Accepted Manuscript

Application of ant colony algorithm to geochemical anomaly detection

Yongliang Chen, Aijun An

PII: S0375-6742(15)30098-4
DOI: doi: 10.1016/j.gexplo.2015.11.011
Reference: GEXPLO 5668

To appear in: *Journal of Geochemical Exploration*

Received date: 1 July 2015
Revised date: 24 November 2015
Accepted date: 25 November 2015

Application of ant colony algorithm to geochemical anomaly detection

[1]Yongliang Chen, [2]Aijun An

[1]Institute of Mineral Resources Prognosis on Synthetic Information, Jilin University, Changchun, Jilin 130026, China

[2]Department of Electrical Engineering and Computer Science, York University, Toronto, Ontario M3J 1P3, Canada

Corresponding Author: chenyongliang2009@hotmail.com

ABSTRACT

The methods for geochemical anomaly detection are commonly based on statistical models that require assumption of the sample population to satisfy a particular distribution. In practice, the assumption of a particular distribution may degrade the performance of geochemical anomaly detection. In this paper, an ant colony algorithm is used to detect geochemical anomalies. The new method does not require assumption that the geochemical data satisfy a particular distribution. Applying this method to detect geochemical anomalies, we only need to put a number of "virtual ants" randomly into a geochemical grid map and let each ant complete its iterative search process. When the algorithm gets converged, the ants tend to aggregate in the geochemical anomalous regions where geochemical element concentration values are significantly greater than surrounding background. The number of times each grid point is visited by ants can be recorded in ant density data for geochemical anomaly identification. The ant density data are almost not affected by regional variations of geochemical background, thus they are suitable for identifying geochemical anomalies using a threshold method. As an illustration, the ant colony algorithm is applied to detect geochemical anomalies in interpolated concentration data of Au, Ag, Cu, Pb, and Zn in the Altay district in northern Xinjiang in China. The results show that the ant colony algorithm can properly identify geochemical anomalies. Anomalies detected by the ant colony algorithm occupy 9.5% of the study area and contain 36% of the known mineral deposits; and anomalies identified using the Youden index method occupy 16.4% of the study area and contain 56% of the known mineral deposits. Therefore, the ant colony algorithm can serve as a feasible swarm intelligence paradigm for geochemical anomaly detection.

Keywords: Ant colony algorithm; Swarm intelligence; Heuristic search; The Youden index; Geochemical anomaly detection

1. Introduction

Swarm intelligence has become a new artificial intelligence field inspired by insect swarms that display intelligence on the swarm level with very simple interacting individuals (Zhuang, 2004). In many aspects, the self-organization of insects into a swarm are similar to the self-organization of neurons into brain-like structures. These features could lead to important developments in pattern recognition systems, where perceptive capabilities can emerge and evolve from the interaction of many simple local rules. The emergence of a collective behavior pattern is largely controlled by competition among all possible behavior patterns, in which the pattern most fitting for the environment will prevail.

The ant colony algorithm (Dorigo, 1992) is one of the swarm intelligent models. It is a parallel computational paradigm that allows exploitation of positive feedback as a search mechanism. The collective behavior that emerges reinforces itself, where the more ants follow a trail, the more attractive that trail becomes for being followed. This heuristic was first applied to the travelling salesman problem (Dorigo, 1992; Dorigo and Gambardella, 1997) and then extended to other optimization problems such as quadratic assignment (Maniezzo et al., 1994; Stutzle and Hoos, 1998; Liu, 2007), vehicle routing (Bullnheimer et al., 1997; Gambardella et al., 1999), and graph coloring (Costa and Hertz, 1997). Over the past 10 years, the ant colony algorithm was extended to deal with digital image habitats, in which virtual ants were able to react to their environment and perceive it. The evolution of pheromone fields suggest that artificial ant colonies could react and adapt to any type of digital habitat. Since then several studies have been conducted to apply this recent paradigm to real case problems with successful results (Ramos and Almeida, 2000).

Geochemical grid maps and digital images have similar features, thus some digital image processing methods are also suitable for processing a geochemical grid map. Inspired by image feature extraction based on the ant colony algorithm, we used the ant colony algorithm to detect geochemical anomalies in a map of interpolated element concentration data. This new approach is quite different from the commonly used geochemical anomaly detection methods. Many of the commonly-used methods are based on statistical models that analyze geochemical data with assumption of a particular distribution; for example, the mean ± 2σ method (Galuszka, 2007; Hawkes and Webb, 1962) and univariate analysis (Govett et al., 1975; Singer and Kouda, 2001) are used to handle data that exhibit a Gaussian distribution, multivariate data analysis

(Cheng et al., 1996; Garrett, 1989; Miesch, 1981; Stanley, 1988; Stanley and Sinclair, 1989; Geranian et al., 2015) deal with data with multivariate Gaussian distribution, and fractal and multifractal methods (Cheng, 1995, 2000, 2006, 2007, 2008; Cheng and Agterberg, 1995; Cheng et al., 1994, 2000; Deng et al., 2010; Li and Cheng, 2004; Zuo et al., 2009; He et al., 2013; Luz et al., 2014) cope with data that follow a power law distribution.

Chen et al. (2014) developed a multivariate geochemical anomaly identification method based on a restricted Boltzmann machine and got successful results. The method can identify multivariate geochemical anomalies from data with an unknown distribution. However, it is unsuitable for detecting univariate geochemical anomalies. As alternative, this paper provides an ant colony algorithm for identifying univariate geochemical anomalies from interpolated element concentration data with unknown distribution. It is a local heuristic search paradigm with the following characteristics: (a) it is versatile, in that it can be applied to similar versions of the same problem; (b) it is robust, i.e., it can be applied with only minimal changes to other problems; and (c) it is a population based approach. The last property allows exploitation of positive feedback as a search mechanism. It also makes the algorithm amenable to parallel implementations. This method uses an iterative search process to transform geochemical element concentration data into corresponding ant density data, from which geochemical anomalies can be identified using a threshold method. The iterative search process is completed through a set of parallel-executed local heuristic search steps that are not significantly influenced by regional variations of geochemical background. Thus, this method can separate geochemical anomalies from a regionally-variable geochemical background.

For demonstration purposes, the ant colony algorithm is applied to detect geochemical anomalies from interpolated concentration data of Au, Ag, Cu, Pb, and Zn in the Altay district in northern Xinjiang in China. This paper therefore seeks to present the application of the ant colony algorithm in geochemical anomaly detection. We compare the performance of the ant colony algorithm with that of the Youden index as applied to geochemical anomaly detection.

2. Overview on the ant colony algorithm

The ant colony algorithm proposed by Dorigo (1992) aims to search for an optimal path in a graph, based on the behavior of ants seeking a path between their colony and a source of food. In the natural world, ants initially wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely to follow the trail, returning and reinforcing it if they eventually find food. Over time, however, the pheromone trail starts to evaporate. The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate. A short path, by comparison, gets marched over more frequently, and thus the pheromone density becomes higher on shorter paths than on longer ones. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, exploration of the solution space would be constrained. Thus, when one ant finds a good (i.e., short) path from the colony to a food source, other ants are more likely to follow that path, and positive feedback eventually leads to all the ants following a single path. The idea of the ant colony algorithm is to mimic this behavior with "virtual ants" walking around the graph representing the problem to solve.

In the ant colony algorithm, an ant is a simple computational agent that iteratively constructs a solution for the problem at hand. The intermediate solutions are called solution states. At each iteration, an ant moves from state s_i to adjacent state s_j, corresponding to a more complete intermediate solution. Thus, each ant k computes set A_k of feasible expansions to its current state in each iteration, and moves to one of these in probability. For ant k at iteration t, the probability of moving from state s_i to adjacent state s_j depends on the combination of two values, i.e., the attractiveness $\tau_{ij}(t)$ of the move, as computed by some heuristic indicating the priori desirability of that move and the pheromone trail η_{ij} of the move, indicating how proficient it has been in the past to make that particular move. In general, ant k moves from state s_i to adjacent state s_j with the following probability (Dorrigo et al., 1999)

$$p_{ij}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{s_l \in A_k} [\tau_{il}(t)]^{\alpha} \cdot [\eta_{il}]^{\beta}} & if\ s_j \in A_k \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

where α and β are positive constants used to express the importance of $\tau_{ij}(t)$ and η_{ij}, respectively, in the above probability computation.

At each iteration t, pheromone trails are updated usually when all ants have completed their solution, increasing or decreasing the pheromone trails corresponding to moves that were part of "good" or "bad" solutions, respectively. The pheromone trail on the path from state s_i to adjacent sate s_j is updated as (Dorigo et al., 1999)

$$\tau_{ij}(t+1) = \rho\tau_{ij}(t) + \Delta\tau_{ij}(t) \qquad (2)$$

where $\tau_{ij}(t)$ and $\tau_{ij}(t+1)$ are the pheromone trail on the path from state s_i to adjacent sate s_j before and after the update, respectively; ρ is the pheromone evaporation coefficient expressed by a constant within interval (0, 1); and $\Delta\tau_{ij}(t)$ is the following pheromone trail updated by all the ants (Dorigo et al., 1999)

$$\Delta\tau_{ij}(t) = \sum_{k}^{m} \Delta\tau_{ij}^{k}(t) \quad (3)$$

where m is the number of ants; and $\Delta\tau_{ij}^{k}(t)$ is the following pheromone trail updated by ant k (Dorigo et al., 1999)

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} 1/L_k & \text{if ant } k \text{ moves from } s_i \text{ to } s_j \text{ at step } t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $1/L_k$ is the reciprocal of the path length experienced by ant k. Therefore, the shorter the path is, the larger value the pheromone trail is enhanced.

Before starting the ant colony algorithm, the following parameters need to be initialized: (a) the set of starting states; (b) the set of goal states; (c) the number of ants; (d) the termination condition of iterations for each ant; and (e) the definition of path length, i.e., the cost of the solution.

3. Geochemical anomaly detection using ant colony algorithm

3.1 Theory

The ant colony algorithm can be used to extract image features (Ramos and Almeida, 2000; Zhuang, 2004). A grid element map, in which element concentration values are recorded by interpolation in regularly spaced grid points, is similar to a digital image. A grid element map can be viewed as a two-dimensional discrete space comprised of regularly spaced grid points that correspond to pixels in a digital image. Detecting geochemical anomalies in a grid element map is similar to extracting image features from a digital image. Thus, the ant colony algorithm for image feature extraction can be modified to detect geochemical anomalies in a grid element map.

The Golden Software Surfer can be used to generate grid element data by interpolating element concentration values at regularly-spaced grid points based on element concentration data collected from irregularly scattered samples. The grid element data are usually stored in a GRID file, which can be processed as a digital image, in which the ant colony algorithm can be used to detect geochemical anomalies.

Detecting geochemical anomalies in a grid element map aims to identify subareas comprised of grid points at which the element concentration values are significantly greater than the surrounding background. Suppose that a virtual ant colony resides in a grid element map where each ant initially occupies randomly one grid point. Then, the anomalous grid points in the map are the goal points which all the ants search for heuristically among the adjacent grid points starting from their randomly occupied grid points. This process is similar to utilizing a virtual ant colony to react and adapt accordingly to different digital image habitats. The successful results of digital image feature extraction using the ant colony algorithm (Zhuang, 2004) provide a useful reference for developing it into a geochemical anomaly detect method.

The ant colony algorithm for detecting geochemical anomalies in a grid element map can be designed as an iterative heuristic search process from the grid points occupied initially by ants to the anomalous grid points. At each iteration, all the ants in a colony simultaneously move one step from the current grid points to the adjacent grid points and each ant tends to choose the neighboring grid point with the highest favorability. The favorability of a neighboring grid point is determined by both the pheromone trail intensity on the path from the current grid point to the neighboring grid point and the interpolated element concentration value of the neighboring grid point.

After each iteration, the pheromone trail intensity on the path from each grid point to its each neighbor is updated in order to record the accumulated experience of the ant colony during the iterative search process. The modified quantity of the pheromone trail intensity is determined by both the difference between the element concentration values of two adjacent grid points and the number of the ants that have experienced the path between two grid points in the current iteration.

In order to avoid visiting one grid point repeatedly during the iterative search process, each ant is required to memorize the grid points it has visited recently. The data-structure used for recording these visited grid points is called Taboo List. The length of Taboo List is determined upon the practical situation. In edge detection of a digital image, if several tiny edges exists in a digital image, the length of Taboo List is defined short so that those tiny edges can be detected properly. Similarly, the length of Taboo List should be defined short if small geochemical anomalies are expected to be properly identified in a grid element map. A small geochemical anomaly usually consists of only a few of anomalous grid points. In our case study, a geochemical anomaly is regarded as small one if it is comprised of less than four anomalous grid points.

The pheromone trail intensity on a path is required to decrease automatically over time. This mandatory rule can make the pheromone trail intensity on the path maintain at a reasonable level during the iterative search process to prevent it from converging too quickly on a sub-optimal region. In image feature extraction, a pheromone evaporation coefficient within (0, 1) is empirically determined to describe the decreasing speed of pheromone trail intensity on a path. The same way can be used in geochemical anomaly detection.

After the algorithm gets converged, all the ants in the colony become stationary during iteration and tend to gather in geochemical anomalies in the grid element map. In this case, whether a grid point belongs to a geochemical anomaly can be determined as follows: a grid point belongs to an anomaly if the number of times it has been visited by ants is more than an empirically determined threshold value. The thing to stress here is that the number of iterations must be large enough (e.g., more than 3000) so that the ant colony algorithm can get completely converged.

3.2 Algorithm for geochemical anomaly detection

The ant colony algorithm for geochemical anomaly detection in a grid element map comprises: (a) initialization; (b) stepwise heuristic search; (c) pheromone trail update; and (d) geochemical anomaly identification.

During initialization, m ants are randomly put into the grid element map and the pheromone trail on the path from each grid point to its each neighbor is initialized as the same small positive constant value. In image feature extraction, the allowed maximum number of ants put into an image is equal to the number of pixels in the image and the optimal number of ants is usually chosen approximately equal to the square root of the number of pixels in the image (Zhuang, 2004). Similarly in geochemical anomaly detection, if the grid element map is $M \times N$ in size and m ants are put into it, then m must satisfy $m \leq M \times N$.

In order to find the optimal number of ants for the algorithm initialization in geochemical anomaly detection, 10, 20, 123, 500, and 1000 ants were tested to initialize the ant colony algorithm in Au anomaly detection in our case study. The arithmetic average of the Au concentration values collected from the grid points occupied by ants in each iteration should become larger and larger and eventually tend to a stable value after the algorithm gets converged. This average value is called average Au concentration value for monitoring the iterative search process in our case study. The diagrams of average Au concentration value varying with iterations for the five different number of ants are shown in Fig. 1. These diagrams show that the iterative search process converges to (a) around value 10 given $m \geq 500$, (b) a value between 9 and 11 given $m \leq 20$, and (c) a value more than 12 given $m = 123 \approx \sqrt{151 \times 100}$. Here, 151×100 is the size of geochemical grid element map in our case study. When the iterative search process has been completed, the bigger the converged value is, the more number of ants gather in anomalous areas. Therefore, using different number of ants to initialize the ant colony algorithm affects its Au anomaly detection performance. The optimal number of ants for the algorithm initialization is approximately equal to $\sqrt{M \times N}$. This is the same as the conclusion in image feature extraction.

In the stepwise heuristic search, Taboo List TL_k of ant k records the grid points that ant k has visited recently. The TL_k is used to define, for ant k, the set of neighboring grid points that ant k is located at grid point i still has to visit. By exploiting TL_k, ant k can build feasible solutions, that is, it can avoid to visit a grid point twice.

To investigate how a given length of Taboo List impacts the algorithm performance in geochemical anomaly detection, we used seven different lengths of Taboo List. These Taboo Lists can record respectively 10, 20, 30, 40, 50, 60, and 70 grid points that have been visited recently by an ant during the iterative search process. Fig. 2 shows that the average Au concentration value varies with iterations when different lengths of Taboo List are used in the ant colony algorithm. These diagrams reveal that the iterative search process converges to (a) a higher value after more than 1000 iterations using length of Taboo List less than 30, (b) a lower value after less than 500 iterations using length of Taboo List more than 50, and (c) a moderate value (which is the same as the converged value given m = 123) after 800 to 1000 iterations using length of Taboo List between 30 and 50. Therefore, the longer the length of Taboo List is, the faster the iterative search process gets converged and the lower value it converges to. The optimal length of Taboo List should be between 30 and 50 in Au anomaly detection. A shorter Taboo List can make the iterative search process converge to a higher value, which results in all the ants tending to gather around anomaly centers and small anomalies after the algorithm gets converged. If one wants to locate anomaly centers or detect small anomalies in geochemical exploration, the length of Taboo List should be less than 30. However, using a Taboo List of length more than 50 is always not suggested for Au anomaly detection.

In a grid element map, weak fluctuation of interpolated element concentration values among neighboring grid points may be a random variation. When an ant at current grid point chooses a neighboring grid point to move to, such random variation must be ignored when judging the result. In our research, a fluctuation limitation was defined for filtering random variations between any two neighboring grid points. If the maximum difference of element concentration values between a current grid point and its neighbors is less than the fluctuation limitation, the fluctuation in element concentration value is regarded as random variation. In this case, an ant at a current grid point will randomly choose a neighboring grid point to move to.

How the fluctuation limitation affects the performance of the ant colony algorithm in geochemical anomaly detection has been investigated in our case study. Fig. 3 shows that the average Au concentration value varies with iterations when different fluctuation limitations are used in the iterative search process. This figure illustrates that the iterative search process converges to a value (a) more than 12 (which is the same as the converged value given m = 123) after more than 1000 iterations using fluctuation limitation of 0.001 and (b) less than 11 after less than 1000 iterations using fluctuation limitations of 0.0, 0.01, or 0.1. Therefore, the optimal fluctuation limitation is 0.001 for Au anomaly detection in our case study. A larger (i.e., 0.1) or a smaller (i.e., 0.0) fluctuation limitation may make the process quickly converge to a lower value. In other words, either a lower or a higher fluctuation limitation may degrade the performance of the ant colony algorithm in Au anomaly detection.

For computation simplicity, the probability from grid point *i* to its neighboring grid point *j* for ant *k* at step *t* is replaced by the favorability which is based on functional composition of pheromone trail and local heuristic value:

$$a_{ij}^{k}(t) = \left[\tau_{ij}(t)\right]^{\alpha} \cdot \left[\eta_{ij}(t)\right]^{\beta} \quad \forall j \in A; j \notin TL_k \qquad (5)$$

where $\tau_{ij}(t)$ is the pheromone trail intensity on the path from grid point *i* to its neighbor *j*; $\eta_{ij}(t)$ is the local heuristic value expressed by the geochemical element concentration value at grid point *j*; *A* is the set of eight neighbors of grid point *i*; TL_k is ant *k*'s Taboo List; and α and β are two parameters that control the relative weight of pheromone trail $\tau_{ij}(t)$ and local heuristic value $\eta_{ij}(t)$. The favorability is non-regularized probability. Its function is the same as the corresponding probability and it can be easily computed in the iterative search process.

The roles of the parameters α and β are the following. If $\alpha = 0$, the grid points with highest value are more likely to be selected; this corresponds to a classical stochastic greedy algorithm (with multiple starting points since ants are initially randomly distributed on the grid points). If $\beta = 0$, only pheromone amplification is at work; this method leads to rapid convergence of a stagnation, that is, a situation in which all ants make the same tour that, in general, is strongly sub-optimal (Dorigo et al., 1996). An appropriate trade-off has to be set between pheromone trail intensity and heuristic value. In geochemical anomaly detection, we suggest that α and β should be almost same numbers.

At iteration *t* ($t = 1, 2, \ldots, t_{max}$), ant *k* ($k = 1, 2, \ldots, m$) completes its one-step local heuristic search and deposits the following quantity of pheromone on the path from grid point i to its neighbor j:

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} (v_j - v_i)/c & \text{if ant } k \text{ moves from grid point } i \text{ to its neighbor } j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where v_i and v_j are element concentration values at grid points i and j, respectively; c is the regularization constant. In our research, c is defined as 10 times the maximum value of element concentration data.

The pheromone evaporation is considered in the update of pheromone trail to avoid occurring too high pheromone trails that may degrade the performance of the heuristic search process. The pheromone trail on the path from grid point i to its neighbor j is updated as follows:

$$\tau_{ij}(t+1) = \rho\tau_{ij}(t) + (1-\rho)\sum_{1\le k\le m}\Delta\tau_{ij}^{k}(t) \quad (7)$$

where $\tau_{ij}(t)$ and $\tau_{ij}(t+1)$ are the pheromone trails on the path before and after the update, respectively; ρ is the pheromone evaporation coefficient expressed by a constant within interval (0, 1); m is the number of ants randomly put into the grid element map; and $\Delta\tau_{ij}^{k}(t)$ can be computed using Eq. (6).

Whether different ρ-values affect the performance of the iterative searching process in geochemical anomaly detection is investigated in our research. Five different ρ-values are used for the pheromone trail update in Au anomaly detection in our case study. Fig. 4 shows that the average Au concentration value varies with iterations for ρ = 0.01, 0.05, 0.08, 0.1, and 0.5. This figure reveals that the iterative search process converges to a value (a) more than 12 given $0.05 \le \rho \le 0.08$ and (b) less than 11 given $\rho < 0.05$ or $\rho > 0.08$. Therefore, different ρ-values can affect the performance of the ant colony algorithm in Au anomaly detection. The optimal ρ-value for the algorithm is between 0.05 and 0.08 in Au anomaly detection.

The algorithm is stopped when all the ants become stationary or the maximum number of iterations has been completed. Then a user-defined threshold can be used to identify simulated geochemical anomalies from the ant density data that are the number of times each grid point has been visited by ants. A grid point belongs to an element concentration anomaly if the number of times it has been visited by ants is more than a predefined threshold value.

It should be pointed out that if the ant colony algorithm is executed more than once and the same set of parameters is used to initialize the algorithm, different geochemical anomaly detection results may be obtained due to the randomness of the initial positions of *m* ants. In order to reduce the impact of the randomness of the initialized positions of *m* ants, the results obtained from different executions can be synthesized into a comprehensive one by averaging the ant density data derived from different executions, and geochemical anomalies can be identified from the comprehensive ant density data using a threshold method. The Python pseudo-code for geochemical anomaly detection based on the ant colony algorithm is listed in Table 1.

4. Case Study

The Altay district in northern Xinjiang in China is chosen as a study area. Data of Au, Ag, Cu, Pb, and Zn concentration values collected from irregularly distributed stream sediment samples were transformed into concentration values at regularly spaced grid points by interpolating with the Golden Software Surfer and saved in five

GRID files. The ant colony algorithm was used to detect anomalous Au, Ag, Cu, Pb, and Zn concentration values in each of the five GRID files.

4.1 Geological setting and mineral deposits

The Altay orogenic belt is the amalgamated part of the Siberia plate and the Kazakhstan-Junggar plate (Li, 1996; Li and Zhao, 2002). The collision and amalgamation of the Siberia, Khazakstan, and Junggar blocks during Devonian and Early Carboniferous periods resulted in the evolution of geotectonic environment that provided permissive conditions for Au, Ag, Cu, Pb, and Zn mineralizations (Zeng et al, 2005). The widely distributed intermediate-acid intrusive rocks and Devonian marine volcanic-sedimentary rocks are genetically related to Au, Ag, Cu, Pb, and Zn mineralizations and the lower−middle marine volcaniclastic-sedimentary sequence of the Devonian Ashele Formation are the primary ore-host strata (Ye et al., 1996; Zhang et al., 2014). The Ashele Cu-Zn deposit, Duolanasayi Au deposit, and other dozens of mineral deposits have been discovered in the study area (Fig. 5). The polymetallic mineralization in the study area was closely related to the Altay orogenic process (Li, 1996; Li and Zhao, 2002). The Ashele Cu-Zn deposit was formed in Ashele volcanic-sedimentary basin in the foreland during the orogenic intermittent extensional period. U-Pb dating of zircons reveals that the mineralization in the sedimentary exhalative period occurred in the late Early Devonian (388 - 387 Ma) and the magmatic hydrothermal mineralization took place in the Middle Devonian (379.4 ± 0.8 Ma) (Yang et al, 2013). The Duolanasayi, Saidu, and Axile Au deposits were formed in the southern margin of the Altay orogenic belt during the main orogenic period. The mineralization style, mineral assemblage, and ore-forming fluid are similar to those of epizonal orogenic Au deposits defined by Groves et a1. (1998). Isotopic ages of the Au deposits are clustered in the range of 310 - 270 Ma, corresponding to Late Carboniferous to Early Permian (Yan et al., 2006).

The known mineral deposits in the study area belong to the same metallogenic series (Wang and Chen, 2001) controlled regionally by the orogenic structures, the volcaniclastic-sedimentary sequence of the Devonian Ashele Formation, and the intermediate-acid intrusive rocks. According to our statistics, 96% of the known mineral deposits in the study area occur in the Devonian Ashele Formation. In this case study, the known mineral deposits serve as the binary target variable for validation of the geochemical anomaly detection.

4.2 Geochemical exploration data

A geochemical stream sediment survey has been conducted in the study area. A total of 1623 stream sediment samples were collected along the drainage systems in the study area (Fig. 6), which is located in an arid region where there are no stream flows within most drainage basins. This posed difficulty in sampling stream sediments in the study area. In order to ensure the sampling density of the geochemical survey, sediment samples were collected at a site no matter whether the stream is active or not (i.e., dry). As a result, some samples are located close the boundaries of drainage catchments (Fig. 6). Concentrations of Au, Ag, Cu, Pb, and Zn in each sample were measured in different units: parts per billion (ppb) for Au; and parts per million (ppm) for the other elements. Using the element concentration data of the irregular distributed stream sediment samples, regularly spaced grid element concentration values were generated by interpolation with inverse distance to a power using the Golden Software Surfer. In data interpolation, a power value should usually fall between one and three. Accordingly, a power value of 2 was used in our case study. The grid point spacing is 0.2539 km in east-west direction and 0.2536 km in north-south direction. There are 100 × 151 grid points in all and 5518 grid points are located within the blank area which is covered by Gobi desert. Fig. 7 shows the contour maps of grid element concentration values on which the known mineral deposits are superimposed.

4.3 Metallogenic indicator assessment

Receiver operating characteristic (ROC) curve (Zou et al., 2007) and area under the curve (*AUC*) (Flach et al., 2011; Chen, 2014) are widely used classification performance measures in machine learning. For a binary classification system, the steeper an ROC curve is toward the upper left corner in the ROC space, the better the binary classification system performs. A perfect classification corresponds to the upper left corner of the ROC space. The *AUC* value is an overall performance measure, which integrates the performance measure of an ROC curve into one metric. The *AUC* value is between 0.5 and 1; an *AUC* value of 0.5 means the classification is equivalent to a random guess while an *AUC* value of 1 means the classification is perfect. On the basis of *AUC* value, statistics Z_{AUC} can be computed (Chen, 2014). Z_{AUC} satisfies standard normal distribution. It can be used to test at α-level ($\alpha = 0.01$, 0.05, or 0.1) to determine if an *AUC* value is significantly different from 0.5 (Chen, 2014).

In geochemical exploration, anomalous concentration values of each geochemical indicator element must be spatially associated with the known (discovered) mineral deposits (the undiscovered mineral deposits are ignored). In other words, indicator element concentration values at regularly-spaced grid points can be used to differentiate between deposit-bearing and non-deposit-bearing grid points. In our case study, deposit-bearing and non-deposit-bearing grid points are defined as follows: a deposit-bearing grid point is located near a known mineral deposit (the distance between a deposit-bearing grid point and a mineral deposit is less than 0.1794 km) while a non-deposit-bearing one is located more than 0.1794 km from any known mineral deposit. The critical distance of 0.1794 km is computed using the following function of the grid point spacing in both east-west and north-south directions: $\sqrt{(\frac{0.2539}{2})^2 + (\frac{0.2536}{2})^2} \approx 0.1794$ (km).

It should be stressed that there are geochemical anomalies located near undiscovered mineral deposits in a study area and these anomalies are exactly what geochemical exploration is seeking. In metallogenic indicator assessment, the undiscovered mineral deposits are ignored when defining deposit-bearing and non-deposit-bearing grid points. As the result, grid points near the undiscovered mineral deposits (i.e., deposit-bearing grid points) are incorrectly defined as non-deposit-bearing ones. However, these incorrectly defined non-deposit-bearing grid points usually account for a very small proportion of total non-deposit-bearing grid points. Thus, the ROC curve and *AUC* value for an indicator element does not significantly affected by ignoring the undiscovered mineral deposits.

In order to test whether Au, Ag, Cu, Pb, and Zn can serve as geochemical indicator elements in our case study, 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. Each of the discretized thresholds was used to discriminate between deposit-bearing and non-deposit-bearing grid points. Grid points with element concentration values more than the threshold are predicted as deposit-bearing while grid points with element concentration values less than the threshold are predicted as non-deposit-bearing. A binary classification system can be established with respect to all the discretized thresholds, and then the corresponding ROC curve and the *AUC* value can be obtained based on the predicted deposit-bearing and non-deposit-bearing grid points. The ROC curves for Au, Ag, Cu, Pb, and Zn are shown in Fig. 8. The estimated *AUC*s and their standard deviations as well as statistics Z_{AUC}s for Au, Ag, Cu, Pb, and Zn are listed in Table 2.

Fig. 8 indicates that concentration values of Au, Ag, Cu, Pb, and Zn can differentiate between deposit-bearing and non-deposit-bearing grid points. Table 2 shows that the estimated *AUC* values of Au, Ag, Cu, Pb, and Zn are significantly different from 0.5 at level $\alpha = 0.1$. Therefore, Au, Ag, Cu, Pb, and Zn can serve as geochemical indicator elements in our case study.

4.4 Ant density data generation

The ant colony algorithm is an iterative search process whereby all the ants in a colony move simultaneously toward adjacent grid points in a grid element map until the termination condition is satisfied. The ant density data, which record the movement tracks of ants in the grid element map, can be obtained after the iterative search process. In each iteration, an ant heuristically searches only the neighboring grid points. Thus, the ant density data are not obviously impacted by regional variations of geochemical background. In other words, transforming element concentration data into the ant density data can bate the impact of regional variation of geochemical background on geochemical anomaly detection. As a result, low or strong geochemical anomalies can be differentiated from geochemical background based on ant density data using a threshold method. The ant density data can also slightly enhance small and weak anomalies in the grid element map if the length of Taboo List is defined properly. Therefore, the ant density data are superior to the corresponding interpolated element concentration data in regard to geochemical anomaly identification.

The following parameters were empirically determined in our case study: m =123; $t_{\max}$ = 5000; $repeat_{\max}$ = 3; α = 2.5; β = 3.0; ρ = 0.08; τ_0 = 0.001; $\tau_{\min}$ = 1.0E-38; *Taboo* = 20 for Pb and *Taboo* = 40 for other elements; $c = z_{\max}*10.0$; $d_{\min}$ = 0.001 for Au; $d_{\min}$ = 0.00001 for Ag; and $d_{\min}$ = 0.1 for Cu, Pb, and Zn. Where m is the number of ants, $t_{\max}$ is the number of iterations, $repeat_{\max}$ is the number of times that the iterative searching process is repeated, α is the weight coefficient of pheromone trail, β is the weight coefficient of element concentration value, ρ is the pheromone evaporation coefficient, τ_0 is the initialization value of pheromone trail, $\tau_{\min}$ is the allowed minimum value of pheromone trail, *Taboo* is the length of Taboo List, c is the regularization constant, d_{min} is fluctuation limitation, and z_{max} is the maximum element concentration value in a grid element map.

According to the above predefined parameters, 123 artificial ants were randomly put into each grid element map and then the ant colony completed 5000 heuristic search steps in a grid element map. This procedure was repeated one to ten times. By comparing the geochemical identification results, we found that increasing the number of repetitions could not obviously improve the performance of geochemical anomaly detection. For the results shown in this paper, only three times of repetition were used. The ant density data for the grid element map were obtained by computing the average of the ant density values obtained from the three repetitions.

4.5 Geochemical anomaly identification

A user-defined threshold can be used to identify geochemical anomalies from the ant density data generated by the ant colony algorithm. The Youden index (Youden, 1950; Ruopp et al., 2008) can be used to determine the optimal threshold in geochemical anomaly identification (Chen, 2014). In medical statistics, the Youden index is a single statistic that captures the performance of a diagnostic test. It is defined as the difference of true positive rate minus false positive rate. Its value ranges from 0 to 1, and has a zero value when a diagnostic test gives the same proportion of positive results for groups with and without the disease, i.e, the test is useless. A value of 1 indicates that there are no false positives or false negatives, i.e., the test is perfect. The index gives equal weight to false positive and false negative values, so all tests with the same value of the index give the same proportion of total misclassified results. In geochemical anomaly identification, the Youden index can be used to express the association between the recognized geochemical anomalies and the known mineral deposits; the bigger the value of the Youden index is, the stronger the association of recognized geochemical anomalies is with the known mineral deposits.

We used the Youden index to determine optimal thresholds for Au, Ag, Cu, Pb, and Zn anomaly identification from both the ant density data and the interpolated element concentration data. The maximum Youden indices, the optimal thresholds, and the percentage of the study area and the percentage of the known mineral deposits in the identified anomalies are listed in Table 3. The geochemical anomalies identified from the interpolated Au, Ag, Cu, Pb, and Zn concentration data and from the corresponding ant density data are shown in Fig. 9.

4.6 Discussion

Table 3 shows the following statistical results: (a) the Au anomalies identified from the ant density data occupy 4.2% of the study area and contain 24% of the known mineral deposits, and the Au anomalies identified from the interpolated Au concentration data occupy 10.1% of the study area and contain 36% of the known mineral deposits; (b) the Ag anomalies identified from the ant density data occupy 2.7% of the study area and contain 24% of the known mineral deposits, and the Ag anomalies identified from the interpolated Ag concentration data occupy 5.7% of the study area and contain 40% of the known mineral deposits; (c) the Cu anomalies identified from the ant density data occupy 16.4% of the study area and contain 40%

of the known mineral deposits, and the Cu anomalies identified from the interpolated Cu concentration data occupy 28.7% of the study area and contain 64% of the known mineral deposits; (d) the Pb anomalies identified from the ant density data occupy 10.1% of the study area and contain 40% of the known mineral deposits, and the Pb anomalies identified from the interpolated Pb concentration data occupy 9.0% of the study area and contain 60% of the known mineral deposits; and (e) the Zn anomalies identified from the ant density data occupy 13.8 % of the study area and contain 52% of the known mineral deposits, and the Zn anomalies identified from the interpolated Zn concentration data occupy 28.5 % of the study area and contain 80% of the known mineral deposits. Thus, compared with the geochemical anomalies identified from the element concentration data, those identified from the ant density data usually have a higher Youden index value, are smaller in size, and contain more of the known mineral deposits.

Fig. 9 shows that: (a) robust anomalies identified from the element concentration data coincide with those identified from the ant density data; (b) each large-scale robust anomaly identified from the element concentration data usually corresponds to several small robust ones identified from the ant density data, in other words, the ant colony algorithm tends to separate a large-scale robust anomaly into several small ones; (c) some small weak anomalies identified from the element concentration data correspond to one anomaly identified from the ant density data; (d) some anomalies identified from the ant density data can not be identified from the element concentration data, i.e., the ant colony algorithm can identify some anomalies that can not be identified from the element concentration data; and (e) some small weak anomalies identified from the element concentration data can not be identified from the ant density data, i.e., the ant colony algorithm can omit some small weak anomalies that can be identified from the element concentration data.

The identified Au, Ag, Cu, Pb, and Zn anomalies are distributed spatially around intermediate-acid intrusive rocks and the robust anomalies occur mainly in the volcaniclastic-sedimentary sequence of the Devonian Ashele Formation. As mentioned in Section 4.1, most of the known mineral deposits occur in this geological formation. Therefore, Au, Ag, Cu, Pb, and Zn anomalies that occur in the volcaniclastic-sedimentary sequence of the Devonian Ashele Formation around the intermediate-acid intrusive rocks are prospective targets for further mineral exploration.

5. Conclusion

The ant colony algorithm can correctly identify geochemical anomalies from grid element data from the Altay district in northern Xinjian in China, and the spatial distribution of the identified geochemical anomalies strongly coincide with the ore-related geological formations in the study area. Anomalous areas detected by the ant colony algorithm occupy 9.5% of the study area and contain 36% of the known mineral deposits; and anomalous areas identified using the Youden index method occupy 16.4% of the study area and contain 56% of the known mineral deposits. Thus, the ant colony algorithm is a feasible method for geochemical anomaly detection.

The configuration parameters, to some extent, affect the performance of the ant colony algorithm in geochemical anomaly detection. The values of these parameters need to be empirically determined in practice. The case study shows that the ant colony algorithm performs best if (a) the number of ants is approximately equal to the square root of the total number of grid points, (b) the length of the Taboo List is between 20 and 50, and (c) the regularization constant is approximately 10 times the maximum value of the interpolated element concentration data.

Acknowledgments

We are grateful to Guest Editor, John Carranza, for his time and kindness in modifying and editing our manuscript. His help benefits us a lot. We are also grateful to the two reviewers for their comments which greatly improved our manuscript. This research is financially supported by the National Natural Science Foundation of China (Grant nos.41272360, 41472299, and 61133011).

References

Bullnheimer, B., Hartl, R. F., Strauss, C., 1997. An improved ant system algorithm for the vehicle routing problem. Technical Report. POM-10/97. Institute of Management Science, University of Vienna, Austria.

Chen, Y.L., 2015. Mineral potential mapping with a restricted Boltzmann machine. Ore Geology Reviews 71, 749-760.

Chen, Y.L., Lu, L.J., and Li, X.B., 2014. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. Journal of Geochemical Exploration 140, 56-63.

Cheng, Q., 1995. The perimeter-area fractal model and its application to geology.

Mathematical Geology 27, 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, 3, p. 204.

Cheng, Q., 2006. Singularity-generalized self-similarity-fractal spectrum (3S) model. Earth Sciences-China University of Geosciences 31, 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 Geology Reviews 32, 314–324.

Cheng, Q., 2008. Non-linear theory and power-law models for information integration and mineral resources quantitative assessments. Mathematical Geology 40, 503–532.

Cheng, Q., Agterberg, F.P., 1995. Multifractal modeling and spatial point processes. Mathematical Geology 27, 831–845.

Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. Journal of Geochemical Exploration 51, 109–130.

Cheng, Q., Agterberg, F.P., Bonham-Carter, G.F., 1996. A spatial analysis method for geochemical anomaly separation. Journal of Geochemical Exploration 56, 183–195.

Cheng, Q., Xu, Y., Grunsky, E., 2000. Multifractal power spectrum–area method for geochemical anomaly separation. Natural Resources Research 9, 43–51.

Costa, D., Hertz, A., 1997. Ants can color graphs. Journal of the Operational Research

Society, 48, 295-305.

Deng, J., Wang, Q., Yang, L., Wang Y., Gong Q., Liu H., 2010. Delineation and explanation of geochemical anomalies using fractal models in the Heqing area, Yunnan Province, China. Journal of Geochemical Exploration 105, 95–105.

Dorigo, M., 1992. Optimization, learning and natural algorithms, PhD Thesis, Politecnico di Milano.

Dorigo, M., Gambardella, L.M., 1997. Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation 1, 1, 53-66.

Dorigo, M., Di Caro, G., Gambardella, L., 1999. Ant colony optimization: new meta-heuristic. Proceedings of the Congress on Evolutionary Computation, 1470-1471.

Dorigo, M., Maniezzo, V., Colorni, A., 1996. The ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics–Part B, 26, 1, 29–41.

Flach, P.A., Hernandez-Orallo, J., Ferri, C., 2011. A coherent interpretation of AUC as a measure of aggregated classification performance. Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 657-664.

Galuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environmental Geology 52, 5, 861–870.

Gambardella, L. M., Taillard, E. D., Agazzi, G., 1999. MACS-VRPTW: A multiple ant colony system for vehicle routing problems with time windows. In: Corne, D., Dorigo, M., Glover, F. (eds), New Ideas in Optimization. McGraw-Hill.

Garrett, R.G., 1989. A cry from the heart. Explore (AEG Newsl.) 66, 18–20.

Govett, G.J.S., Goodfellow, W.D., Chapman, A., Chork, C.Y., 1975. Exploration geochemistry distribution of elements and recognition of anomalies. Mathematical Geology 7, 415-446.

Geranian, H., Tabatabaei, S.H., Asadi, H.H., Carranza, E.J.M., 2015. Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay Gold Deposit, NW Iran. Natural Resources Research, DOI: 10.1007/s11053-015-9271-2.

Groves, D.I., Goldfarb, R.J., Gebre-Mariam, M., Hagemann, S.G., Robert, F., 1998. Orogenic gold deposits: A proposed classification in the context of their crustal distribution and relationship to other gold deposit types. Ore Geo1ogy Reviews 13, 7-27.

Hawkes, H.E., Webb, J.S., 1962. Geochemistry in Mineral Exploration. Harper and Row, New York, NY.

He., J., Yao, S., Zhang, Z., You, G., 2013. Complexity and productivity differentiation models of metallogenic indicator elements in rocks and supergene media around Daijiazhuang Pb–Zn deposit in Dangchang County, Gansu Province. Natural Resources Research 22, 19-36.

Li, Q., Cheng, Q., 2004. Fractal singular-value (eigen-value) decomposition method for geophysical and geochemical anomaly reconstruction. Earth Sciences-China

University of Geosciences 29, 109–118 (In Chinese with English Abstract).

Liu, H., 2007. An hybrid fuzzy variable neighborhood particle swarm optimization algorithm for solving quadratic assignment problems. Journal of Universal Computer Science13, 9, 1309-1331.

Li, Z.C., 1996. The evolution and formation of the left echelon volcanic-sedimentary basins in Altay and the tracing of their metallogenic activities. Geotectonica Et Metallogenia 20, 3, 189-200 (In Chinese with English Abstract).

Li, Z.C., Zhao, Z.H., 2002. Creation of the Altay orogenic belt and the Altay MTS. Tectono-metallogenic province. Chinese Journal of Geology 37, 4, 483-490 (In Chinese with English Abstract).

Luz, F., Mateus, A., Matos, J.X., Gonçalves, M.A., 2014. Cu- and Zn-soil anomalies in the NE Border of the South Portuguese Zone (Iberian Variscides, Portugal) identified by multifractal and geostatistical analyses. Natural Resources Research 23, 195-215.

Maniezzo, V., Colorni, A., Dorigo, M., 1994. The ant system applied to the quadratic assignment problem. Technical Report. IRIDIA/94-28. Universit´e Libre de Bruxelles, Belgium.

Miesch, A.T., 1981. Estimation of the geochemical threshold and its statistical significance. Journal of Geochemical Exploration 16, 49–76.

Ramos, V., Almeida, F., 2000. Artificial ant colonies in digital image habitats – a mass behavior effect study on pattern recognition. Proceedings of ANTS2000 – International Workshop on Ant Algorithms (From Ant Colonies to Artificial Ants), Brussels, Belgium, pp.113-116.

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. Biomedical Journal 50, 3, 419-430.

Singer, D.A., Kouda, R., 2001. Some simple guides to finding useful information in exploration geochemical data. Natural Resources Research 10, 137-147.

Stanley, C.R., 1988. Comparison of data classification procedures in applied geochemistry using Monte Carlo simulation. Ph.D. Thesis, University of British Columbia, Vancouver, pp. 256.

Stanley, C.R., Sinclair, A.J., 1989. Comparison of probability plots and gap statistics in the selection of threshold for exploration geochemistry data. Journal of Geochemical Exploration 32, 355–357.

Stützle, T., Hoos, H., 1998. Max–min ant system and local search for combinatorial optimization problems. In: Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization. Voß, S., Martello, S., Osman, I.H., Roucairol, C. (eds), Kluwer Academics, Boston, 137-154.

Wang, D.H., Chen, Y.C., 2001. A preliminary study on the metallogenic series type of Fe-Cu-Pb-Zn ore deposits related to submarine volcanism and its origin. Mineral Deposits 20, 2, 112-118 (In Chinese with English Abstract).

Xue, Y., Dai, T.G., Deng, H.J., Xia, H.D., Wang, D., 2008. Extraction of hydroxyl alteration information based on the ant colony algorithm-A case study of the Awa area, Tongren County, Qinhai, China. Geological Bulletin of China 27, 5, 657-661(In

Chinese with English Abstract).

Yan, S.H., Wang, Y.T., Zhang, Z.C., Chen, B.L., Chen, W., 2006. Metallogenic type, tectonic setting, and resource potential of Ertix gold ore belt in northern Xinjiang. Mineral Deposits 25, 6, 693-704 (In Chinese with English Abstract).

Yang, F.Q., Li, F.M., Qin, J.H., Zheng, K.P., Liu, F., 2013. LA-MC-ICP-MS U-Pb dating of zircons from volcanic rocks and subvolcanic intrusions in Ashele Cu-Zn ore district, Xinjiang, and its geological implications. Mineral Deposits 32, 5, 869-883 (In Chinese with English Abstract).

Ye, Q.T., Fu, X.J., Zhang, X.H., 1996. Geological characteristics and genesis of the Ashele copper-zinc massive sulfide deposit, Xinjiang. Mineral Deposits 6, 2, 97-106 (In Chinese with English Abstract).

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 1, 32-35.

Zeng, Q.S., Chen, G.H., Wang, H., Li, P.C., 2005. Polygenetic compound metallogenic characteristics of the Ashele copper deposit in northern Xinjiang. Geotectonica Et Metallogenica 29, 545-550 (In Chinese with English Abstract).

Zhang, Z.X., Yang, F.Q., Liu, F., Li, Q., Geng, X.X., 2014. Helium and argon isotopes tracing for sources of ore-forming fluid in the Ashele volcanogenic-host massive sulfide copper-zinc deposit, Chinese Altay. Geological Review 60, 1, 222-230 (In Chinese with English Abstract).

Zhuang, X., 2004. Image feature extraction with the perceptual graph based on the ant colony system. IEEE International Conference on Systems, Man and Cybernetics.

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., 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. Journal of Geochemical Exploration 101, 355–357.

Table and Figure captions:

Table 1 Python pseudo-codes for the ant colony algorithm of geochemical anomaly detection.

Table 2 AUCs, the standard deviations of *AUC*s, and Z_{AUC}s for Au, Ag, Cu, Pb, and Zn.

Table 3 The Youden index, optimal threshold, anomaly area percentage, and the percentage of mineral deposits in anomaly area.

Fig. 1. Average Au concentration value for iterations at different number of ants.

Fig. 2. Average Au concentration value for iterations at different lengths of Taboo List.

Fig. 3. Average Au concentration value for iterations at different fluctuation limitations.

Fig.4. Average Au concentration value for iterations at different ρ-values.

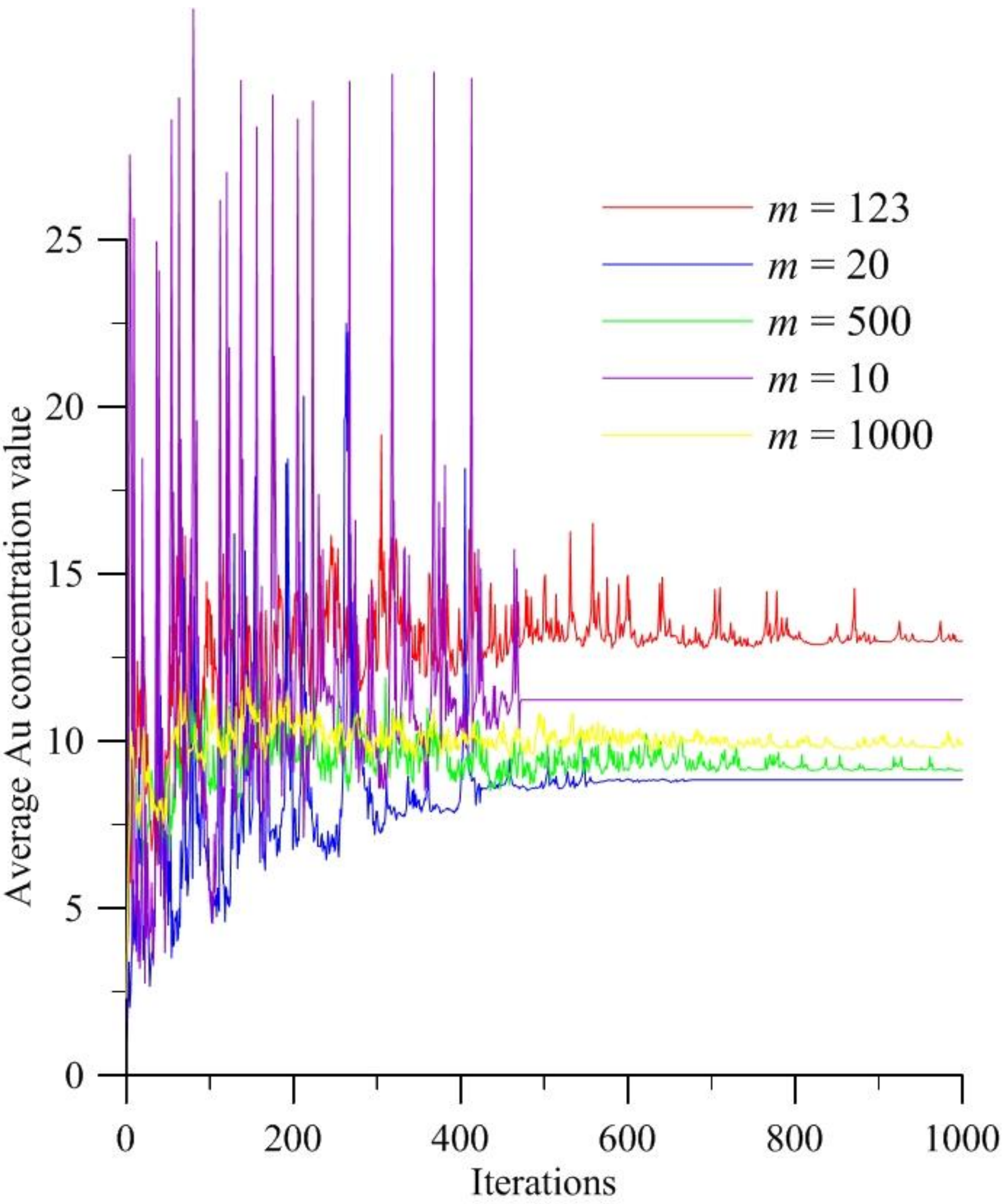
Fig. 5. Simplified geologic map on which the known mineral deposits are superimposed.

Fig. 6. Drainage basins on which the stream sediment sample locations are superimposed.

Fig. 7. Contour maps of element concentration values on which the known mineral deposits are superimposed.

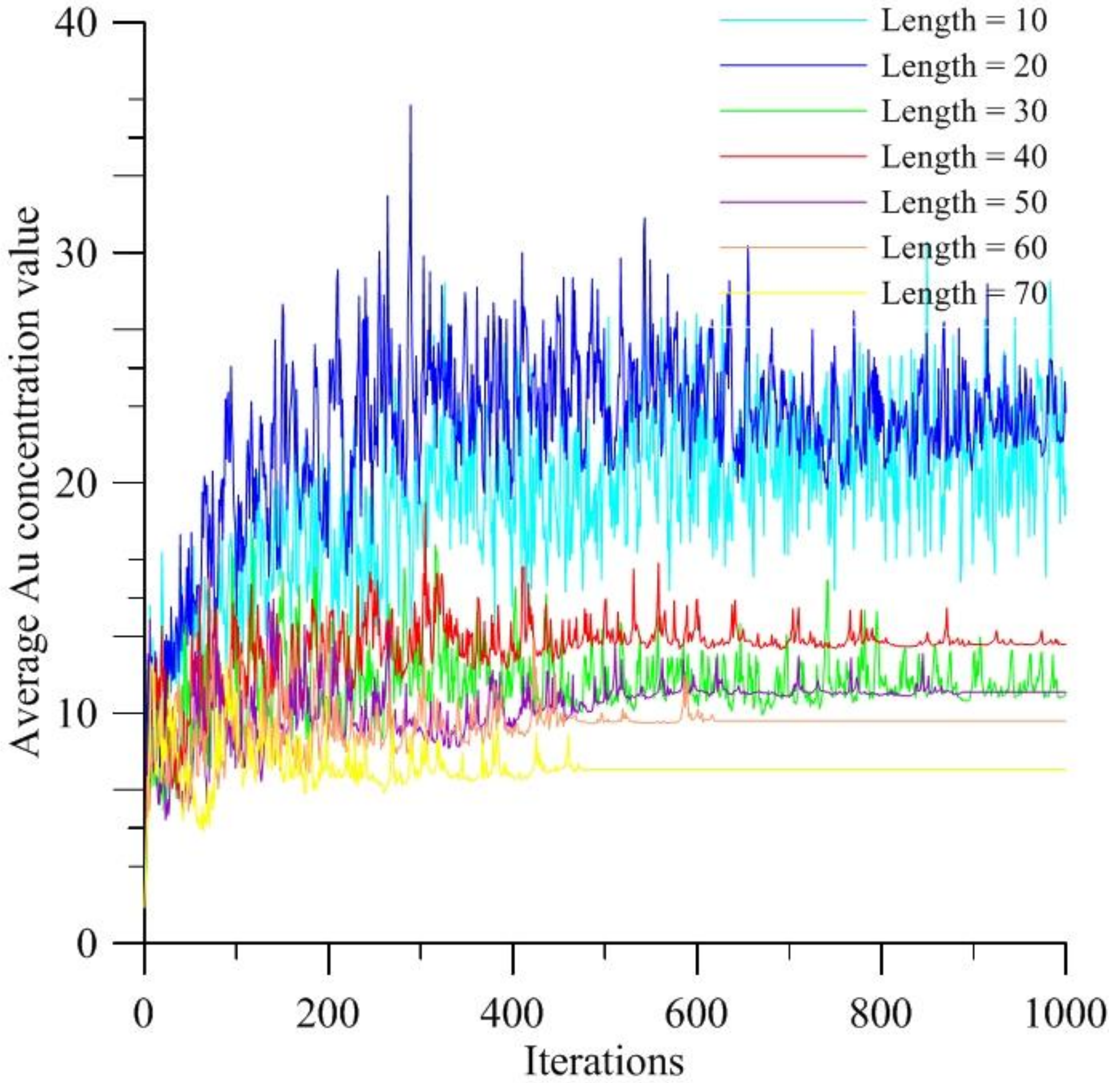
Fig. 8. ROC curves for Au, Ag, Cu, Pb, and Zn.

Fig. 9. Geochemical anomaly contour maps on which the known mineral deposits are superimposed.



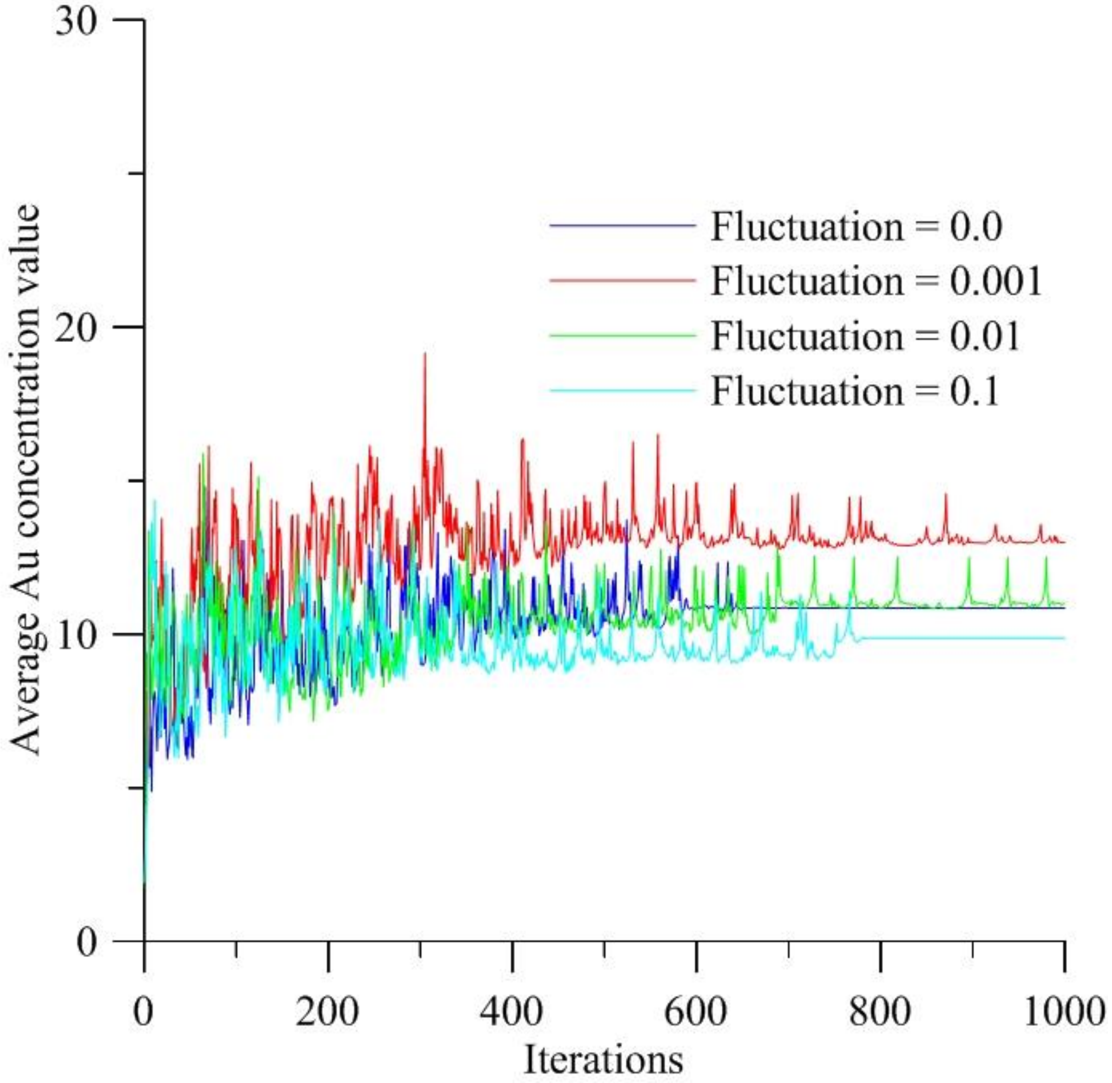
m = 123
m = 20
m = 500
m = 10
m = 1000
Average Au concentration value
0
5
10
15
20
25
0
200
400
600
800
1000
Iterations

Figure 1



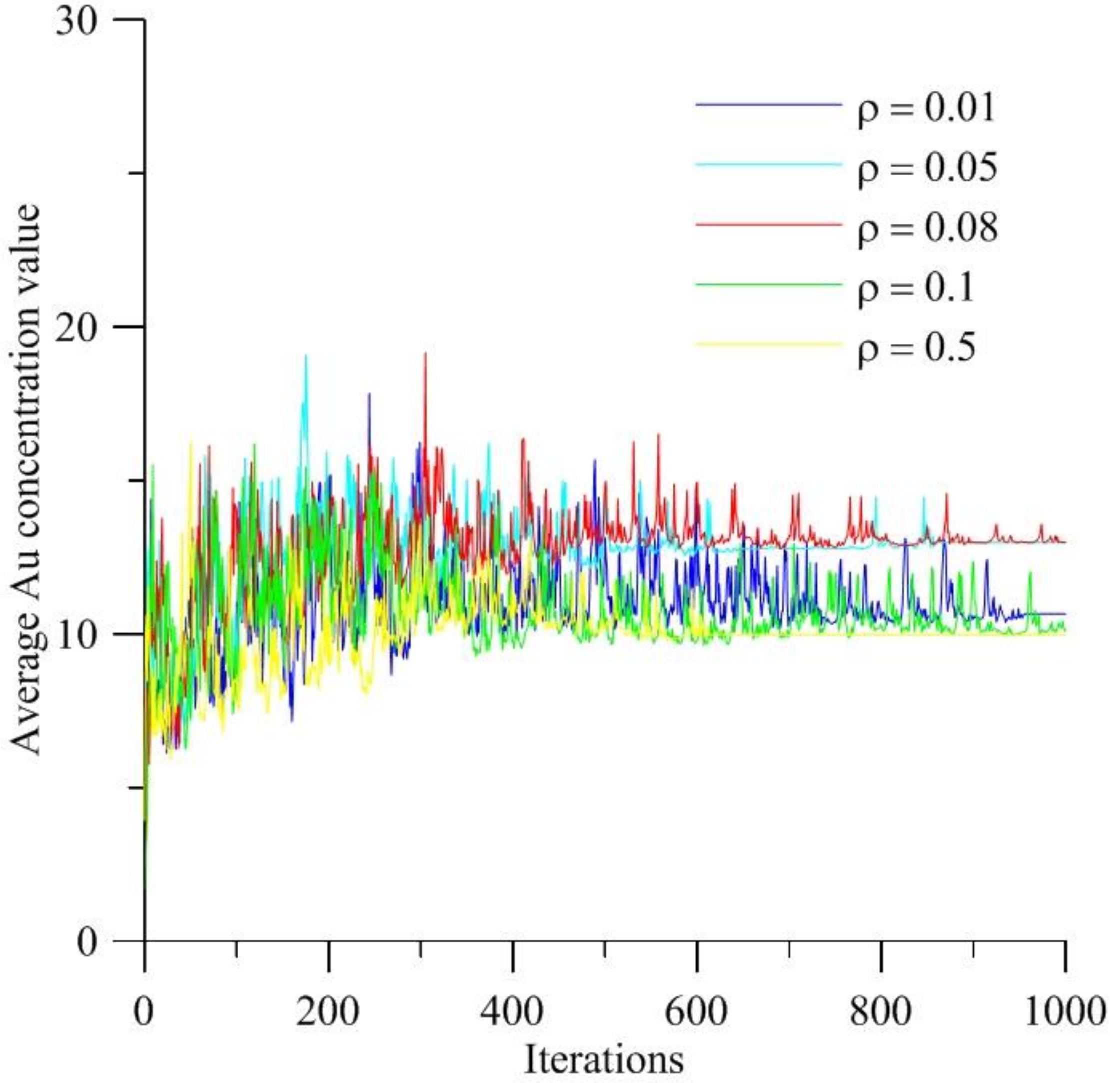
Length = 10
Length = 20
Length = 30
Length = 40
Length = 50
Length = 60
Length = 70
Average Au concentration value
Iterations
0
10
20
30
40
0
200
400
600
800
1000

Figure 2



30
20
10
0
Average Au concentration value
Fluctuation = 0.0
Fluctuation = 0.001
Fluctuation = 0.01
Fluctuation = 0.1
0
200
400
600
800
1000
Iterations

Figure 3



30
20
10
0
Average Au concentration value
0
200
400
600
800
1000
Iterations
ρ = 0.01
ρ = 0.05
ρ = 0.08
ρ = 0.1
ρ = 0.5

Figure 4

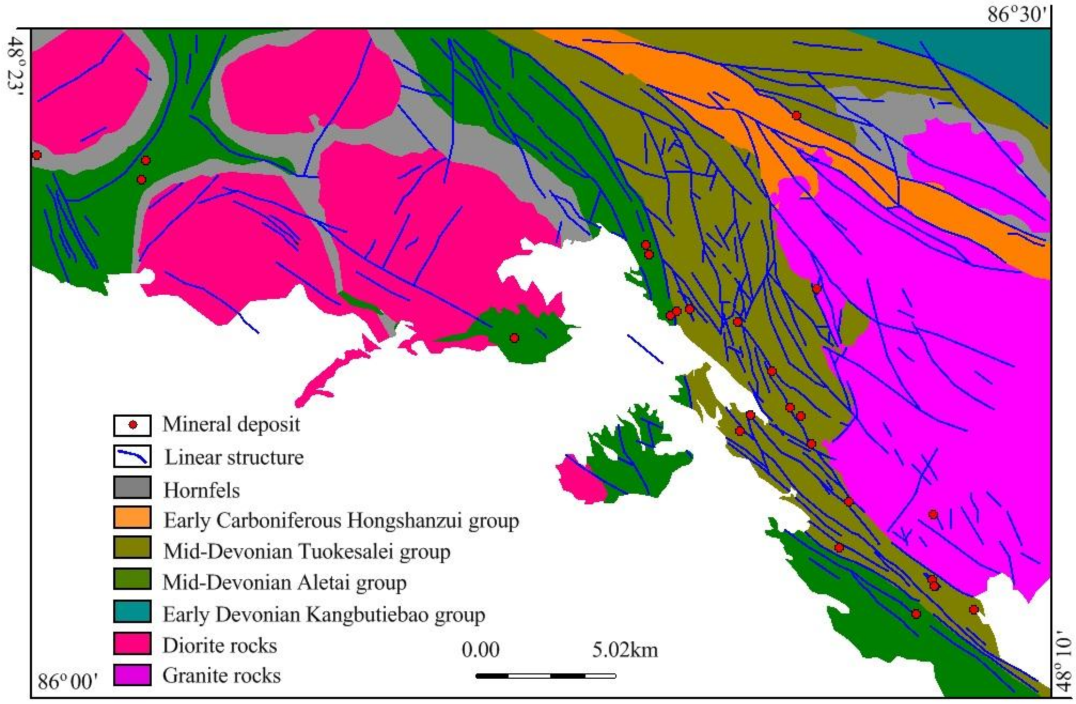

86°30'
48°23'
Mineral deposit
Linear structure
Hornfels
Early Carboniferous Hongshanzui group
Mid-Devonian Tuokesalei group
Mid-Devonian Aletai group
Early Devonian Kangbutiebao group
Diorite rocks
Granite rocks
0.00
5.02km
86°00'
48°10'

Figure 5

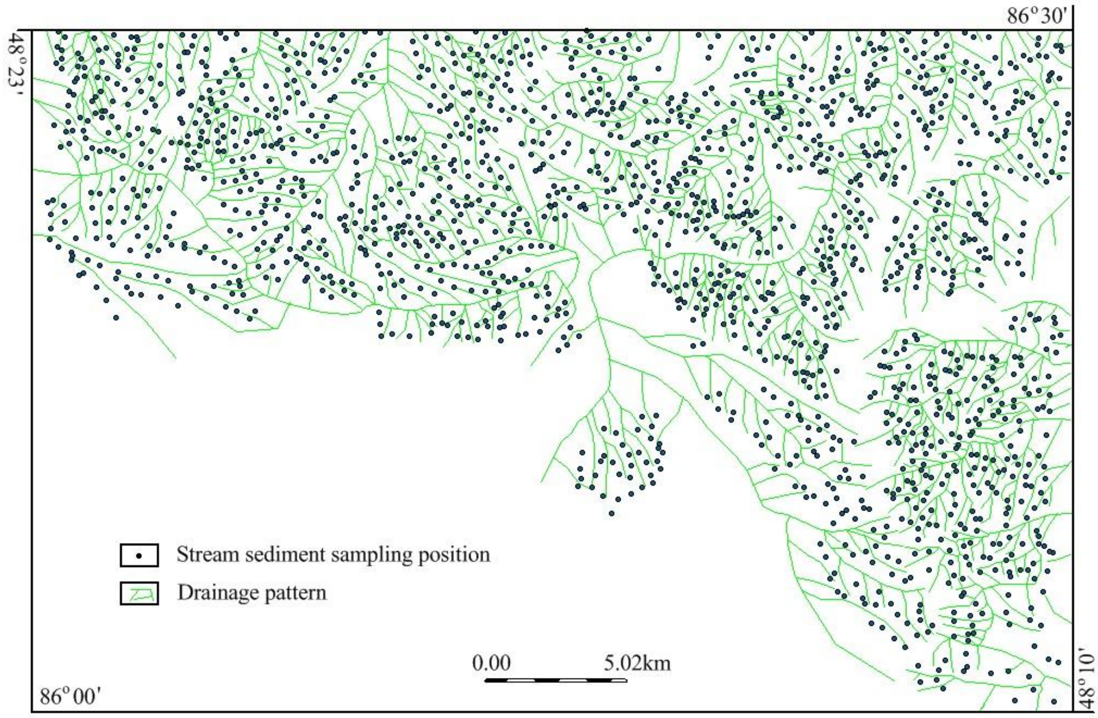

86°30'
48°23'
Stream sediment sampling position
Drainage pattern
0.00
5.02km
86°00'
48°10'

Figure 6

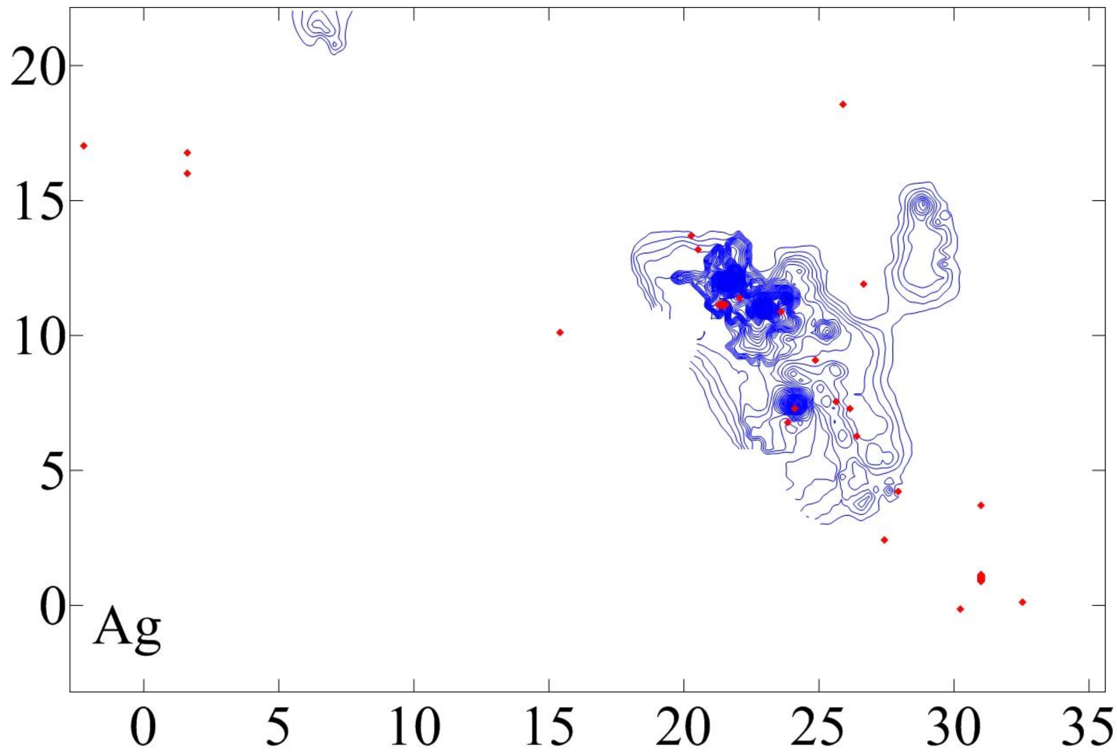


Figure 7ag

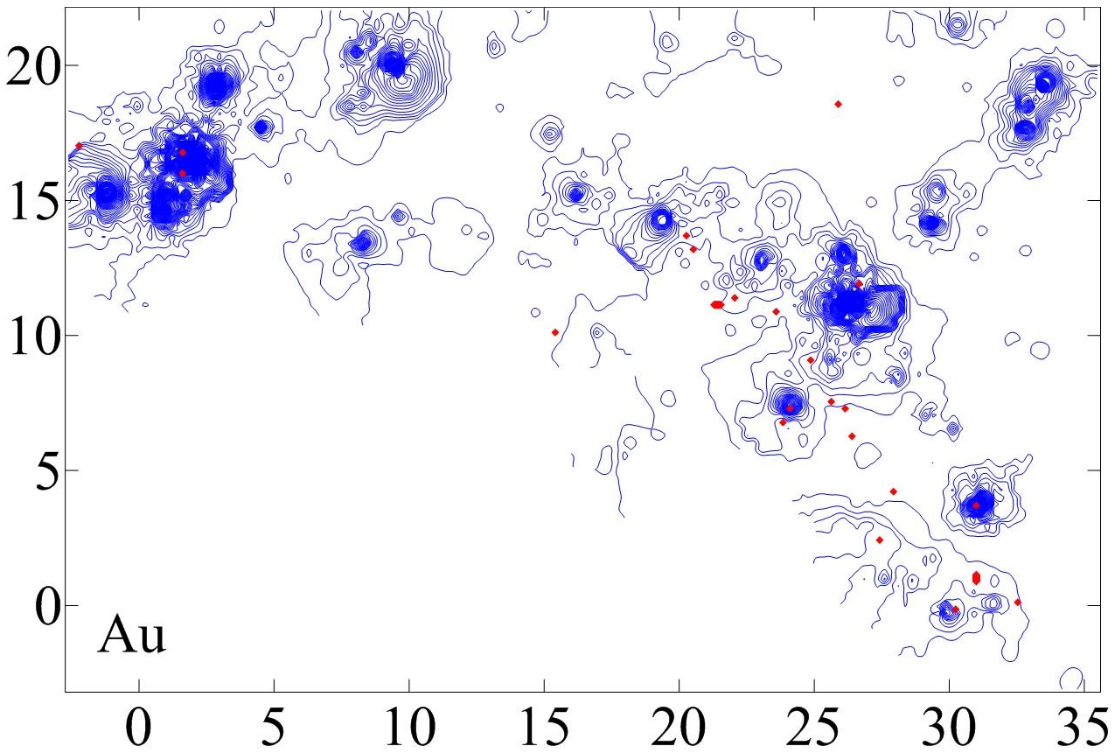


Figure 7au

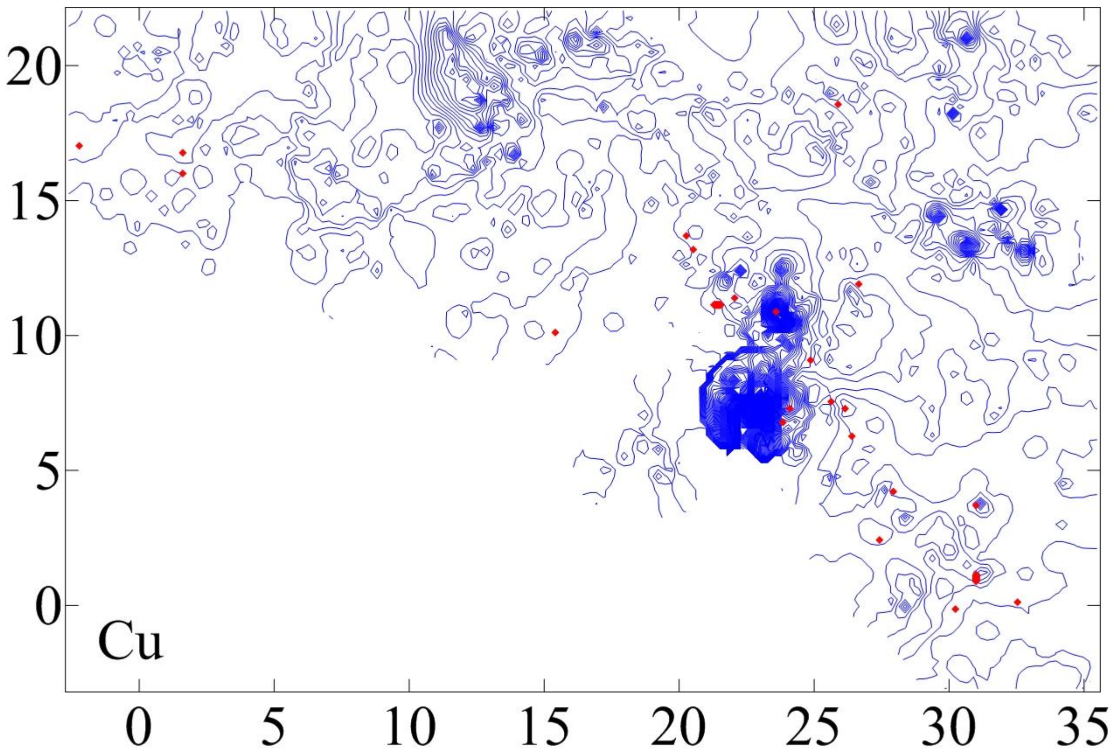


Figure 7cu

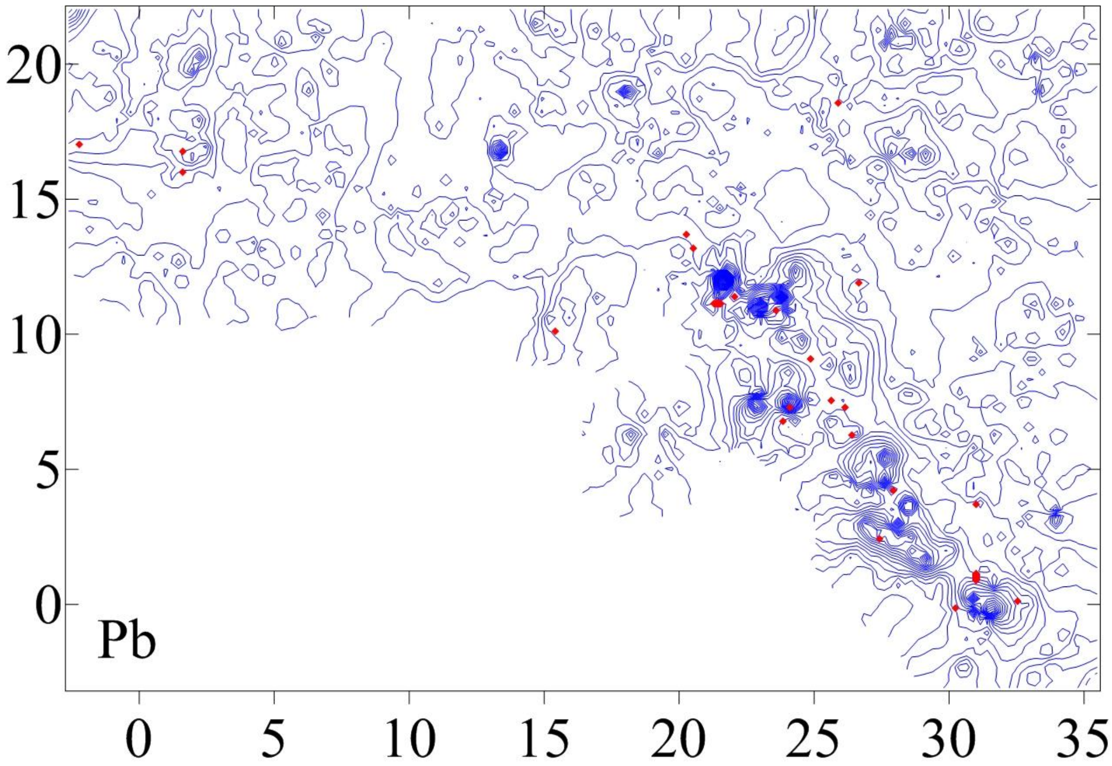


Figure 7pb

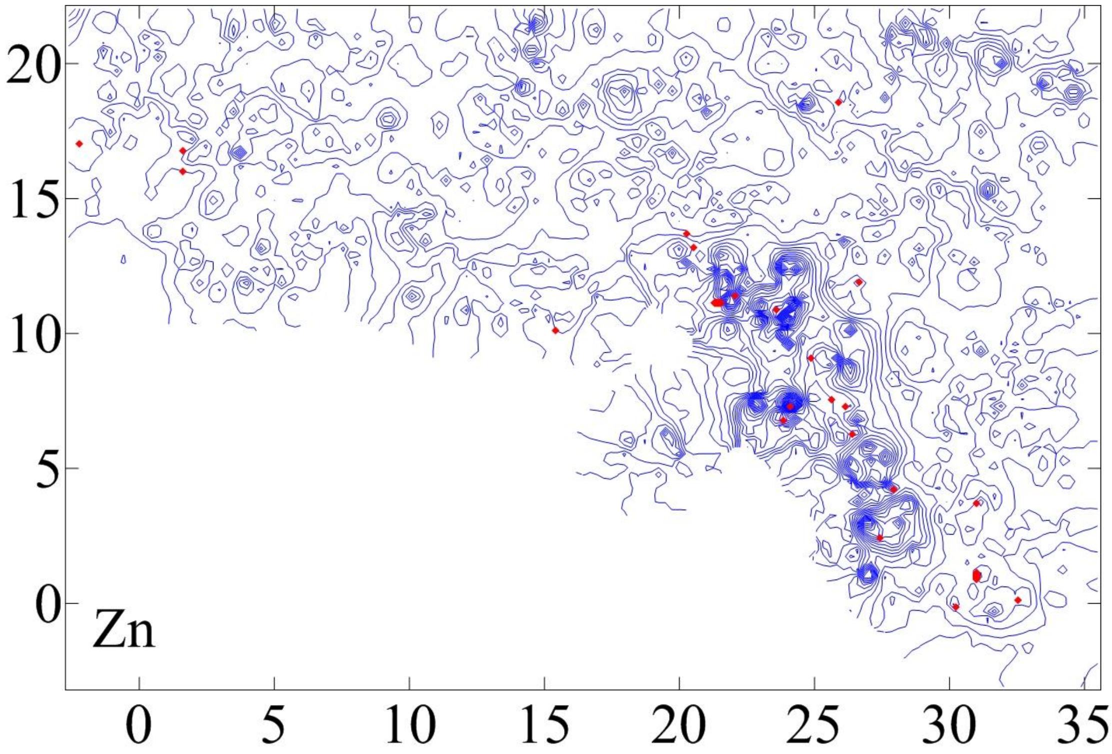


Figure 7zn

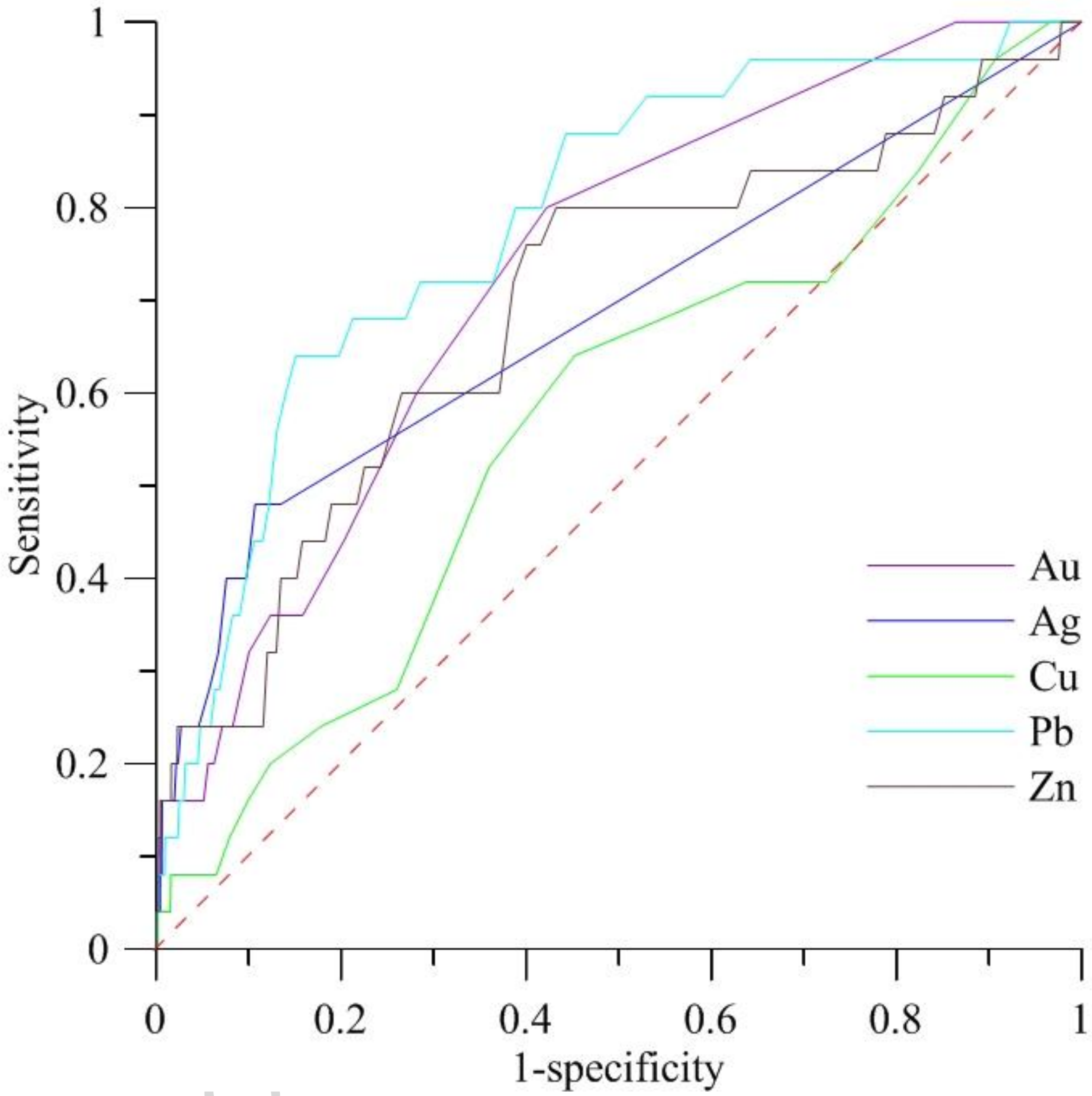

1
0.8
0.6
0.4
0.2
0
Sensitivity
0
0.2
0.4
0.6
0.8
1
1-specificity
Au
Ag
Cu
Pb
Zn

Figure 8

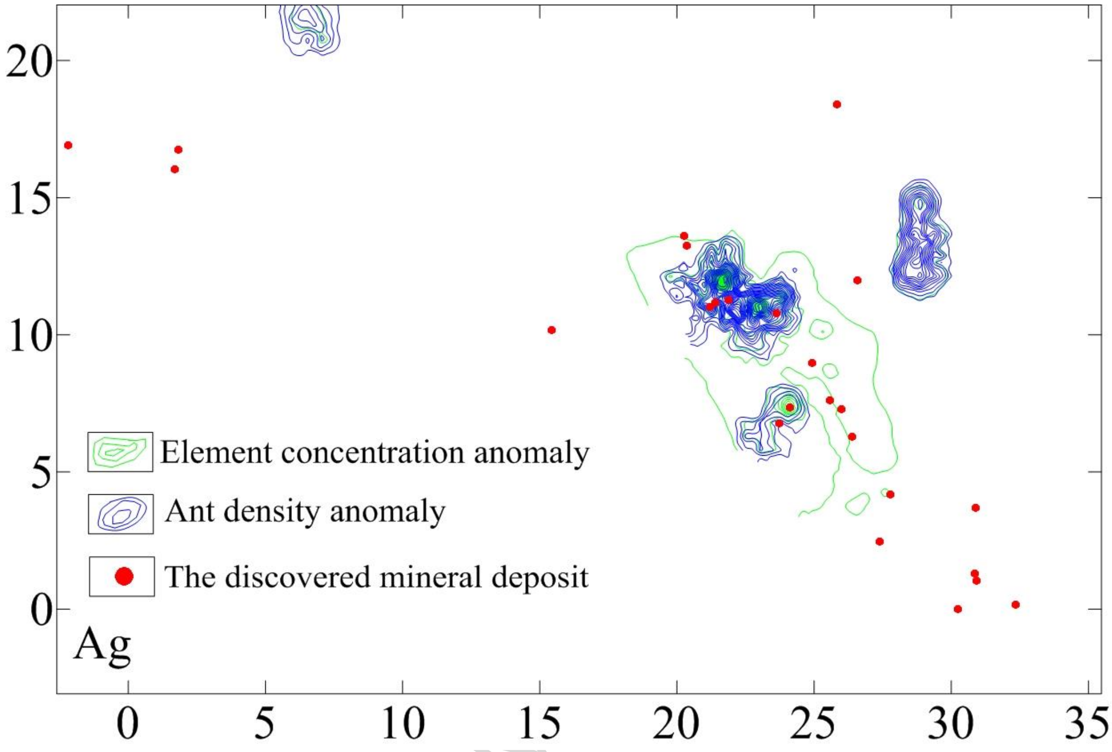

Element concentration anomaly
Ant density anomaly
The discovered mineral deposit
Ag
0
5
10
15
20
25
30
35

Figure 9ag

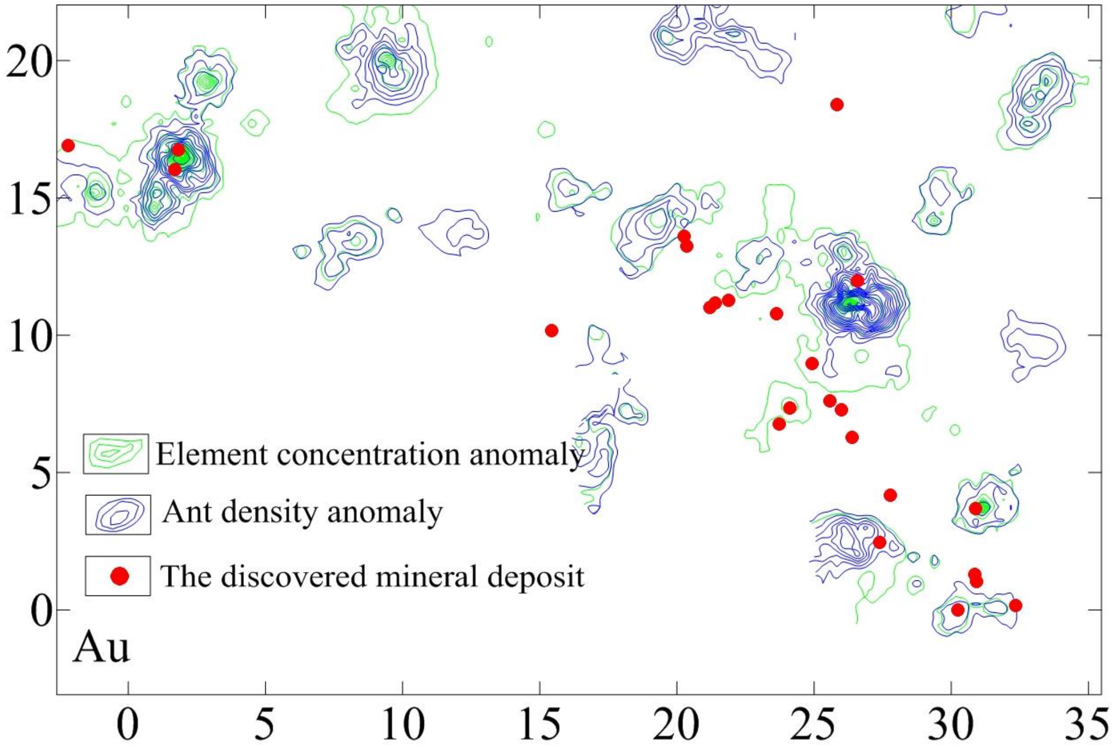

Element concentration anomaly
Ant density anomaly
The discovered mineral deposit
Au
0
5
10
15
20
0
5
10
15
20
25
30
35

Figure 9au

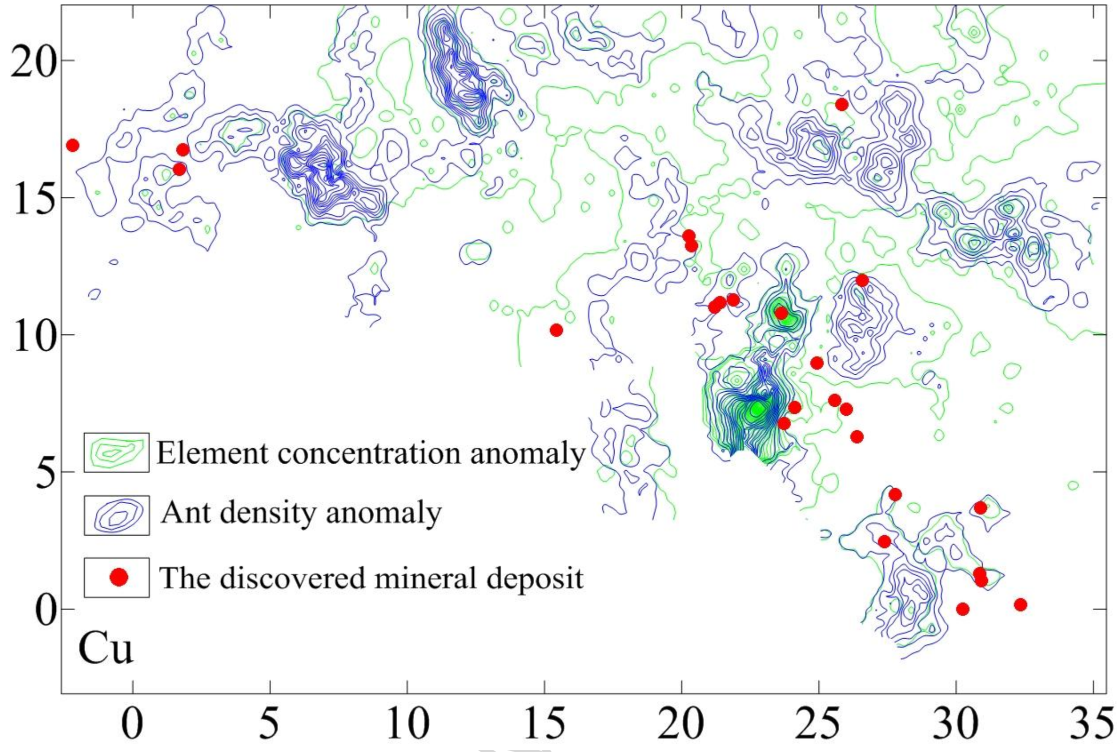
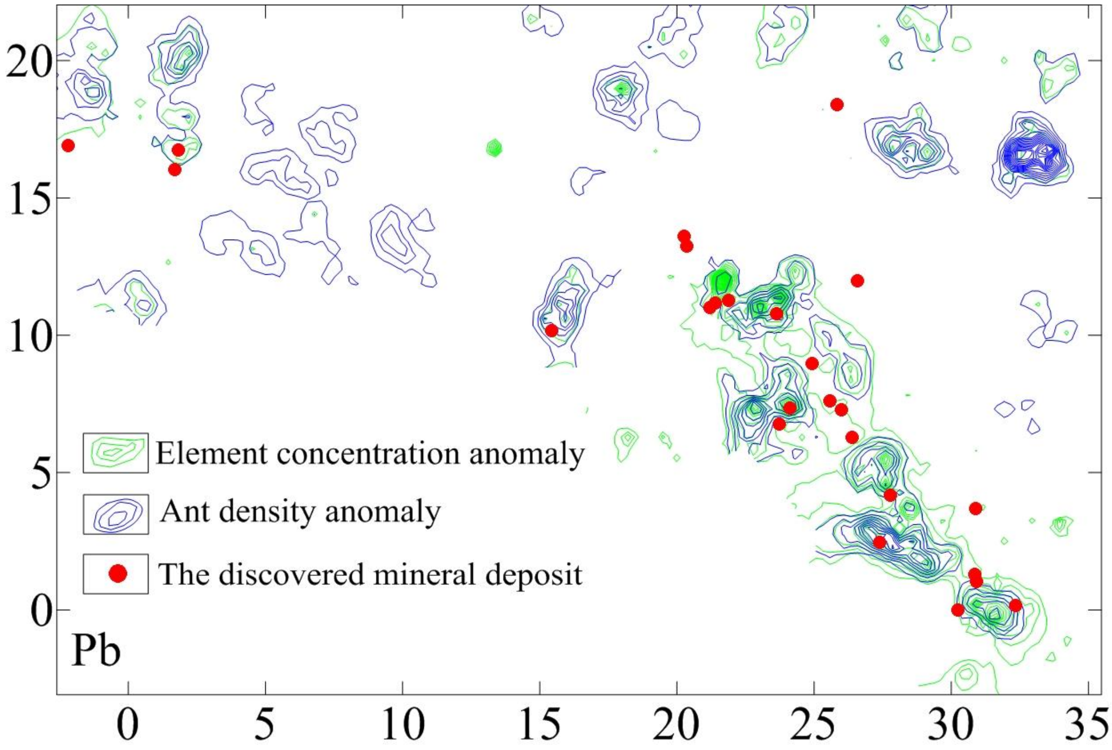

Element concentration anomaly
Ant density anomaly
The discovered mineral deposit
Cu
0
5
10
15
20
0
5
10
15
20
25
30
35

Figure 9cu



20
15
10
5
0
0
5
10
15
20
25
30
35
Element concentration anomaly
Ant density anomaly
The discovered mineral deposit
Pb

Figure 9pb

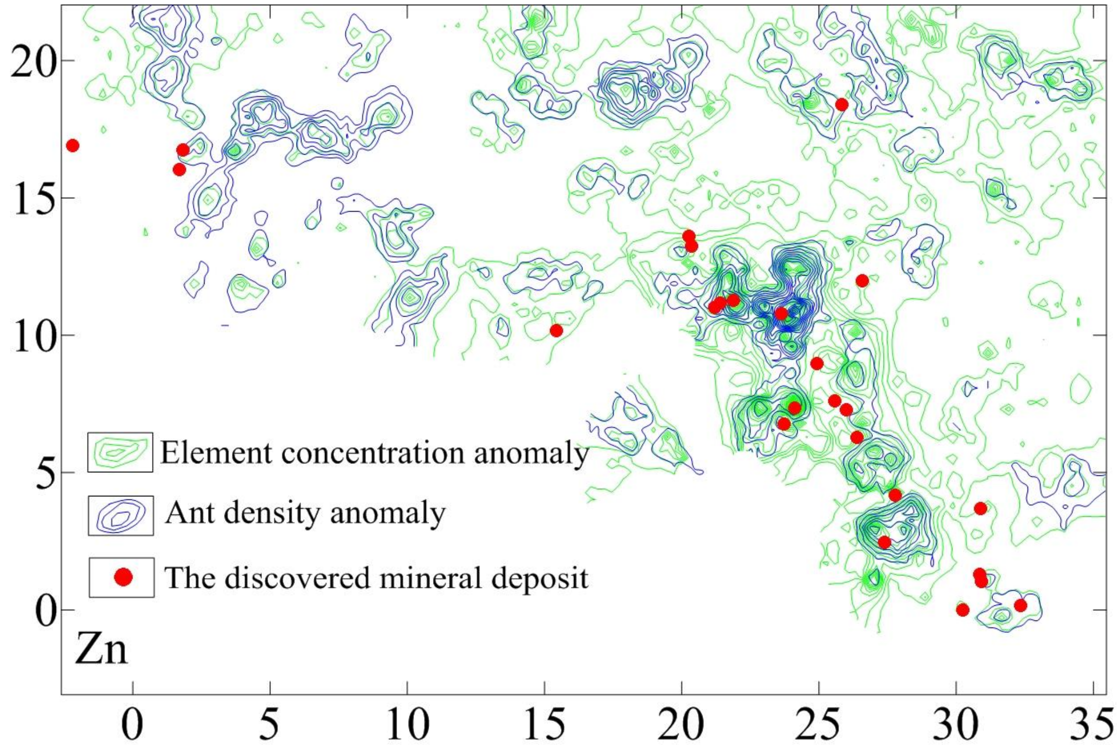

20
15
10
5
0
Zn
0
5
10
15
20
25
30
35
Element concentration anomaly
Ant density anomaly
The discovered mineral deposit

Figure 9zn

Table 1 Python pseudo-code for the ant colony algorithm of geochemical anomaly detection

Input a grid map of *nrow* × *ncol* into buffer [*nrow* × *ncol*]
Initialize n_buffer[*nrow* × *ncol*] to record the number of the ants that passes each grid
Initializing α, β, ρ, *m*, max_*t*, τ_0, min_τ, *c*, min_*d*, *taboo*, max_*repeat*

For *repeat* in range (max_*repeat*):

- Initialize current[*m*] to record the grids that *m* ants locate at step *t*
- Initialize Tao[*nrow** *ncol*, 8] to record τ_{ij}
- Initialize TL[*m*, *taboo*] to record the grids that are visited recently by each ant
- Set *m* ants randomly into the map
- For *t* in range (max_*t*):
    - Initialize deta[*nrow* * *ncol*, 8] to record the updating quantity of $\Delta\tau_{ij}$
    - self.active = 0
    - For *k* in range (*m*):
        - Initialize position[] as null array
        - For *i* in (8):
            - If *i* is not in the TL[*k*, *taboo*]:
                - Add the grid number *i* into position[]
                - Compute ant-routing index using Eq.(5)
                - Record the grid with the biggest ant-routing index
                - *select* = the grid with max index

```
            If position != []:
                Update TL[k, taboo]
                Self.active += 1
                If buffer[select] – buffer[current[k]] > min_d:
                    Compute Δτij (0 ≤ i < nrow* ncol; 0 ≤ j< 8) using Eq. (6)
                    n_buffer[ll] += 1.0
                    current[k] = select
            Else:
                    #Randomly choose one neighbor from position[]
                    select = random.choice(position)
                    Compute Δτij (0 ≤ i < nrow* ncol; 0 ≤ j< 8) using Eq. (6)
                    n_buffer[i] += 1
                    current [k] = select
        For i in range (nrow * ncol):
            For j in range (8):
                Update τij using Eq. (7)
                If τij <= 0.0:
                    τij = min_τ
        if self.active == 0:
            break
For i in range (nrow * ncol):
    n_buffer[i] /= max_repeat
```

End the algorithm

Table 2 AUCs, the standard deviations of *AUC*s, and Z_{AUC}s for Au, Ag, Cu, Pb, and Zn

	Au	Ag	Cu	Pb	Zn
*AUC*s	0.743	0.703	0.574	0.786	0.691
Standard Deviation	0.057	0.059	0.060	0.055	0.059
Z_{AUC}	4.249	3.445	1.242	5.245	3.234

Table 3 The Youden index, optimal threshold, anomaly area percentage, and the percentage of mineral deposits in anomaly area

		Au	Ag	Cu	Pb	Zn
Element concentration data	Youden index	0.202	0.310	0.188	0.460	0.351
	Optimal threshold	2.579	0.050	22.376	26.821	66.727
	Area%	10.1%	5.7%	28.7%	9.0%	28.5%
	Deposit%	36.0%	40.0%	64.0%	60.0%	80.0%
Ant density data	Youden index	0.174	0.198	0.142	0.241	0.304
	Optimal threshold	10.240	8.373	2.827	9.630	4.040
	Area%	4.2%	2.7%	16.4%	10.1%	13.8%
	Deposit%	24.0%	24.0%	40.0%	40.0%	52.0%

Highlights

- Use an ant colony algorithm to detect geochemical element anomalies.
- Use ROC curve to assess the performance of metallogenic indicators.
- Use AUC metric to assess the performance of metallogenic indicators.
- Use the Youden index to determine the optimal threshold of geochemical data.