Computers & Geosciences 86 (2016) 75–82

Contents lists available at ScienceDirect

Computers & Geosciences

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

Case study

Recognition of geochemical anomalies using a deep autoencoder network

Yihui Xiong, Renguang Zuo *

State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form
9 October 2015
Accepted 12 October 2015
Available online 17 October 2015

Keywords:
Autoencoder network
Multivariate geochemical data
Geochemical exploration
Skarn-type iron deposits

ABSTRACT

In this paper, we train an autoencoder network to encode and reconstruct a geochemical sample population with unknown complex multivariate probability distributions. During the training, small probability samples contribute little to the autoencoder network. These samples can be recognized by the trained model as anomalous samples due to their comparatively higher reconstructed errors. The southwestern Fujian district (China) is chosen as a case study area. A variety of learning rates, iterations, and the size of each hidden layer are constructing and training the deep autoencoder networks on all the geochemical samples. The reconstruction error (or, anomaly score) of each training sample is used to recognize multivariate geochemical anomalies associated with Fe polymetallic mineralization. By comparing the results obtained with a continuous restricted Boltzmann machine, we conclude that the autoencoder network can be trained to recognize multivariate geochemical anomalies. Most of the known skarn-type Fe deposits are located in areas with high reconstruction errors or anomaly scores in the anomaly map, indicating that these anomalies may be related to Fe mineralization.

1. Introduction

Separating and recognizing geochemical anomalies from the geochemical background is one of the major tasks in geochemical exploration. In past decades, geochemical anomalies have been recognized using a number of methods (Harris et al., 1999, 2000). These include calculating the mean ± 2 standard deviation (Hawkes and Webb, 1962; Gałuszka, 2007), histograms, quantile–quantile plots, boxplots (Tukey, 1997), probability graphs (Sinclair, 1974, 1991), univariate analysis (Govett et al., 1975), and multivariate data analysis methods (e.g., Ali et al., 2006; Grunsky et al., 2009; Yousefi et al., 2012; Zuo, 2014). Moreover, fractal/multifractal models are powerful tools that have been widely used to recognize geochemical anomalies (Cheng et al., 1994, 2000; Li et al., 2003; Cheng, 2007, 2012; Zuo et al., 2009, 2012, 2013, 2015a; Luz et al., 2014; Carranza, 2010; Zuo, 2011, 2014; Zuo and Wang, 2015; J. Wang and Zuo, 2015). Spatial statistical methods, such as the moving average technique, spatial factor analysis (Grunsky and Agterberg, 1988), and multivariate geostatistics (Wackernagel, 2003) have been used to recognize multivariate geochemical anomalies by taking into account spatial correlations and variations among neighboring samples, in addition to concentration value frequency distributions.

The multivariate methods usually require multivariate geochemical data satisfying a known statistical distribution, such as a multivariate normal distribution. However, various geological events and complex geological processes usually result in highly complex spatial and frequency properties (Cheng, 2007; Zuo et al., 2009), and this means that the complex multivariate probability distribution of any geochemical data is often unknown (Chen et al., 2014). In such cases, geochemical data show neither a normal nor a lognormal distribution (Reimann and Filzmoser, 2000). Therefore, most ordinary multivariate statistical techniques that require assumption of multivariate normal distribution are limited in terms of the kind of geochemical data.

Restricted Boltzmann Machines (RBMs) (Murray, 2001) have been widely applied for a variety of learning tasks, including multivariate distribution modeling, high-dimensional temporal sequence modeling, and the construction of deep architectures (Hinton and Salakhutdinov, 2006; Hinton, 2010). Generally, RBMs work well for binary or Gaussian inputs (Hinton and Salakhutdinov, 2006), but they are limited for other kinds of input variables, such as continuous-value non-Gaussian inputs (Chen and Murray, 2003; Chen et al., 2014). Continuous restricted Boltzmann machines (CRBMs) (Chen and Murray, 2003) can be trained iteratively using the minimizing contrastive divergence (MCD) (Hinton, 2002), and used to model multivariate distributions and complex high-dimensional continuous data. CRBMs have been successfully used to fingerprint orientation fields learning (Sahasrabudhe and Namboodiri, 2013), multivariate geochemical anomaly recognition

* Corresponding author.
E-mail address: zrguang@cug.edu.cn (R. Zuo).

http://dx.doi.org/10.1016/j.cageo.2015.10.006
0098-3004/

(Chen, 2014; Chen et al., 2014) and for construction of deep architectures (Valentine and Trampert, 2012).

Deep Belief Network (DBN) (Hinton et al., 2006; Hinton and Salakhutdinov, 2006) is a probabilistic generative model composed of stacked RBMs. A DBN can recognize high level features of the inputs with the RBMs using a greedy layer-wise unsupervised training algorithm. The deep autoencoder network based on DBN is trained by minimizing the difference between the input and the output data (Hinton and Salakhutdinov, 2006). Hinton and Salakhutdinov (2006) suggested the use of RBMs for deep autoencoder networks with binary inputs and outputs. Furthermore, it is possible to use a CRBM (Chen and Murray, 2003), rather than RBMs as the unsupervised building block of the autoencoder network (Valentine and Trampert, 2012; Sun et al., 2014). During the training of a deep autoencoder network, small probability samples contribute little to the autoencoder network. Consequently, these small probability samples are poorly encoded and reconstructed by the trained model (Valentine and Trampert, 2012; Chen, 2014; Chen et al., 2014; Sakurada and Yairi, 2014). In geochemical exploration, multivariate geochemical anomaly samples are small probability samples, by using a deep autoencoder network they will be poorly encoded and reconstructed and have comparatively higher reconstruction errors. Accordingly, the reconstruction errors can be used to recognize a multivariate geochemical anomaly. The aim of this study is to demonstrate that the deep autoencoder network is a powerful tool for recognition of multivariate geochemical anomalies in a case study from southwestern Fujian district (China).

2. Deep neural network: deep autoencoder network

Multi-layered neural networks have been available for several years already. An autoencoder network is a feed forward neural network with one or several layers designed to minimize the difference between the input and output. The main difference between an autoencoder network and a traditional network is the size of the output layer. The size of an autoencoder's output layer is always the same as the input layer. An autoencoder network is composed of an encoder and a decoder. The encoder network transforms the input data into new code, and the decoder network recovers the data from the code. Fig. 1 shows a so-called 5-4-3-2-3-4-5 autoencoder network. However, training algorithms are unable to train such deep networks effectively (Bengio, 2009; Bengio et al., 2007). This is because the standard learning strategy used for training a neural network is the backpropagation algorithm based on gradient descent and the weights of the network is random initialized (Rumelhart et al., 1988). As the network gets deeper, the propagated error to the first hidden layer becomes very small, thus the parameters of the first few layers change very litter in training (Larochelle et al., 2009). Nevertheless, interest in such deep neural networks has been revived with the proposition of novel approaches (Hinton et al., 2006; Hinton and Salakhutdinov, 2006; Bengio et al., 2007). The autoencoder networks have been applied to detect anomalies based on the reconstruction error (Valentine and Trampert, 2012; Fiore et al., 2013; Sun et al., 2014). The reconstruction error, also called the anomaly score, can be computed as follows:

$$Err(i) = \sqrt{\sum_{j=1}^{D} (R_{ij} - I_{ij})^2}, \tag{1}$$

where D denotes the dimensions of the data and R_i is the reconstructed data of the input data I_i. For geochemical data, the reconstruction error is high for anomalous samples and low for background samples.

Training the deep autoencoder network consists of three phases. In the first phase, pre-training, each CRBM is pre-trained respectively for initialized weights. In the second phase, all CRBMs are unrolled to construct the autoencoder network. Finally, the entire deep network is fine-tuned via backpropagation to adjust all of the parameters simultaneously.

2.1. Network pre-training: CRBMs

CRBMs (Chen and Murray, 2003) are variants of the RBM, and are typically stochastic two-layer neural networks that consist of a visible layer and a hidden layer with symmetric inter-layer connections defined by a weight matrix {W} (Fig. 2). The continuous visible and hidden units are introduced by adding zero-mean Gaussian noise to the input of a sample sigmoid unit.

Let v_i and h_j represent the states of the visible units i and the hidden units j, respectively, and let w_{ij} represent the bi-directional connection between units v_i and h_j. Given the visible units, the hidden units can be updated as follows:

$$h_j = f\left(\sum_i w_{ij} v_i + b_j + \sigma N(0, 1)\right), \tag{2}$$

where b_j denotes the bias of the hidden unit j.

A similar rule can be used to update the visible units with the hidden units:

$$v_i = f\left(\sum_j w_{ji} h_j + b_i + \sigma N(0, 1)\right), \tag{3}$$

where b_i is the bias of the visible unit i.

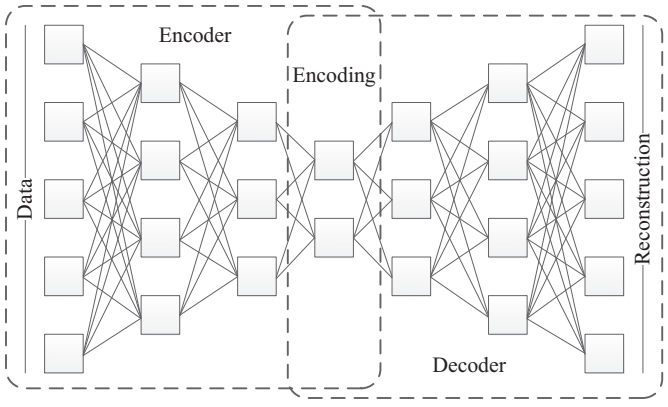

Fig. 1. An architecture of autoencoder network (modified from Valentine and Trampert (2012)).

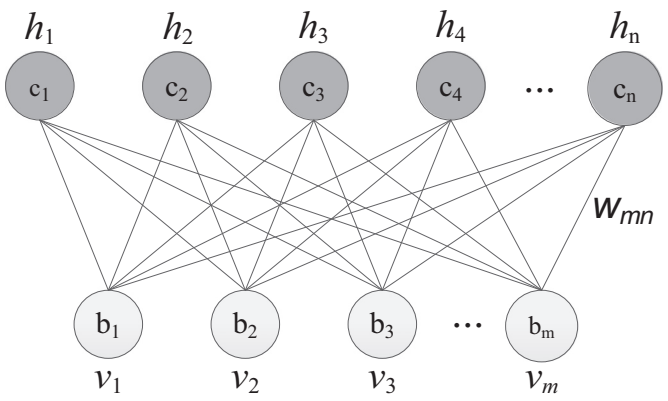

Fig. 2. An architecture of continuous restricted Boltzmann machines (CRBM).

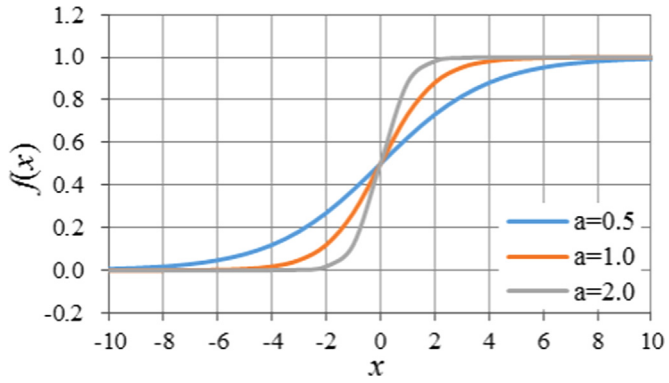

Fig. 3. A plot of sigmoid function $f(x)$ versus x for three values of a, and with $\theta_L=0$, and $\theta_H=1$.

In Eqs. (2) and (3), the constant σ and $N(0, 1)$ together constitute the noise component, which allows a CRBM to perform an analogous probabilistic neural computation without quantization, avoiding any unnecessary loss of information that would result using a binary RBM (Murray, 2001; Hinton, 2002). Here, $N(0, 1)$ is a Gaussian random variable with zero mean and unit variance, and $f(x_i)$ is a logistic function defined as

$$f(x_i) = \theta_L + (\theta_H - \theta_L)\cdot\frac{1}{1+\exp(-a_i x_i)}, \tag{4}$$

where θ_L and θ_H are two constants defining the asymptotes of the sigmoid function and a_i is the "noise-control" parameter that controls the slope of the sigmoid function and the nature and extent of the unit's stochastic behavior. It controls a smooth transition from noise-free, deterministic behavior to binary-stochastic behavior while a_i increases. The plot for the sigmoid function (Fig. 3) approaches a step function as a_i increases.

The CRBM can be trained by a MCD learning rule (Hinton, 2002). The weight (w_{ij}), bias (b_i), and noise-control parameter (a_i) are updated after each epoch. The updated rules can be expressed as follows:

$$\Delta w_{ij} = \eta_w\left(\langle v_i h_j\rangle - \langle \hat{v}_i \hat{h}_j\rangle\right), \tag{5}$$

$$\Delta a_i = \frac{\eta_a}{a_i^2}\left(\langle s_i\rangle^2 - \langle \hat{s}_i\rangle^2\right), \tag{6}$$

and

$$\Delta b_i = \eta_b\left(\langle s_i\rangle - \langle \hat{s}_i\rangle\right), \tag{7}$$

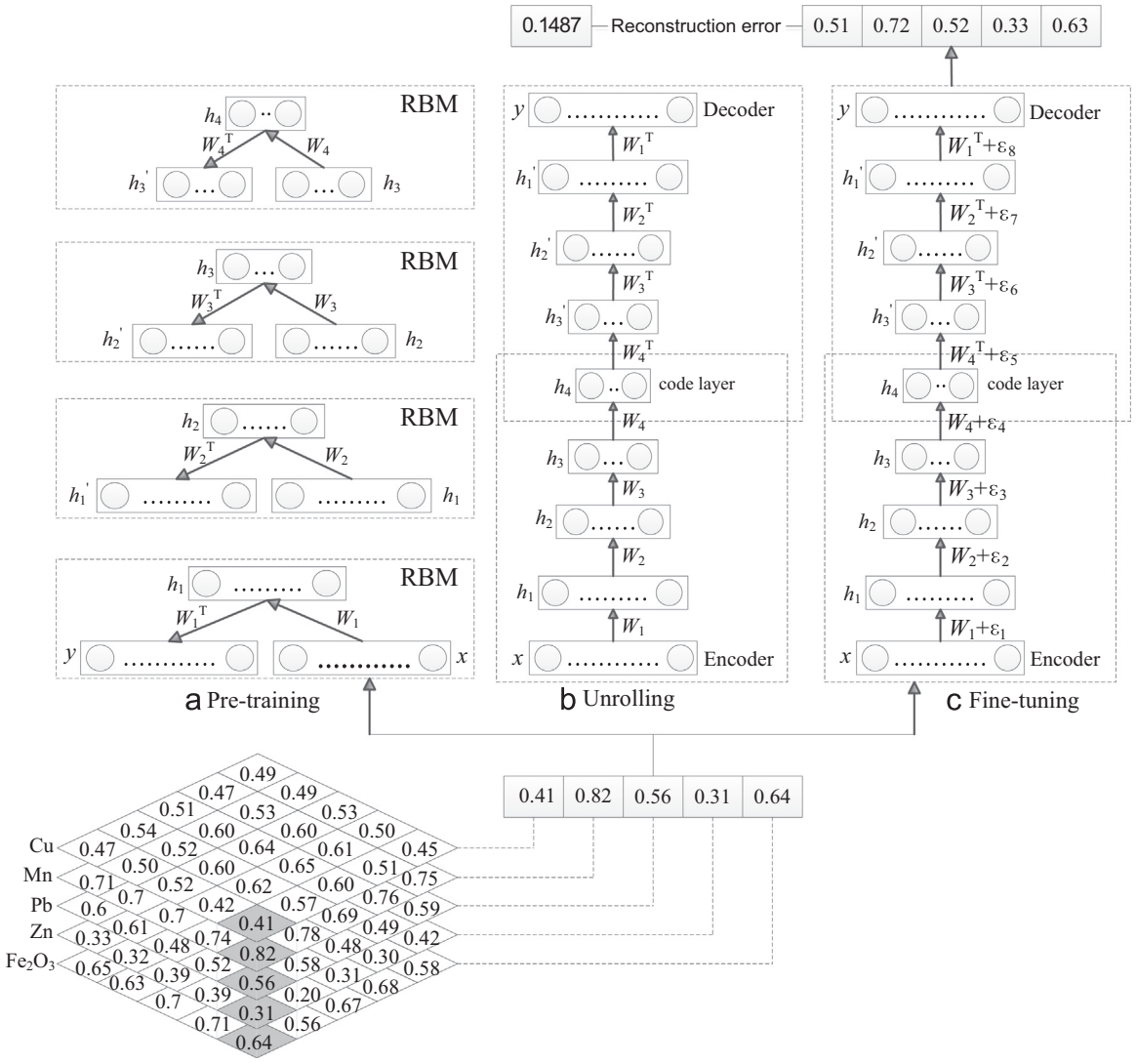

Fig. 4. The construction and calculation process of autoencoder network.

where $\langle \bullet \rangle$ refers to the expected value over all training samples, and η_w, η_a, and η_b denote the learning rate for weights, noise-control, and biases, respectively. Here, $\hat{v}_i$ refers the one-step reconstructed continuous states of the visible units, $\hat{h}_j$ denotes the one-step reconstructed continuous states of the hidden units, and s_i represents v_i for the visible units and h_j for the hidden units.

Training the CRBM involves attempting to find and enhance the correlations between the values of visible and hidden units. Each CRBM contains a fraction of the number of parameters in the autoencoder network (see Fig. 4a). Therefore, the training of CRBM may be a useful technique to pre-train each layer of autoencoder network separately.

2.2. Network fine-tuning

After pre-training, the CRBMs are unrolled to create a deep autoencoder network whose lower layers use the matrices to encode the input, and whose upper layers use the matrices in reverse to decode it (Hinton and Salakhutdinov, 2006) (Fig. 4b). Backpropagation with a gradient descent technique is used to fine-tune the entire network (Bengio et al., 2007; Larochelle et al., 2009). However, fine-tuning with backpropagation does not extract the features that are used in the hidden layers of the autoencoder network. Rather, it merely modifies the features extracted at the pre-training phase with the aim of minimizing reconstruction errors (Hinton and Salakhutdinov, 2006) (Fig. 4c).

For each output unit i in L (i.e., the output layer), the backward pass is as follows:

$$\delta_i^L = (y_i - x_i^L) \cdot a_i^L \cdot f'(\theta_i^L), \tag{8}$$

and for $l = L-1, L-2, L-3, \ldots, 2$, the backward pass in each unit i of layer l can be expressed as

$$\delta_i^l = a_i^l \cdot f'(\theta_i^l) \cdot \sum_{t=1}^{n_{l+1}} \delta_t^{l+1} w_{ti}^{l+1}, \tag{9}$$

The notations for Eqs. (8) and (9) are listed as follows:

y_i: the expected output value for unit i in the output layer;

x_i^l: the actual output for the i-th unit in layer l;

n_l: the number of units in layer l;

a_i^l: the noise-control parameter for the i-th unit in layer l;

θ_i^l: the product of the noise-control parameter and the total input for unit i given by

$$\theta_i^l = a_i^l \left(\sum_{j=1}^{n_{l-1}} w_{ij}^l x_j^{l-1} + b_j + \sigma N(0, 1) \right);$$

$f'(x)$: the derivative of the sigmoid function, where the equation is $f'(x) = f(x)(1 - f(x))$.

w_{ij}^l: the weight between i-th unit in layer l and j-th unit in layer $l-1$;

According to Eqs. (8) and (9), the partial derivatives are computed, and the updating parameters are as follows:

$$\Delta w_{ij}^l = \eta_w \cdot \delta_i^l \cdot x_j^{l-1}, \tag{10}$$

$$\Delta b_i^l = \eta_b \cdot \delta_i^l, \tag{11}$$

and

$$\Delta a_i^l = \eta_a \cdot \frac{\delta_i^l}{(a_i^l)^2} \cdot \theta_i^l, \tag{12}$$

where η_a is the learning rate for the noise-control parameter.

Pre-training a stack of CRBMs (Fig. 4) is needed in order to discover a suitable region of the parameter space, and these are used as the initial parameters for the deep autoencoder network. Based on these parameters, the autoencoder network is fine-tuned using backpropagation algorithm to minimize reconstruction errors and to ensure that the output of the network is approximately equal to the input.

3. Study area and data

3.1. Geological setting and mineralization

The southwestern Fujian district is one of the most important iron polymetallic metallogenic belts in China (Fig. 5). Extensive Yanshanian granitoids were emplaced in this region, and cropped out in the iron deposit districts. Recent zircon U–Pb dating of granites in Makeng, Dapai, Luoyang, and the Zhongjia Fe deposit yielded an age of 125–145 Ma (Zhang et al., 2015a), 134 Ma (Yuan et al., 2013), 131–132 Ma (Zhang et al., 2012), and 99 Ma (Yang et al., 2008), respectively, indicating a close relationship between the Fe mineralization and Yanshanian intrusions. The dominant lithologies in the southwestern Fujian are late Paleozoic marine sedimentary rocks, and the primary ore-hosted strata are middle–lower Carboniferous carbonate and clastic rock formations discontinuously distributed along the regionally NE-controlling Zhenghe–Dapu Fault (Han and Ge, 1983). Other conspicuous structures are NNE-trending Nanping–Ninghua Fault and NW-trending Shanghang–Yunxiao Fault. The three major faults define the boundaries of the southwestern Fujian. A genetic model of these deposits has been built and it indicates that Yanshanian granitoids provided heat and metal-bearing magmatic–hydrothermal fluids for the Fe mineralization, the secondary faults provided the pathway for the migration of the fluids, and the contact zones between granitoids and the late Paleozoic sedimentary rocks or contact zones between the carbonate and clastic rocks provided the space for mineralization (Zhang and Zuo, 2014; Zhang et al., 2015a, 2015b). Previous research indicates that the skarn-type iron deposits are strongly correlated to the element association of Cu–Pb–Mn–Zn–Fe_2O_3 (H. Wang and Zuo, 2015).

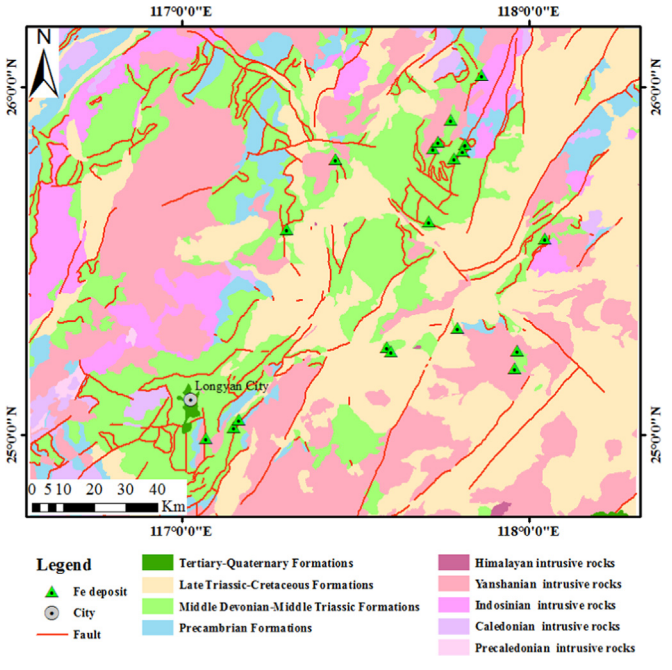

Fig. 5. Simplified geological map of the southwestern Fujian province in China (modified from Geological Survey Institute of Fujian, 2011).

3.2. Preprocessing of geochemical data

In the study, stream sediment geochemical data have been collected at a scale of 1:2000,000 by the Chinese National Geochemical Mapping Project (Xie et al., 1997). The stream sediment samples were collected at a 2 km × 2 km grid within drainage basins, and analyzed for 39 elements which were determined based on inductively coupled plasma-mass spectrometry (ICP-MS), X-ray fluorescence (XRF), and inductively coupled plasma-atomic emission spectrometry (ICP-AES) as the backbone combined with other methods. The concentration value for Fe_2O_3 was determined by XRF. The concentration value for Cu and Pb was determined by ICP-MS. The concentration value for Mn and Zn was determined by ICP-AES. In this study area, the stream sediment data have been applied to recognize geochemical anomalies related to Fe polymetallic deposits (Wang et al., 2015a, 2015b; Zuo et al., 2015b; Zhang et al., 2015c).

The geochemical data are composition data and should be processed prior to data analysis due to the closure problem (Aitchison, 1986; Aitchison et al., 2000; Egozcue et al., 2003; Filzmoser and Hron, 2008; Filzmoser et al., 2009; Carranza, 2011; Zuo, 2014). The isometric logratio transformation (*ilr*) (Egozcue et al., 2003) was applied to process the raw geochemical data to address the closure problem.

Before training the CRBM and the autoencoder network, the measured concentration values of the five elements (Cu, Pb, Mn, Zn, Fe_2O_3) related to Fe polymetallic mineralization were normalized into the range [0, 1], using the following function:

$$\tilde{y}_i = \frac{y_i - y_{min}}{y_{max} - y_{min}},\ i \in [1, 6682], \tag{13}$$

where i denotes the sample number, and y_{max} and y_{min} denote the maximum and the minimum values of y, respectively.

4. Results and discussion

4.1. Autoencoder network architectures selection

Choosing a suitable network architecture is critical to the performance of the autoencoder network. To obtain the optimum network architecture, the learning rates for pre-training and fine-tuning, the number of iterations and the size of each hidden layer with experiments on the geochemical data should be investigated. Before training of the network, several parameters for the autoencoder network and the CRBM must be initialized, such that

$$w_{ij} = \psi(0, 0.01);\quad b_i = \psi(0, 0.01);\quad a_i = 1, \tag{14}$$

where $\psi(\mu, \sigma)$ represents a random sample drawn from a Gaussian distribution.

The relationship between the iterations and reconstruction errors during pre-training and fine-tuning for a number of different choices of learning rates (Fig. 6) show the effects of altering the pre-training and fine-tuning learning rates, and a higher learning rate results in a more rapid reduction in reconstruction errors. However, if the learning rates are too high, the reconstruction errors usually increase dramatically and the weights risk exploding (Hinton, 2010). As expected, when the learning rates become large enough, the existing reconstruction errors fluctuate during the training, and this illustrates the phenomenon of "overfitting": at some point during the training, the network begins to learn the intricacies of the training dataset at the cost of its ability to represent unseen examples (Valentine and Trampert, 2012).

As a consequence, the learning rate for the CRBM was set to 0.3, and the learning rate for fine-tuning was set to 0.8. The

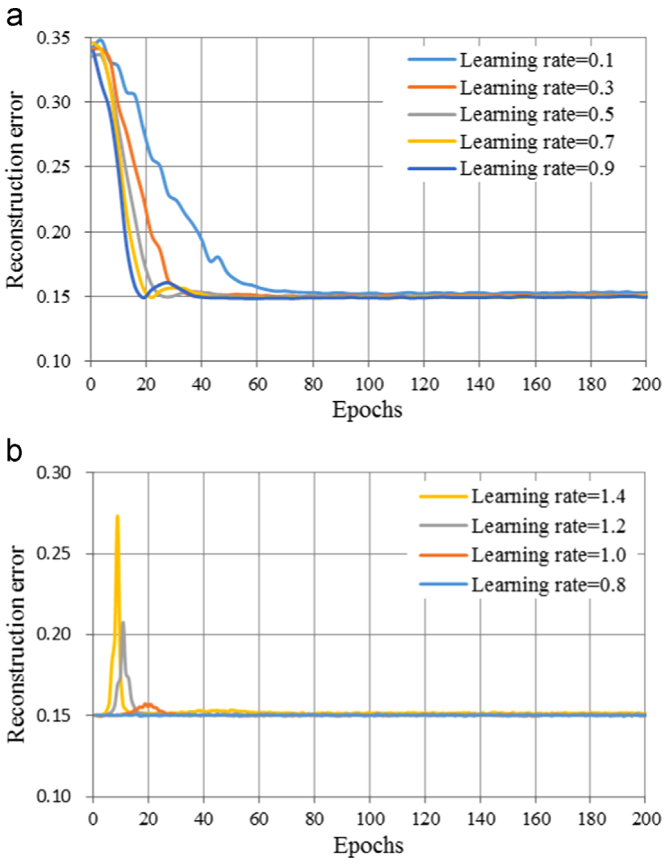

Fig. 6. Plots showing the effects of learning rate on pre-training (a) and fine-tune (b).

reconstruction errors have not decreased significantly after fine-tuning the network, indicating that the network is well-trained during the pre-training phase, and that the parameters for the autoencoder network have undergone only slight changes (Fig. 6a and b).

Different iterations can render the autoencoder network under-trained, well-trained, or over-trained (i.e., overfitted). A well-trained network concentrates on the general solution. However, an under-trained network concentrates on a local solution. Over-trained network contains both general and specific information. Therefore, it is crucial to determine the appropriate number of iterations in order to correctly train the network. The average reconstruction error is used as an indicator to produce a well-trained network. If the average reconstruction error reaches a minimum and keeps stable, the network is considered well-trained. The average reconstruction error remains stable at approximately 80 epochs, meaning that the network is considered "nearly well-trained." (Fig. 6). In these experiments, the multivariate geochemical anomaly recognition was implemented with an architecture at 200 epochs during pre-training.

Because the autoencoder network is stacked with CRBMs, the depth of the network should be considered. It is therefore important to establish whether too many CRBMs will degrade the generalization performance. There are various reasons why this might occur, given that more parameters are added into the formula of the model, while the hidden layer is increased. In such circumstances, more degrees of freedom are provided to fit the model, and this makes it susceptible to overfitting (Larochelle et al., 2009). In this experiment, the autoencoder network was stacked with four CRBMs (see Fig. 4).

After determining the optimal depth for the autoencoder network, the appropriate size for the different layers must be

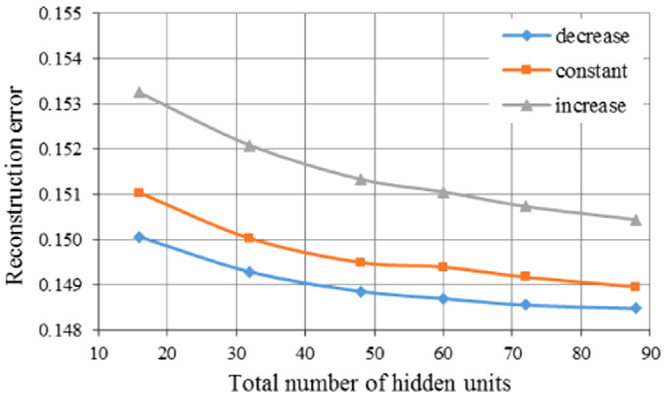

Fig. 7. Plots showing reconstruction errors of autoencoder network for three kinds of architectures.

established. In building the network, if there are an insufficient number of hidden units in the CRBMs, the network will not have enough recourse to learn the general features of the training sample population. If there are too many hidden units, on the other hand, it will increase the training time without necessarily producing a corresponding increase in performance. In fact, too many hidden units can result in overfitting (Chen, 2014; Chen et al., 2014). In the experiment, we considered three general cases, namely, the size of different layers in terms of increasing it (doubles), maintaining a constant size, or decreasing it (halves) while increasing of the layer index (Larochelle et al., 2009). These three cases were explored with a different total number of hidden units in the network. In Fig. 7, we can observe that the best performance from altering the size of the network is the one that resulted from decreasing the size of the hidden layers. Based on the previous discussion, we ultimately created a 40-20-10-5-10-20-40 autoencoder network.

4.2. Recognition of multivariate geochemical anomalies

The geochemical data were normalized within the range [0,1]. Thus, θ_L and θ_H in Eq. (4) were set to 0 and 1, respectively. We created a 40-20-10-5-10-20-40 autoencoder network and set the CRBM learning rate to $\eta^{(C)} = 0.3$ and the fine-tuning learning rate to $\eta^{(A)} = 0.8$. We generated four CRBMs, and trained each with 200 iterations. The following two parameters were chosen empirically: (a) learning cost=0.00001; and (b) learning moment=0.9. The autoencoder network was then used to encode all 6682 geochemical samples. The reconstruction error for each geochemical sample was computed using Eq. (1).

The geochemical anomalies recognized by the autoencoder network (Fig. 8) show that most of the known skarn-type Fe deposits are located in areas associated with high reconstruction errors in the anomaly map, indicating that these anomalies may be related to Fe mineralization. The other areas with high reconstruction errors indicate the potential for discovering Fe deposits in the study area.

The results were compared with the resulting CRBM (Figs. 8 and 9). The spatial distribution of the geochemical anomalies obtained with the autoencoder network and the CRBM were similar in the study area, indicating that both the CRBM and the autoencoder network are useful tools for recognition of multivariate geochemical anomalies.

Student's *t*-values (Agterberg et al., 1990; Bonham-Cater, 1994) were used to quantitatively measure the spatial correlation between the multivariate geochemical anomaly and the locations of the known Fe-deposits. Student's *t*-values that are larger than 1.96

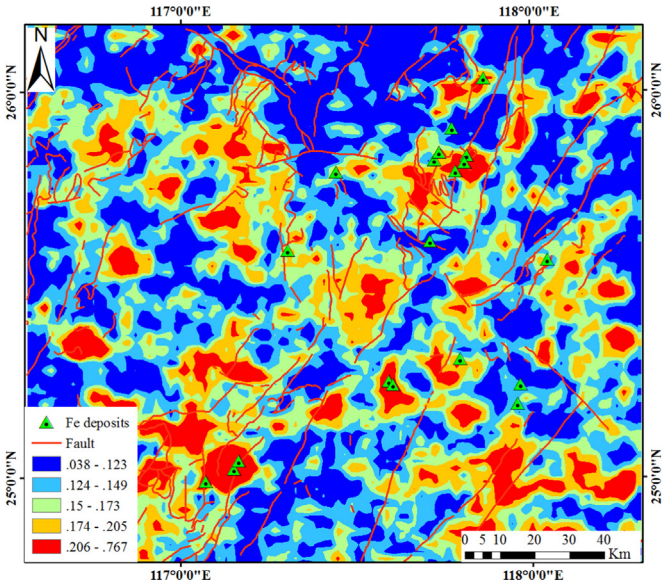

Fig. 8. Geochemical anomaly map obtained by the autoencoder network.

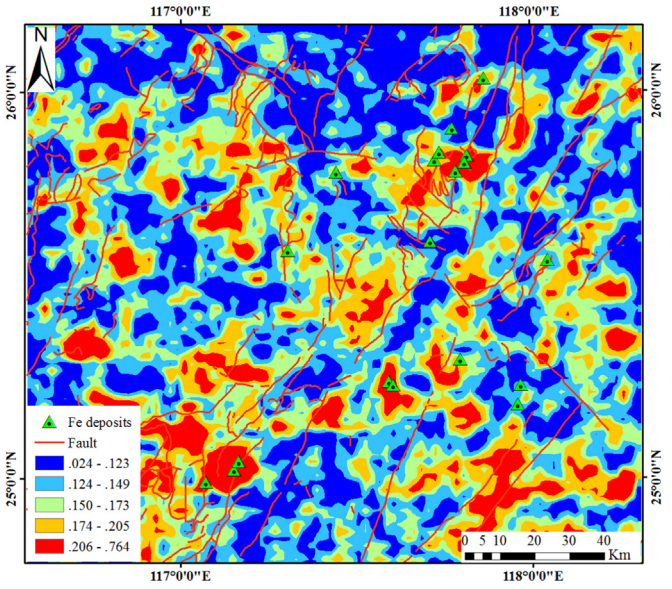

Fig. 9. Geochemical anomaly map obtained by the continuous restricted Boltzmann machines.

indicate a statistically significant spatial correlation. Fig. 10a shows that the maximum *t*-value is 4.48 at geochemical anomaly value of 0.25, indicating the geochemical anomalies have strong spatial correlation with the location of Fe deposits. The *t*-values reach 1.96 at geochemical anomaly value of 0.16. These two values (0.16 and 0.25) divide the anomaly values into three classes: the high anomaly (≥ 0.25), moderate anomaly (0.16–0.25), and low anomaly (≤ 0.16) (Fig. 10b). The high anomaly area occupies 2.4% of the total area and contains 31.5% of the total number of known Fe deposits. The moderate anomaly occupies 34.1% of the total area and contains 68.4% of known Fe deposits.

5. Conclusions

In this paper, the autoencoder network was applied to recognize geochemical anomalies. It is crucial to select a suitable network architecture in order to enhance the performance of the

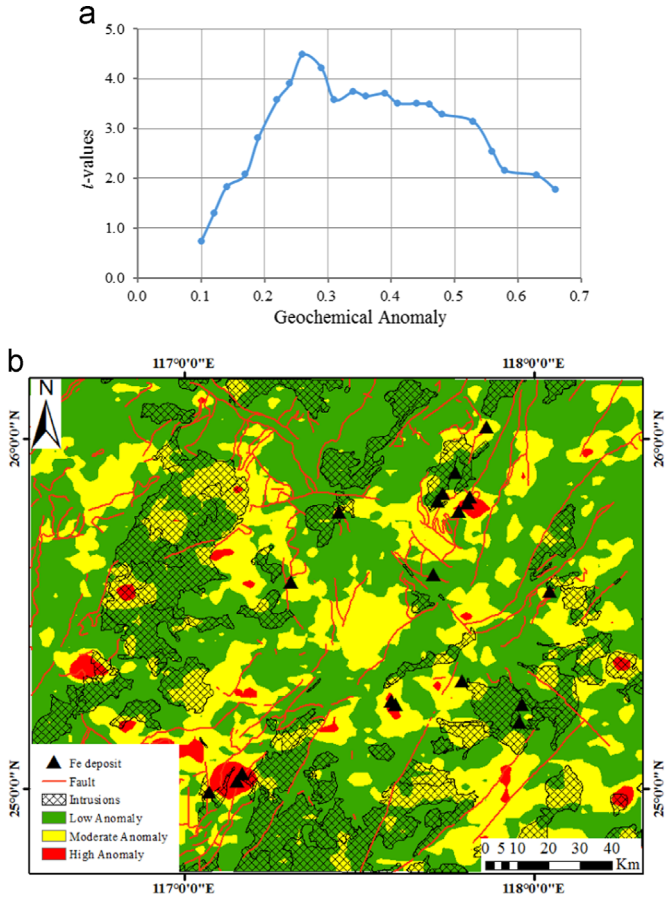

Fig. 10. Plot of Student's *t*-values vs geochemical anomaly (a), and geochemical anomaly map (b).

autoencoder network. However, what counts as an optimal network is likely to vary depending on the precise problem under consideration. In this study, the parameters for the network, such as the learning rate, the number of iterations, and the size of each hidden layer, were determined based on whether they minimized reconstruction errors. When the reconstruction errors are minimal and stable, the autoencoder network and its corresponding parameters are optimal for modeling populations of geochemical samples.

The reconstruction error based on the optimal autoencoder network is a useful indicator of multivariate geochemical anomalies. Based on this indicator, geochemical anomalies were identified in the southwestern Fujian district (China). The results from the autoencoder network were compared with the results from a CRBM, and the spatial distribution of the geochemical anomalies obtained by autoencoder network and the CRBM were similar in the study area. There is a strong spatial correlation with the known Fe deposits with the anomalous areas, indicating that these two methods used in this study are powerful tools for recognition of multivariate geochemical anomalies.

Acknowledgments

The authors thank Dr. John Carranza and two anonymous reviewers for their critical reviews and constructive comments that improved the manuscript. This research benefited from the joint financial support from the National Natural Science Foundation of China (No. 41522206).

References

Agterberg, F.P., Bonham-Carter, G.F., Wright, D.F., 1990. Statistical pattern integration for mineral exploration. In: Gaal, G., Merrian, D.F. (Eds.), Computer Applications in Resource Estimation Prediction and Assessment for Metals and Petroleum. Pergamon Press, Oxford, New York, pp. 1–21.
Aitchison, J., 1986. The statistical analysis of compositional data. Chapman & Hall, London, p. 416.
Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J., Pawlowsky-Glahn, V., 2000. Logratio analysis and compositional distance. Math. Geol. 32, 271–275.
Ali, K., Cheng, Q., Li, W., Chen, Y., 2006. Multi-element association analysis of stream sediment geochemistry data for predicting gold deposits in south-central Yunnan Province, China. Geochem. Explor. Environ. Anal. 6, 341–348.
Bengio, Y., 2009. Learning deep architectures for AI. Found. Trends Mach. Learn. 2, 1–127.
Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2007. Greedy layer-wise training of deep networks. Adv. Neural Inf. Process. Syst. 19, 153–160.
Bonham-Cater, G.F., 1994. Geographic Information Systems for Geoscientist: Modeling with GIS. Pergamon Press, Oxford, p. 398.
Carranza, E.J.M., 2010. Catchment basin modelling of stream sediment anomalies revisited: incorporation of EDA and fractal analysis. Geochem. Explor. Environ. Anal. 10, 365–381.
Carranza, E.J.M., 2011. Analysis and mapping of geochemical anomalies using log-ratio-transformed stream sediment data with censored values. J. Geochem. Explor. 110, 167–185.
Chen, H., Murray, A.F., 2003. Continuous restricted Boltzmann machine with an implementable training algorithm. Vision Image Signal Process. IEEE Process. 150, 153–158.
Chen, Y., 2014. Mineral potential mapping with a restricted Boltzmann machine. Ore Geol. Rev. 71, 749–760.
Chen, Y., Lu, L., Li, X., 2014. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. J. Geochem. Explor. 140, 56–63.
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, 314–324.
Cheng, Q., 2012. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. J. Geochem. Explor. 122, 55–70.
Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. J. Geochem. Explor. 51, 109–130.
Cheng, Q., Xu, Y., Grunsky, E.C., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. Nat. Resour. Res. 9, 43–52.
Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35, 279–300.
Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. Math. Geosci. 40, 233–248.
Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. Environmetrics 20, 621–632.
Fiore, U., Palmieri, F., Castiglione, A., De Santis, A., 2013. Network anomaly detection with the restricted Boltzmann machine. Neurocomputing 122, 13–23.
Gałuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environ. Geol. 52, 861–870.
Govett, G., Goodfellow, W., Chapman, R., Chork, C., 1975. Exploration geochemistry distribution of elements and recognition of anomalies. Math. Geol. 7, 415–446.
Grunsky, E.C., Agterberg, F.P., 1988. Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario. Math. Geol. 20, 825–861.
Grunsky, E.C., Drew, L.J., Sutphin, D.M., 2009. Process recognition in multi-element soil and stream-sediment geochemical data. Appl. Geochem. 24, 1602–1616.
Han, F., Ge, C., 1983. Geological and geochemical features of submarine volcanic hydrothermal-sedimentary mineralization of Makeng iron deposit, Fujian Province. Bull. Inst. Miner. Deposits Chin. Acad. Geol. Sci. 7, 1–118 (in Chinese with English abstract).
Harris, J., Wilkinson, L., Grunsky, E., 2000. Effective use and interpretation of lithogeochemical data in regional mineral exploration programs: application of Geographic Information Systems (GIS) technology. Ore Geol. Rev. 16, 107–143.
Harris, J., Wilkinson, L., Grunsky, E., Heather, K., Ayer, J., 1999. Techniques for analysis and visualization of lithogeochemical data with applications to the Swayze greenstone belt, Ontario. J. Geochem. Explor. 67, 301–334.
Hawkes, H.E., Webb, J.S., 1962. Geochemistry in Mineral Exploration. Harper and Row, New York, NY.
Hinton, G.E., 2002. Training products of experts by minimizing contrastive divergence. Neural Comput. 14, 1771–1800.
Hinton, G.E., 2010. A practical guide to training restricted Boltzmann machines, Technical Report, University of Toronto, version 1.
Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.
Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P., 2009. Exploring strategies for training deep neural networks. J. Mach. Learn. Res. 10, 1–40.
Li, C., Ma, T., Shi, J., 2003. Application of a fractal method relating concentrations and distances for separation of geochemical anomalies from background. J.

Geochem. Explor. 77, 167–175.
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. Nat. Resour. Res. 23, 195–215.
Murray, A.F., 2001. Novelty detection using products of simple experts—a potential architecture for embedded systems. Neural Netw. 14, 1257–1264.
Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environ. Geol. 39, 1001–1014.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning Representations by Back-propagation Errors. MIT Press, Cambridge, MA, USA.
Sahasrabudhe, M., Namboodiri, A.M., 2013. Learning fingerprint orientation fields using continuous restricted Boltzmann machines. In: Proceedings of the Second IAPR Asian Conference on Pattern Recognition, pp. 351–355.
Sakurada, M., Yairi, T., 2014. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the Second Workshop on Machine Learning for Sensory Data Analysis, pp. 4.
Sinclair, A.J., 1974. Selection of threshold values in geochemical data using probability graphs. J. Geochem. Explor. 3, 129–149.
Sinclair, A.J., 1991. A fundamental approach to threshold estimation in exploration geochemistry: probability plots revisited. J. Geochem. Explor. 41, 1–22.
Sun, J., Steinecker, A., Glocker, P., 2014. Application of deep belief networks for precision mechanism quality inspection, Precision Assembly Technologies and Systems. Springer, pp. 87–93.
Tukey, J.W., 1997. Exploratory Data Analysis. Addison-Wesley Publishing Complany, Massachusetts, Reading.
Valentine, A.P., Trampert, J., 2012. Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data. Geophys. J. Int. 189, 1183–1202.
Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications. Springer-Verlag, Berlin, Heidelberg, p. 381.
Wang, H., Zuo, R., 2015. A comparative study of trend surface analysis and spectrum-area multifractal model to identify geochemical anomalies. J. Geochem. Explor. 155, 84–90.
Wang, H., Cheng, Q., Zuo, R., 2015a. Quantifying the spatial characteristics of geochemical patterns via GIS-based geographically weighted statistics. J. Geochem. Explor. 157, 110–119.
Wang, H., Cheng, Q., Zuo, R., 2015b. Spatial characteristics of geochemical patterns related to Fe mineralization in the southwestern Fujian province (China). J. Geochem. Explor. 148, 259–269.
Wang, J., Zuo, R., 2015c. A MATLAB-based program for processing geochemical data using fractal/multifractal modeling. Earth Sci. Inform. . http://dx.doi.org/10.1007/s12145-015-0215-5
Xie, X., Mu, X., Ren, T., 1997. Geochemical mapping in China. J. Geochem. Explor. 60, 99–113.
Yang, Z., Zhang, D., Feng, C., She, H., Li, J., 2008. SHRIMP zircon U–Pb dating of quartz porphyry from Zhongjia tin–polymetallic deposit in Longyan area, Fujian province, and its geological significance. Miner. Depos. 27, 329–335 (in Chinese with English abstract).
Yousefi, M., Kamkar-Rouhani, A., Carranza, E.J.M., 2012. Geochemical mineralization probability index (GMPI): a new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping. J. Geochem. Explor. 115, 24–35.
Yuan, Y., Feng, H., Zhang, D., Di, Y., Wang, C., Ni, J., 2013. Geochronology of Dapai iron–polymetallic deposit in Yongding city, Fujian province and its geological significance. Acta Miner. Sin. (Suppl.), S73–S75 (in Chinese with English abstract).
Zhang, D., Wu, G., Di, Y., Wang, C., Yao, J., Zhang, Y., Lv, L., Yuan, Y., Shi, J., 2012. Geochronology of diagenesis and mineralization of the Luoyang iron deposit in Zhangping city, Fujian province and its geological significance. Earth Sci. J. China Univ. Geosci. 37, 1217–1231 (in Chinese with English abstract).
Zhang, Z., Zuo, R., 2014. Sr–Nd–Pb isotope systematics of magnetite: implications for the genesis of Makeng Fe deposit, southern China. Ore Geol. Rev. 57, 53–60.
Zhang, Z., Zuo, R., Cheng, Q., 2015a. The mineralization age of the Makeng Fe deposit, South China: implications from U–Pb and Sm–Nd geochronology. Int. J. Earth Sci. 104, 663–682.
Zhang, Z., Zuo, R., Cheng, Q., 2015b. Geological features and formation processes of the Makeng Fe deposit, China. Resour. Geol. 65, 266–284.
Zhang, Z., Zuo, R., Xiong, Y., 2015c. A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. Sci. China Earth Sci. (in press)
Zuo, R., 2011. Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal modeling in the Gangdese Belt, Tibet (China). J. Geochem. Explor. 111, 13–22.
Zuo, R., 2014. Identification of geochemical anomalies associated with mineralization in the Fanshan district, Fujian, China. J. Geochem. Explor. 139, 170–176.
Zuo, R., Carranza, E.J.M., Cheng, Q., 2012. Fractal/multifractalmodelling of geochemical exploration data. J. Geochem. Explor. 122, 1–3.
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, 225–235.
Zuo, R., Wang, J., 2015. Fractal/multifractal modeling of geochemical data: a review. J. Geochem. Explor. http://dx.doi.org/10.1016/j.gexplo.2015.04.010
Zuo, R., Wang, J., Chen, G., Yang, M., 2015a. Identification of weak anomalies: a multifractal perspective. J. Geochem. Explor. 148, 12–24.
Zuo, R., Xia, Q., Zhang, D., 2013. A comparison study of the C–A and S–A models with singularity analysis to identify geochemical anomalies in covered areas. Appl. Geochem. 33, 165–172.
Zuo, R., Zhang, Z., Zhang, D., Carranza, E.J.M., Wang, H., 2015b. Evaluation of uncertainty in mineral prospectivity mapping due to missing evidence: a case study with skarn-type Fe deposits in Southwestern Fujian Province, China. Ore Geol. Rev. 71, 502–515.