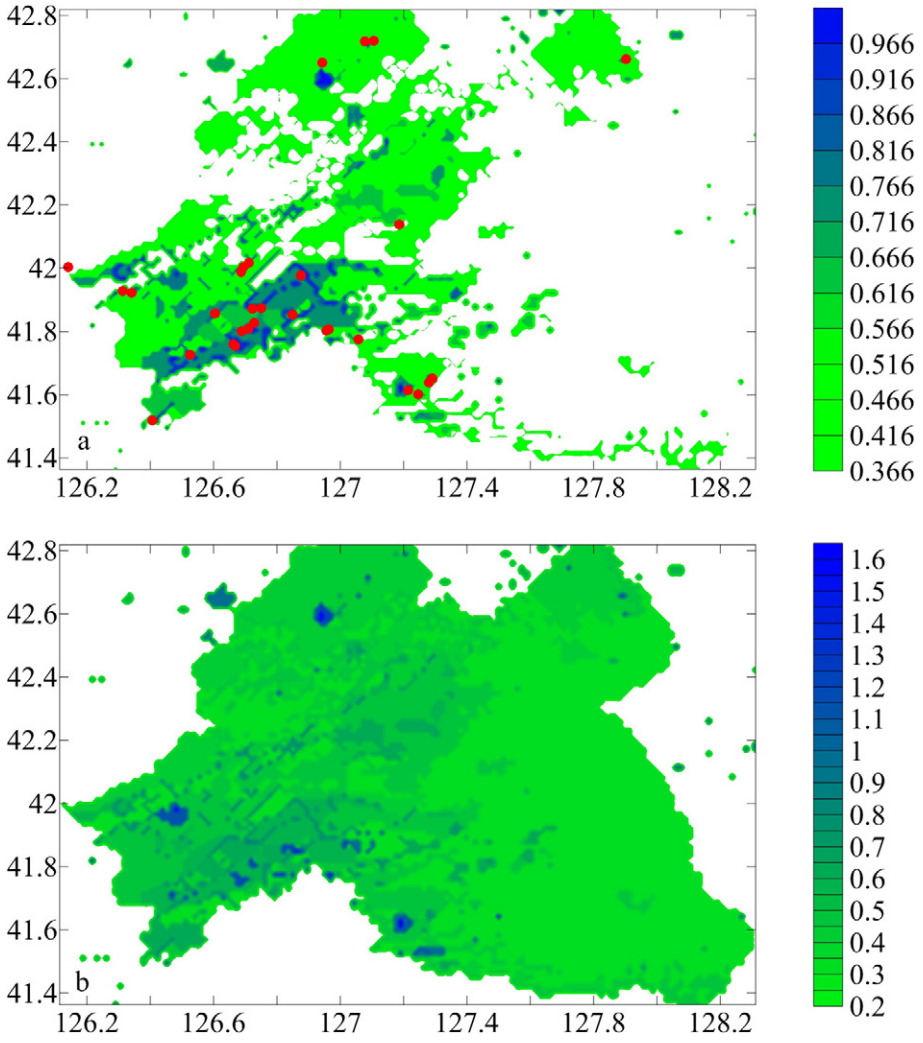

Fig. 4. Maps of posterior probability (a) and posterior probability deviation (b).

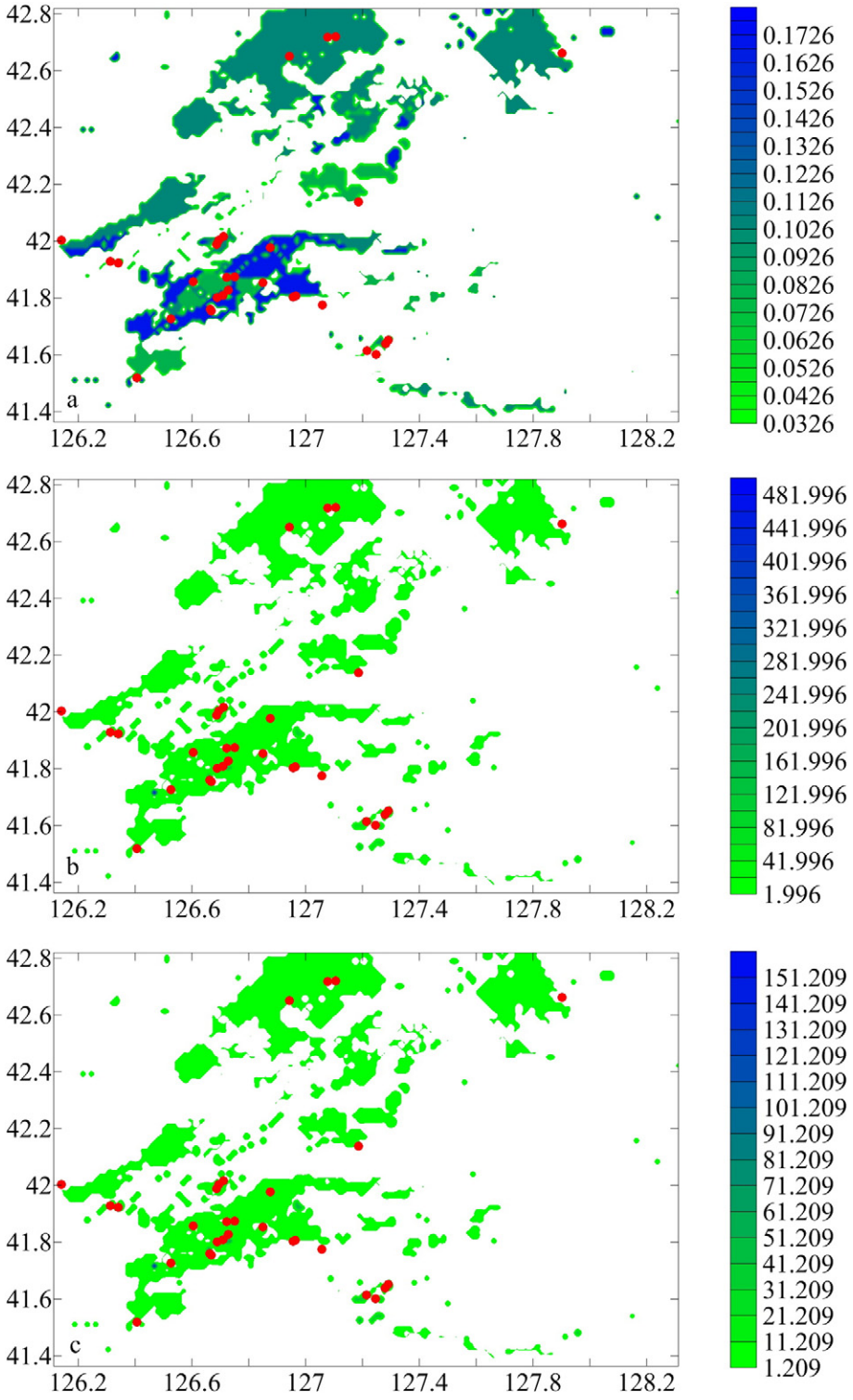

Fig. 5. Maps of Youden indexes (a), likelihood ratios (b) and lift indexes (c).

benefit strategy. This is only for facilitating the comparison between the performance of the three mineral potential indicators and the WofE model. In fact, the Youden index, likelihood ratio, and lift index require the map patterns used in mineral potential mapping to satisfy neither pairwise nor overall conditional independence assumption. Thus, the prospecting cost-benefit strategy can be directly applied to map the mineral potential of a study area without taking the conditional independence tests.

The likelihood ratio performs better than the Youden index in the case study. This is because only the 9 comprehensive map patterns are used in the mineral potential mapping. If the 23 selected map patterns are used, the likelihood ratio performs worse than the Youden index. The reason is that computing the likelihood ratio may lead to the situation where a value is divided by zero. In mineral potential mapping, increasing map patterns may generate many small-sized "unique" conditions whose fprates may equal to zero. In this case, the likelihood

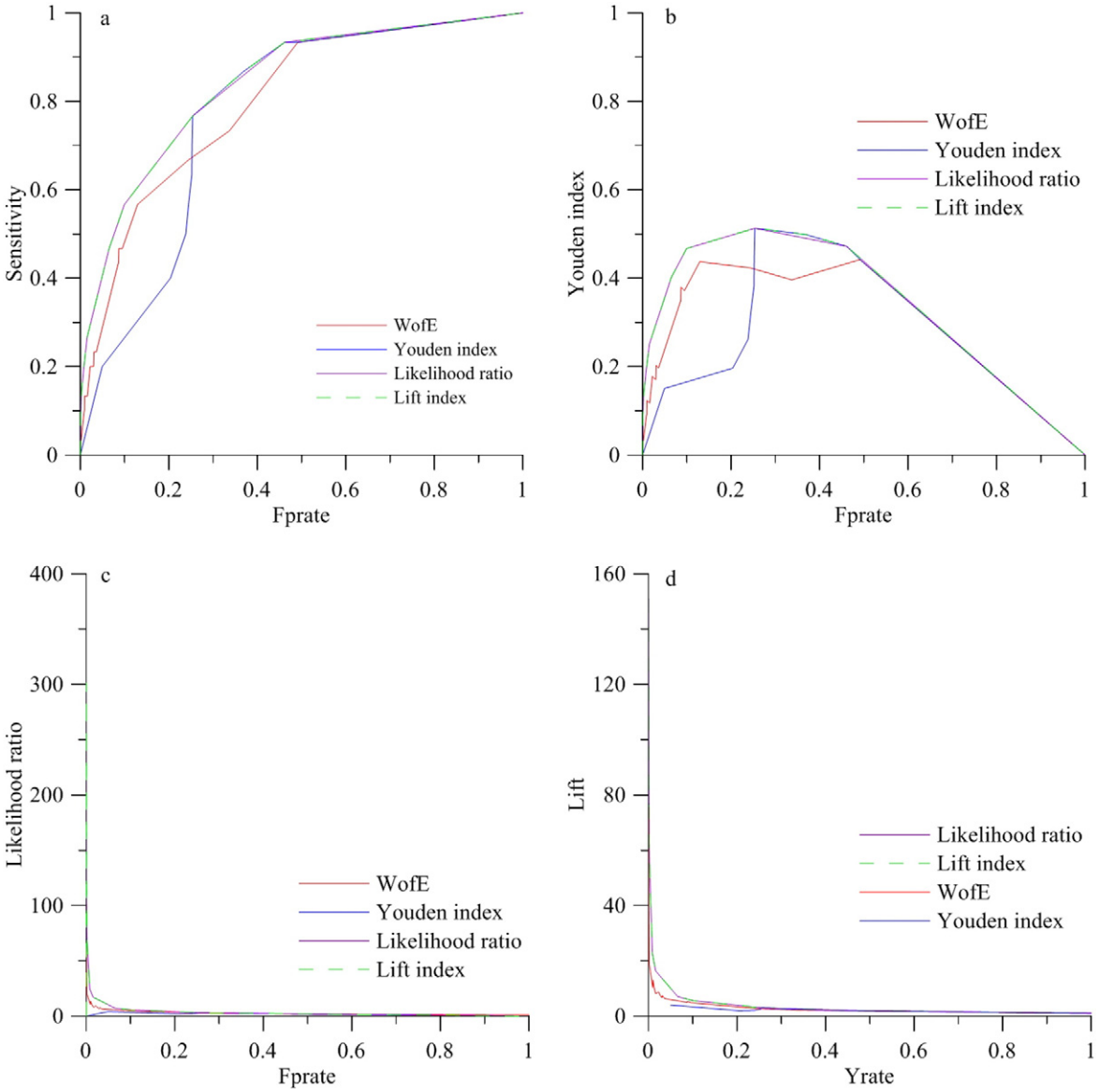

Fig. 6. ROC curves (a), Youden (b) and likelihood ratio (c) diagrams, and cumulative lift charts (d) of the four mineral potential indicators.

ratio is indefinite due to the situation where tprate is divided by zero-valued fprate. This is the drawback of the likelihood ratio in the application.

It should be pointed out that not only the three mineral potential indicators, discussed in this paper, but also other mineral potential indicators, such as discriminant power (Powers, 2012) and Cohen kappa (Powers, 2012), can be constructed on the basis of the prospecting cost and benefit, when applying the prospecting cost-benefit strategy to mineral potential mapping. Thus, the prospecting cost-benefit strategy can use a series of mineral potential indicators in both mineral potential mapping and model performance evaluation.

In mineral exploration, the applicability of the prospecting cost-benefit strategy for mineral potential mapping depends on whether the deposit-bearing and non-deposit-bearing cells can be defined in a study area. The exploration degree of a study area has a great influence on the exactness of the defined deposit-bearing and non-deposit-bearing cells. Thus, the prospecting cost-benefit strategy is suitable for application in a study area where the exploration degree is high and some mineral deposits have been discovered. In practice, it is very often the case that some deposit-bearing cells (i.e., the cells contain undiscovered mineral deposits) are defined incorrectly as non-deposit-bearing cells. However, the overwhelming majority of the grid cell population belong to non-deposit-bearing cells in a study area. The incorrectly defined deposit-bearing cells are always only a tiny fraction of the whole non-deposit-bearing cells. Thus, the inexactness of defined deposit-bearing and non-deposit-bearing cells does not significantly affect the result of mineral potential mapping of the prospecting cost-benefit strategy.

The prospecting cost-benefit strategy usually uses binary map patterns to map the mineral potential of a study area. It can also use categorical map patterns in mineral potential mapping if required. In this case, the "unique condition" searching algorithm needs to be slightly modified. It is quite expected that the prospecting cost-benefit strategy can be modified to optimally combine a set of continuous map patterns into a mineral potential map. The detailed algorithm is now under investigation.

Table 7
AUC, *AUL*, and Z_{AUC} values for the four mineral potential mapping results.

Indicator	*AUC*	*AUL*	Z_{AUC}
Post-probability	0.8002	0.799	6.164
Youden index	0.773	0.772	5.399
Likelihood ratio	0.841	0.840	7.542
Lift index	0.841	0.840	7.542

The prospecting cost-benefit strategy can be straightforwardly used to assess the mineral potential mapping performance of different detection techniques which may individually provide potentially different patterns of prospectivity. Different detection techniques may use different sampling techniques which often generate different cell populations. Thus, the deposit-bearing and non-deposit-bearing cells for each detection technique need to be identified individually. But their ROC curves (or lift charts, or Youden and likelihood ratio diagrams) can be plotted in the same space for their performance assessment.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

The authors are grateful to the two reviewers for their constructive comments which greatly improved this manuscript. This research is supported by the National Natural Science Foundation of China (Grant nos. 41272360, 41472299, and 61133011).

References

Abedi, M., Norouzi, G.H., Bahroudi, A., 2012a. Support vector machine for multi-classification of mineral potential areas. Comput. Geosci. 46 (3), 272–283.
Abedi, M., Torabi, S.A., Norouzi, G.H., Hamzeh, M., Elyasi, G.R., 2012b. PROMETHEE II: a knowledge-driven method for copper exploration. Comput. Geosci. 46 (3), 255–263.
Abedi, M., Torabi, S.A., Norouzi, G.H., Hamzeh, M., 2012c. ELECTRE III: a knowledge-driven method for integration of geophysical data with geological and geochemical data in mineral potential mapping. J. Appl. Geophys. 87 (12), 9–18.
Abedi, M., Gholami, A., Norouzi, G.H., 2013a. A stable downward continuation of airborne magnetic data: a case study for mineral potential mapping in central Iran. Comput. Geosci. 52 (1), 269–280.
Abedi, M., Norouzia, G.H., Fathianpourb, N., 2013b. Fuzzy outranking approach: a knowledge-driven method for mineral potential mapping. International Journal of Applied Earth Observation and Geoinformation 21 (4), 556–567.
Agterberg, F.P., 1974. Automatic contouring of geological maps to detect target areas for mineral exploration. Math. Geol. 6 (4), 373–395.
Agterberg, F.P., 1989. LOGDIA-FORTRAN 77 program for logistic regression with diagnostics. Comput. Geosci. 15 (4), 599–614.
Agterberg, F.P., 1990. Combining indicator patterns for mineral resource evaluation: China University of Geosciences. Proceedings of International Workshop on Statistical Prediction of Mineral Resources, Wuhan, China. 1, pp. 1–15.
Agterberg, F.P., 1992. Combining indicator patterns in weights of evidence modeling for resource evaluation. Nonrenewable Resources 1 (1), 39–50.
Agterberg, F.P., Bonham-Carter, G.F., Wright, D.F., 1990. Statistical pattern integration for mineral exploration. In: Gaal, G., Merriam, D.F. (Eds.), Computer Applications for Mineral Exploration in Resource Exploration. Pergamon Press, Oxford, pp. 1–21.
Agterberg, F.P., Bonham-Carter, G.F., 1999. Logistic regression and weights of evidence modeling in mineral exploration. Proceedings of the 28th International Symposium on Applications of Computer in the Mineral Industry (APCOM), Golden, Colorado, pp. 483–490.
Agterberg, F.P., Cheng, Q.M., 2002. Conditional independence test for weights-of-evidence modeling. Nat. Resour. Res. 11 (4), 249–255.
An, P., Moon, W.M., 1993. Evidential reasoning structure for integrating geophysical, geological and remote sensing data. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2141–2144.
An, P., Moon, W.M., Rencz, A.N., 1991. Application of fuzzy theory for integration of geological, geophysical and remotely sensed data. Can. J. Explor. Geophys. 27 (1), 1–11.
Anjum, S., 2014. Composite indicators for data mining: a new framework for assessment of prediction classifiers. Journal of Economics, Business and Management 2 (1), 62–67.
Barreno, M., Cardenas, A.A., Tygar, J.D., 2008. Optimal ROC curve for a combination of classifiers. Advances in Neural Information Processing Systems (NIPS). MIT Press, pp. 57–64.
Bekkar, M., Djemaa, H.K., Alitouche, T.A., 2013. Evaluation measures for models assessment over imbalanced data sets. Journal of Information Engineering and Applications 3 (10), 27–38.
Bergmann, R., Ludbrook, J., Spooren, W.P.J.M., 2000. Different outcomes of the Wilcoxon–Mann–Whitney test from different statistics packages. Am. Stat. 54 (1), 72–77.
Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. Photogramm. Eng. Remote. Sens. 54 (11), 1585–1592.
Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1989. Weights-of-evidence modelling: a new approach to mapping mineral potential. In: Agterberg, F.P., Bonham-Carter, G.F. (Eds.), Statistical Applications in the Earth Sciences. Paper 89–9. Geological Survey of Canada, 171–183.
Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G., 2000. Artificial neural networks: a new method for mineral potential mapping. Aust. J. Earth Sci. 47 (4), 757–770.
Carranza, E.J.M., 2010. Improved wildcat modelling of mineral potential. Resour. Geol. 60 (2), 129–149.
Carranza, E.J.M., Hale, M., 2001a. Geologically constrained fuzzy mapping of gold mineralization potential, Baguio district, Philippines. Nat. Resour. Res. 10 (2), 125–136.
Carranza, E.J.M., Hale, M., 2001b. Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines. Explor. Min. Geol. 10 (3), 165–175.
Carranza, E.J.M., Hale, M., 2002a. Where are porphyry copper deposits spatially localized? A case study in Benguet province, Philippines. Nat. Resour. Res. 11 (1), 45–59.
Carranza, E.J.M., Hale, M., 2002b. Wildcat mapping of gold potential, Baguio district, Philippines. Transactions Institute of Mining and Metallurgy (Applied Earth Science) 111 (2), 100–105.
Carranza, E.J.M., Hale, M., 2003. Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. Ore Geol. Rev. 22 (1), 117–132.
Carranza, E.J.M., Mangaoang, J.C., Hale, M., 1999. Application of mineral exploration models and GIS to generate mineral potential maps as input for optimum land-use planning in the Philippines. Nat. Resour. Res. 8 (2), 165–173.
Carranza, E.J.M., Woldai, T., Chikambwe, E.M., 2005. Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi district, Zambia. Nat. Resour. Res. 14 (1), 47–63.
Chen, C.H., Dai, H.Z., Liu, Y., He, B.B., 2011. Mineral potential mapping integrating multi-source geology spatial data sets and logistic regression modeling. Proceedings of IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), pp. 214–217.
Chen, Y.L., 2003. Indicator pattern combination for mineral potential mapping with the general C-F model. Math. Geol. 35 (3), 301–321.
Chen, Y.L., 2004. MRPM: three visual basic programs for mineral resource potential mapping. Comput. Geosci. 30 (9–10), 969–983.
Chen, Y.L., 2014. Mineral potential mapping with a restricted Boltzmann machine. Ore Geol. Rev. in press.
Chen, Y.L., Lu, L.J., Li, X.B., 2014. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. J. Geochem. Explor. 140 (1), 56–63.
Cheng, Q., 1995. The perimeter-area fractal model and its application to geology. Math. Geol. 27 (1), 69–82.
Cheng, Q., 2000. GeoData Analysis System (GeoDAS) for mineral exploration: user's guide and exercise manual. Material for the Training Workshop on GeoDAS held at York University, Toronto, Canada. 1, p. 204.
Cheng, Q., 2006. Singularity-generalized self-similarity-fractal spectrum (3S) model. Earth Sciences—China University of Geosciences 31 (3), 337–348 (In Chinese with English Abstract).
Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan province, China. Ore Geol. Rev. 32 (1), 314–324.
Cheng, Q., 2008. Non-linear theory and power-law models for information integration and mineral resources quantitative assessments. Math. Geol. 40 (5), 503–532.
Cheng, Q.M., Chen, Z.J., Khaled, A., 2007. Application of fuzzy weights of evidence method in mineral resource assessment for gold in Zhenyuan District, Yunnan Province, China. Earth Science — Journal of China University of Geosciences 32 (2), 175–184 (In Chinese with English Abstract).
Cheng, Q.M., Zhao, P.D., Chen, J.G., Xia, Q.L., Chen, Z.J., Zhan, S.Y., Xu, D.Y., Xia, X.Y., Wan, W.L., 2009a. Application of singularity theory in prediction of tin and copper mineral deposits in Gejiu District, Yunnan, China: weak information extraction and mixing information decomposition. Earth Science — Journal of China University of Geosciences 34 (2), 232–242 (In Chinese with English Abstract).
Cheng, Q.M., Zhao, P.D., Zhan, S.Y., Xia, Q.L., Chen, Z.J., Chen, J.G., Xu, D.Y., Wan, W.L., 2009b. Application of singularity theory in prediction of tin and copper mineral deposits in Gejiu District, Yunnan, China: information integration and delineation of mineral exploration targets. Earth Science — Journal of China University of Geosciences 34 (2), 243–252 (In Chinese with English Abstract).
Chung, C.F., Moon, W.M., 1990. Combination rules of spatial geoscience data for mineral exploration. Geoinformatics 2, 159–169.
D'Ercole, C., Groves, D.I., Knox-Robinson, C.M., 2000. Using fuzzy logic in a geographic information system environment to enhance conceptually based potential analysis of Mississippi valley-type mineralization. Aust. J. Earth Sci. 47 (5), 913–927.
Galuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environ. Geol. 52 (5), 861–870.
Hawkes, H.E., Webb, J.S., 1962. Geochemistry in Mineral Exploration. Harper and Row, New York, NY.
Hernandez-Orallo, J., 2013. ROC curves for regression. Pattern Recogn. 46 (12), 3395–3411.
Johnson, N.P., 2004. Advantages to transforming the receiver operating characteristic (ROC) curve into likelihood ratio co-ordinates. Stat. Med. 23 (14), 2257–2266.
Karimi, M., Menhaj, M.B., Mesgari, M.S., 2008. Preparing mineral potential map using fuzzy logic in GIS environment. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing. I, pp. 1263–1270 (Part B8).
Knox-Robinson, C.M., 2000. Vectorial fuzzy logic: a novel technique for enhanced mineral potential mapping, with reference to the orogenic gold mineralization potential of the Kalgoorlie Terrane, Western Australia. Aust. J. Earth Sci. 47 (5), 929–941.
Leite, E.P., de Souza Filho, C.R., 2009a. Artificial neural networks applied to mineral potential mapping for copper–gold mineralization in the Carajás Mineral Province, Brazil. Geophys. Prospect. 57 (6), 1049–1065.
Leite, E.P., de Souza Filho, C.R., 2009b. Probabilistic neural networks applied to mineral potential mapping for platinum group elements in the Serra Leste region, Carajás Mineral Province, Brazil. Comput. Geosci. 35 (3), 675–687.
Mansour, Z., Ali, P., Mahdi, Z., 2009. A computational optimized extended model for mineral potential mapping based on WofE method. Am. J. Appl. Sci. 6 (2), 200–203.

Li, B., Yang, Z., Wang, Y., 2010. Geological characteristics and genesis of Huanggoushan and Banmiaozi gold deposits in Laoling metallogenic belt of southern Jilin. Global Geology 29 (3), 392–399 (In Chinese with English Abstract).
Li, H., 2009. On the geological features of Huanggoushan Au deposit. Jilin Geology 28 (2), 29–31 (In Chinese with English Abstract).
Liu, W., Deng, J., Chu, X.L., Zhai, Y.S., Xu, G.Z., Li, X.J., 2000. Characteristics and geological background of formation of large and giant ore deposits within the northern margin of the north China platform. Prog. Geophys. 15 (2), 67–78 (In Chinese with English Abstract).
Molan, Y.E., Behnia, P., 2013. Potential mapping of Pb–Zn SEDEX mineralization using remote-sensing data in the Behabad area, Central Iran. Int. J. Remote Sens. 34 (4), 1164–1179.
Moon, W.M., 1989. Integration of remote sensing and geophysical/geological data using Dempster-Shafer approach. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 838–841.
Moon, W.M., 1990. Integration of geophysical and geological data using evidence theory function. IEEE Trans. Geosci. Remote Sens. 28 (4), 711–720.
Moon, W.M., 1993. On mathematical representation and integration of multiple spatial geoscience data sets. Can. J. Remote. Sens. 19 (1), 63–67.
Moon, W.M., So, C.S., 1995. Information representation and integration of multiple sets of spatial geoscience data. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2141–2144.
Nykänen, V., 2008. Radial basis functional link nets used as a potential mapping tool for orogenic gold deposits within the central Lapland greenstone belt, northern Fennoscandian Shield. Nat. Resour. Res. 17 (1), 29–48.
Nykänen, V., Groves, D.I., Ojala, V.J., Gardoll, S.J., 2008. Combined conceptual/empirical potential mapping for orogenic gold in the northern Fennoscandian Shield, Finland. Aust. J. Earth Sci. 55 (1), 39–59.
Pan, G.C., 1996. Extended weights of evidence modeling for the pseudo-estimation of metal grades. Nonrenewable Resources 5 (1), 53–76.
Porwal, A., Carranza, E.J.M., Hale, M., 2003. Artificial neural networks for mineral potential mapping. Nat. Resour. Res. 12 (3), 155–171.
Porwal, A., Carranza, E.J.M., Hale, M., 2006a. A hybrid fuzzy weights-of-evidence model for mineral potential mapping. Nat. Resour. Res. 15 (1), 1–15.
Porwal, A., Carranza, E.J.M., Hale, M., 2006b. Bayesian network classifiers for mineral potential mapping. Comput. Geosci. 32 (1), 1–16.
Powers, D.M.W., 2012. The problem with Kappa. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 345–355.
Rad, A.R.J., Busch, W., 2011. Porphyry copper mineral potential mapping using interval valued fuzzy sets TOPSIS method in central Iran. J. Geogr. Inf. Syst. 3 (3), 312–317.
Ruopp, M.D., Perkins, N.J., Whitcomb, B.W., Schisterman, E.F., 2008. Youden index and optimal cut-point estimated observations affected by a lower limit of detection. Biom. J. 50 (3), 419–430.
Skabar, A., 2003. Mineral potential mapping using feed-forward neural networks. Proceedings of the International Joint Conference on Neural Networks 3. IEEE Press, Portland, OR, the United States, pp. 1814–1819.
Skabar, A., 2007. Mineral potential mapping using Bayesian learning for multilayer perceptrons. Math. Geol. 39 (5), 439–451.
Su, X.J., Zang, X.Y., 2010. Geological characteristics and genetic analysis of Banmiaozi gold deposit in Baishan city, Jilin province. Contributions to Geology and Mineral Resources Research 25 (4), 326–330 (In Chinese with English Abstract).
Swets, J.A., 1996. Signal Detection Theory and ROC Analysis in Psychology and Diagnostics. Collected Papers. Lawrence Erlbaum Associates, Mahwah, NJ.
Tangestani, M.H., Moore, F., 2001. Porphyry copper potential mapping using the weights-of-evidence model in a GIS, northern Shahr-e-Babak, Iran. Aust. J. Earth Sci. 48 (5), 913–927.
Tukey, J.W., 1997. Exploratory Data Analysis. Addison-Wesley Publishing Company, Massachusetts, Reading.
Wu, F., Lin, J., Wilde, S.A., Zhang, Q., Yang, J., 2005. Nature and significance of early Cretaceous giant igneous event in eastern China. Earth Planet. Sci. Lett. 233 (1-2), 103–119.
Yang, Y.C., Feng, B.Z., Liu, P.E., 2001. Dahenglu type of cobalt deposit in Laoling area, Jilin Province—a SedEx deposit with late reformation. Journal of Changchun University of Science and Technology 31 (1), 40–45 (In Chinese with English Abstract).
Yang, Y.C., Ye, S.Q., Feng, B.Z., 1999. The Huanggoushan typed hot-water deposition and superimposed reformation gold deposit in Laoling mineralization belt of South Jilin province. Gold 6 (20), 1–4 (In Chinese with English Abstract).
Zou, K.H., O'Malley, A.J., Mauri, L., 2007. Receiver operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 115 (5), 654–657.
Zuo, R., Carranza, E.J.M., 2011. Support vector machine: a tool for mapping mineral potential. Comput. Geosci. 37 (12), 1967–1975.
Zuo, R., Cheng, Q., Agterberg, F.P., 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. Geochem. Explor. 101 (3), 355–357.