

# Big Data Analytics of Identifying Geochemical Anomalies Supported by Machine Learning Methods

Renguang Zuo<sup>1,2</sup> and Yihui Xiong<sup>1</sup>

Received 2 August 2017; accepted 12 October 2017

Big data analytics brings a novel way for identifying geochemical anomalies in mineral exploration because it involves processing of the whole geochemical dataset to reveal statistical correlations between geochemical patterns and known mineralization. Traditional methods of processing exploration geochemical data mainly involve the identification of positive geochemical anomalies related to mineralization, but ignore negative geochemical anomalies. Therefore, the identified geochemical anomalies do not completely reflect the sought geochemical signature of mineralization, leading to uncertainty in geochemical prospecting. In this study, data for 39 geochemical variables from a regional stream sediment geochemical survey of southwest Fujian Province of China were subjected to big data analytics for identifying geochemical anomalies related to skarn-type Fe polymetallic mineralization through deep autoencoder network. The receiver operating characteristic (ROC) and areas under curve (AUC) were applied to evaluate the performance of big data analytics. The AUC of the anomaly map obtained using all the geochemical variables is larger than the AUC of the anomaly map obtained using only five selected elements known to be associated with the mineralization (i.e., Fe<sub>2</sub>O<sub>3</sub>, Cu, Pb, Zn, Mn). This indicates that big data analytics, with the support of machine learning methods, is a powerful tool for identifying multivariate geochemical anomalies related to mineralization.

**KEY WORDS:** Big data analytics, Machine learning, Mineral exploration, Geochemical anomalies, GIS, Fujian Province.

## INTRODUCTION

Humanity has entered the era of big data as large volumes of complex data are increasingly generated by various sources (Gantz and Reinsel 2011; Manyika et al. 2011; Wang et al. 2016). There are different definitions of big data, varying from the product-oriented, the process-oriented, the cognition-oriented, and to the social-oriented perspective (Ekbia et al. 2015). The ideas of big data include

(Mayer-Schonberger and Cukier 2013): (1) that they are based on data science and let the data speak for themselves; (2) that they do not take into account causal relations, but statistical correlations among all the available data; and (3) that they are based on full samples rather than partial samples. Big data imply not only a large volume of data, but also other features that differentiate them from the concepts of “massive data” and “very large data” (Hu et al. 2014).

The initial applications of big data technologies and methodologies mainly focused on Internet and business intelligence (Fan et al. 2014; Guo et al. 2014), but applications to solving scientific and engineering problems have been increasing (Boyd

<sup>1</sup>State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China.

<sup>2</sup>To whom correspondence should be addressed; e-mail: zrguang@cug.edu.cn

and Crawford 2012; M. Chen, et al. 2014). Accordingly, there are increasing numbers of publications related to big data in various journals. Two special issues on big data have been published in *Nature* (Howe et al. 2008) and *Science* (Finlayson 2011). Big data analytics can bring new ideas for geological studies (e.g., Zhang et al. 2017). Zhai (2017) reported that big data analytics of granite geochemical data could lead to spectacular results. For example, Wang et al. (2017) found that the components of N-mid-ocean ridge basalts (MORB) highly varied and they pointed out that the discrimination diagrams for the basalt tectonic environment should be further investigated by analysis of MORB big data.

Machine learning (ML) methods provide effective tools for big data analytics and can reveal underlying patterns and make predictions based on big data (M. Chen, et al. 2014). The focus of ML is on building a computer system to analyze and learn from data for pattern recognition and prediction in various fields (Carbonell et al. 1983; Kotsiantis et al. 2007; Jordan and Mitchell 2015). A framework of ML for big data analytics proposed by Zhou et al. (2017) consists of ML, big data, user, domain, and system. ML plays a vital role in such framework and interacts with the other four components. The challenges of ML for big data analytics include (Zhou et al. 2017): (1) improving the efficiency of iterations; (2) minimizing the feedback/communication from/with classifiers; (3) improving the speed and various aspects of big data in ML; (4) solving the increased problems of complexity; and (5) other issues such as the complexity of setting objective functions.

Various ML methods have been applied to recognize geochemical anomalies related to mineralization (e.g., Twarakavi et al. 2006; Y. Chen, et al. 2014; O'Brien et al. 2015; Gonbadi et al. 2015; Kirkwood et al. 2016; Zhao et al. 2016; Xiong and Zuo 2016; Chen and Wu 2017). For instance, support vector machine has been applied for mapping arsenic concentration using data of gold concentration in sediments (Twarakavi et al. 2006) and random forest has been used to identify gahnite compositions as guides for exploration of Pb–Zn–Ag deposits in Broken Hill of Australia (O'Brien et al. 2015). The advantages of ML for processing of exploration geochemical data rely on its ability to model the complex and unknown multivariate geochemical distribution and extract meaningful elemental associations related to mineralization or environmental pollution. While some ML methods

are black-box techniques such that the inner structures of multielements are invariably unknown to geoscientists, thereby leading to lack of information for proper geological interpretation (Y. Chen, et al. 2014; Xiong and Zuo 2016; Zuo 2017).

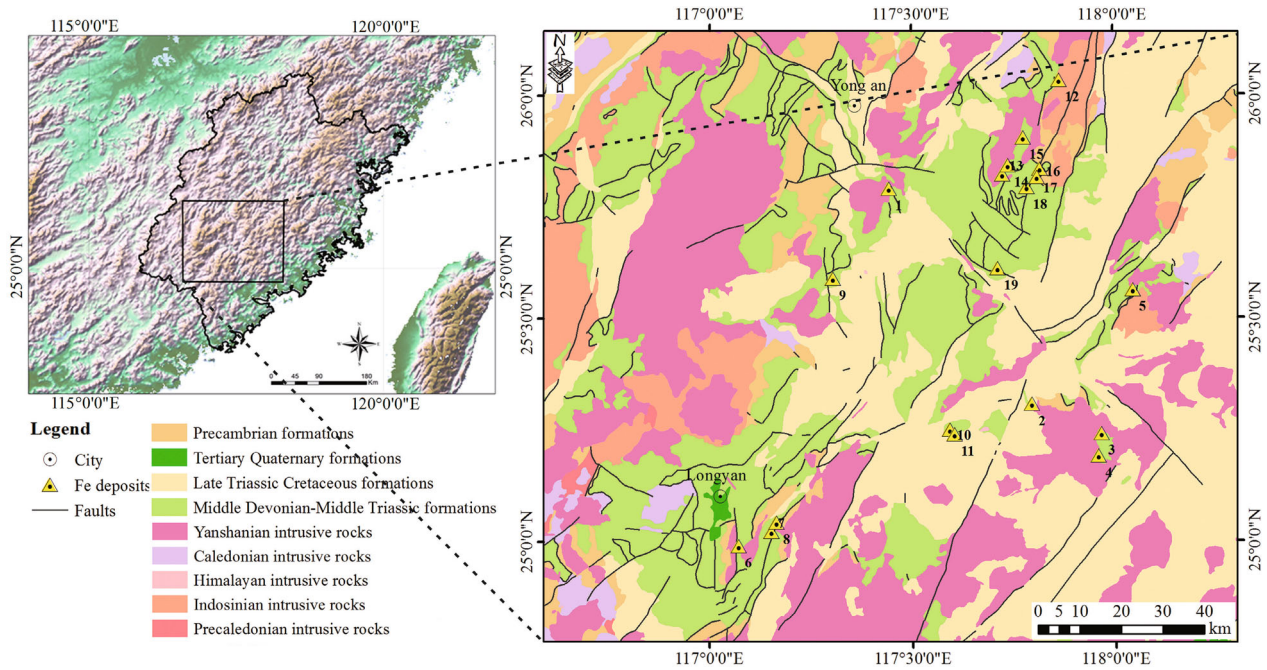
The concept of big data does not only refer to large volume and complex structure of data, but also represents a novel idea for data processing. In this study, data for all geochemical variables that provide positive/negative geochemical anomalies and the statistical correlations between geochemical anomalies and known mineralization are considered to identify geochemical signatures of Fe polymetallic mineralization in southwest Fujian Province of China. These aspects represent two of the three ideas of big data mentioned above and are markedly different from the traditional methods for processing exploration geochemical data. Therefore, the main aim of this paper is to demonstrate the applicability and robustness of big data analytics for extraction of significant anomalies from exploration geochemical data supported by ML methods.

## STUDY AREA AND DATA

The study area, located in the southwest part of Fujian Province, is one of the important iron polymetallic belts in China. Several Fe polymetallic deposits have been discovered in this region, such as the Makeng, Luoyang, Zhongjia, and Pantian Fe deposits (Fig. 1). Previous geological and geochemical studies and field observations revealed that these Fe polymetallic deposits are skarn-type mineralization (Ge et al. 1981; Zhang and Zuo 2014, 2015; Z. Wang, et al. 2015; Zhang et al. 2015a, b, 2016; Zuo et al. 2015b; Zuo 2016). Detailed information on the geological setting and mineral deposit model of skarn-type Fe polymetallic deposits in the study area can be found in Zhang et al. (2015a, b) and Zuo et al. (2015b).

The datasets used in this study consist of a regional geological map, locations of known Fe polymetallic deposits, and regional stream sediment geochemical data. The geological map comprised lithologic formations, intrusions, and faults mapped at 1:200,000 scale. The mineral deposits data include 19 skarn-type Fe polymetallic deposits associated with Yanshanian intrusions. The geochemical data consist of concentrations of 39 major and trace elements compiled from the national geochemical mapping of China at 1:200,000 scale.

## Big Data Analytics of Identifying Geochemical Anomalies



**Figure 1.** Simplified geological map of southwest Fujian Province in China (modified from Fujian Institute of Geological Survey 2011).

Geochemical data points used in this study were equally distributed at a 2-km spatial resolution. The concentrations of Bi, Cd, Co, Cu, La, Mo, Nb, Pb, Th, U, and W were determined by inductively coupled plasma mass spectrometry. The concentrations of Al, Cr, Fe, K, P, Si, Ti, Y, and Zr were determined by X-ray fluorescence. The concentrations of Ba, Be, Ca, Li, Mg, Mn, Na, Ni, Sr, V, and Zn were measured by inductively coupled plasma atomic emission spectrometry. The concentrations of Ag, B, and Sn were measured by emission spectrometry. The concentrations of As and Sb were measured by hydride generation atomic fluorescence spectrometry. The concentrations of Au, Hg, and F were measured by graphite furnace atomic absorption spectrometry, cold vapor atomic fluorescence spectrometry, and ion selective electrode, respectively (Xie et al. 2008). The sampling, analysis, detection limit, and quality control of geochemical data are detailed in Xie et al. (1997) and Wang et al. (2011).

The stream sediment geochemical data used in this study have been investigated to exploit the spatial relationships between geochemical patterns and known skarn-type Fe polymetallic deposits and identify geochemical anomalies related to mineralization (H. Wang and R. Zuo 2015; Wang et al. 2015a, b; Zuo et al. 2015b; Zhang et al. 2016; Xiong

and Zuo 2016, 2017). In these existing studies, only five elements including  $\text{Fe}_2\text{O}_3$ , Cu, Pb, Zn, and Mn were processed and other elements were not involved.

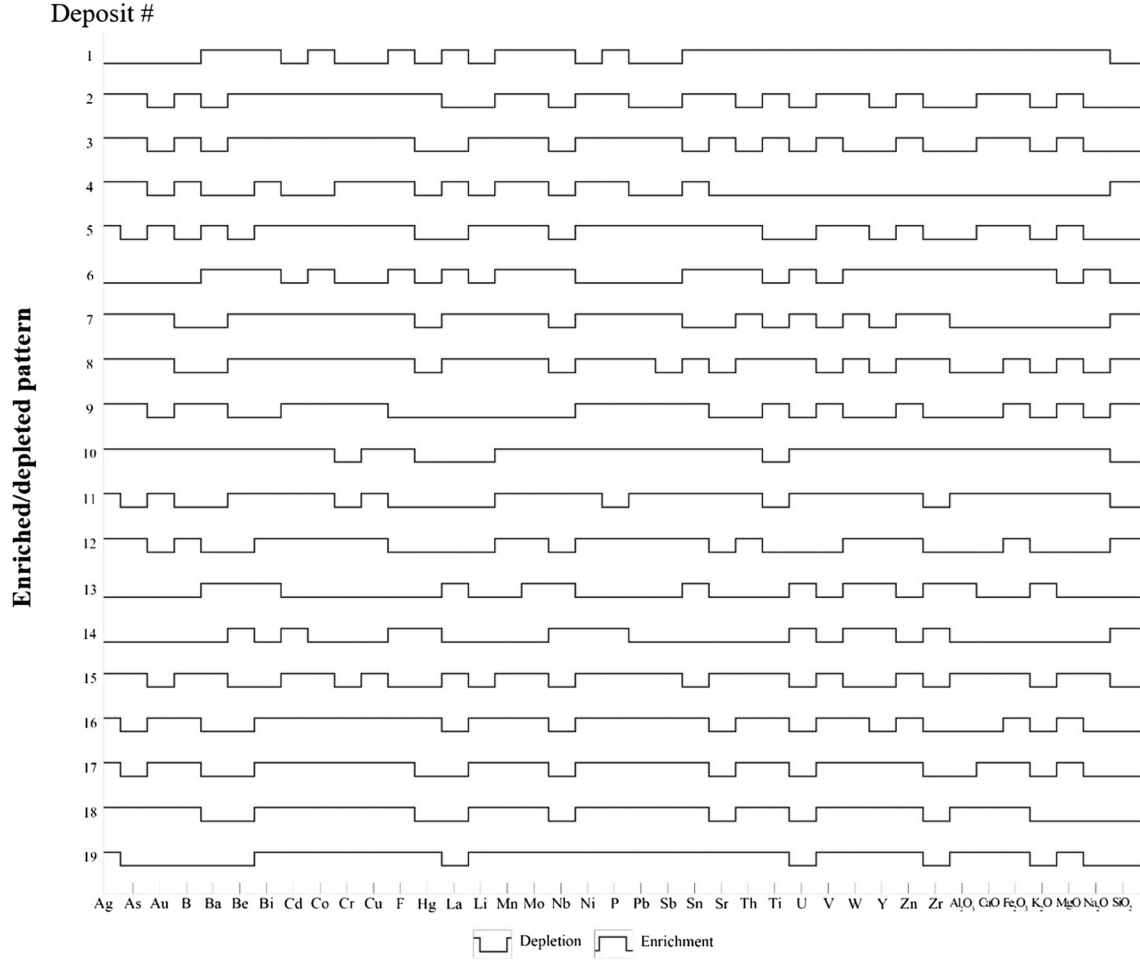
## METHODS

### Local Singularity Analysis

The local singularity analysis (LSA) proposed by Cheng (2007) was demonstrated to be a powerful tool for characterizing local enrichment/depletion patterns of geochemical elements (Cheng 2007, 2012; Zuo and Cheng 2008; Zuo et al. 2009, 2013, 2015a, 2016; Zuo and Wang 2016). A power-law model is typically used to calculate the desired exponent within a local window,

$$\mu(A) = cA^{a/2}, \quad (1)$$

where  $\mu(A)$  denotes the concentration of an element within a window of area  $A$ ,  $c$  is a constant, and  $a$  represents the local singularity index, which can be estimated using the ratio of the logarithmic transformations of the measure  $\mu$  to the area  $A$  as follows:



**Figure 2.** Enrichment and depletion characteristics of the studied 39 elements for each of the known skarn-type Fe polymetallic deposits based on the singularity index. The number in the Y-axis denotes the known skarn-type Fe polymetallic deposits shown in Figure 1.

$$a = \log(\mu_1/\mu_2) / \log[(A_1/A_2)^{1/2}].$$

Here,  $a > 2$  and  $a < 2$  represent local depletion and enrichment, respectively;  $a = 2$  represents a normal distribution.

### Deep Autoencoder Network

The deep autoencoder network (DAN), which was developed based on the deep belief network (Hinton et al. 2006; Hinton and Salakhutdinov 2006), can be used to detect anomalies based on the reconstruction error that can be computed as:

$$E = \sqrt{\sum_{i=1}^n (O_i - I_i)^2},$$

where  $n$  represents the dimensions of the data and  $O$  is the reconstructed data of the input data  $I$ . The principle of DAN for detecting anomalies is that small probability samples like geochemical anomalies related to mineralization contribute little to the autoencoder network and will be linked to higher reconstruction errors because they are poorly encoded and reconstructed during the training of a DAN. Detailed information on DAN and its construction and calculation process can be found in Hinton and Salakhutdinov (2006) and Xiong and Zuo (2016).



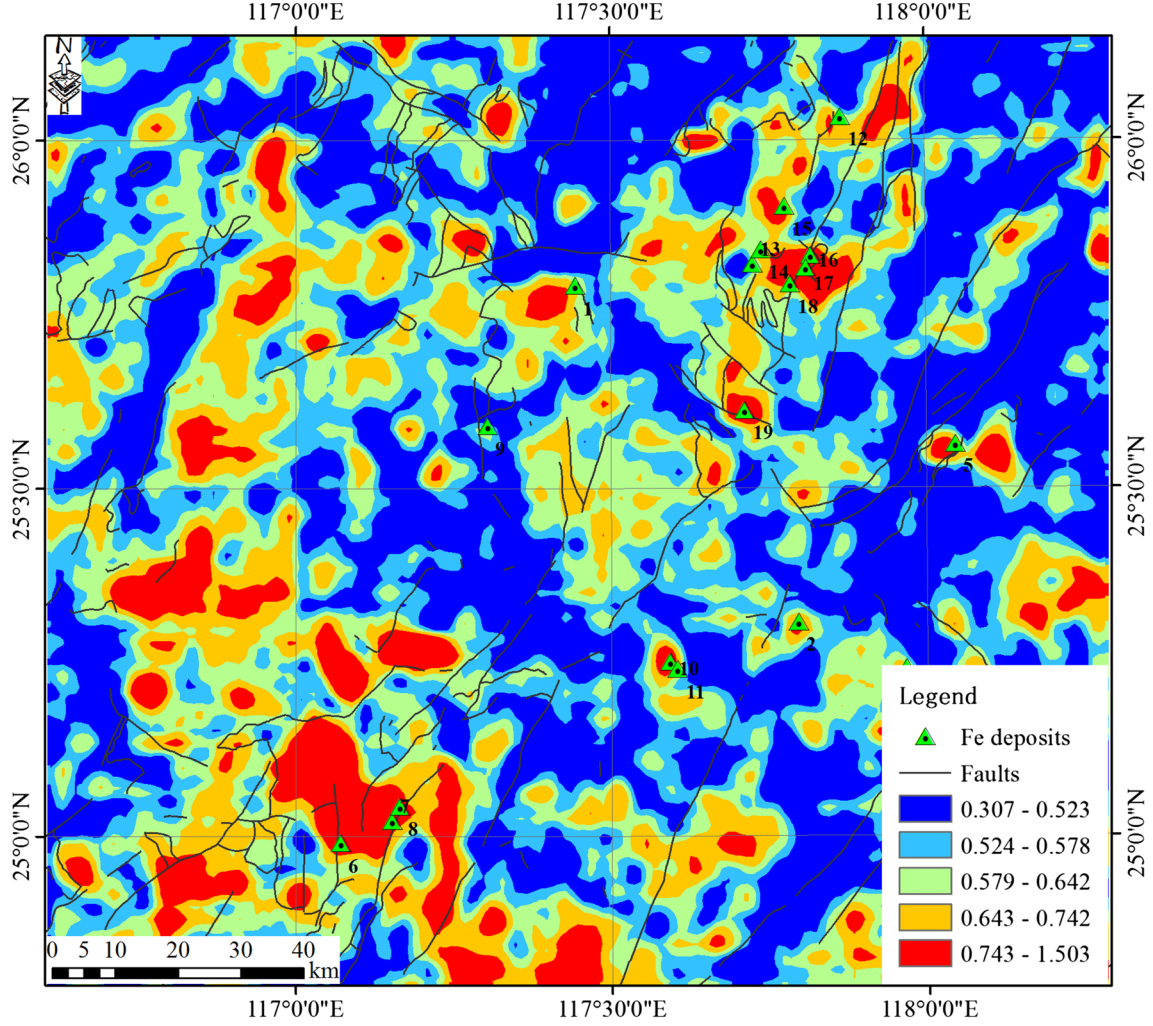


Figure 3. Geochemical anomaly map based on all the 39 elements obtained by the deep autoencoder network.

### Receiver Operating Characteristic

The receiver operating characteristic (ROC) and areas under curve (AUC) (Fawcett 2006) were applied to assess the performance of DAN. The ROC has become a popular method for evaluating the performance of tools for mapping mineral prospectivity and geochemical anomalies (e.g., Nykänen et al. 2017; Chen and Wu 2016, 2017; Parsa et al. 2017; Xiong et al. 2017). In this study, 19 randomly selected points were used as non-deposit for ROC analysis.

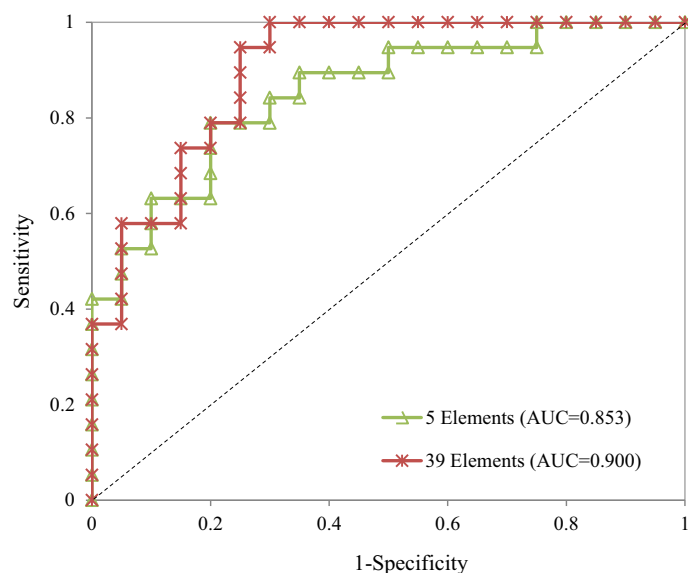
For analysis of AUC,  $x_i$  ( $i = 1, 2, \dots, p$ ) represents the predictive value of the  $i$ th true positive

sample and  $y_j$  ( $j = 1, 2, \dots, n$ ) represents the predictive value of the  $j$ th true negative sample. The AUC can be calculated as (Bergmann et al. 2000):

$$AUC = \frac{1}{p \times n} \sum_{i=1}^p \sum_{j=1}^n \phi(x_i, y_j), \quad (4)$$

$$\text{where } \phi(x_i, y_j) = \begin{cases} 1, & x_i > y_j \\ 0.5, & x_i = y_j \\ 0 & x_i < y_j \end{cases} \quad (5)$$

The range of AUC is from 0 to 1. An  $AUC > 0.5$  indicates good performance for a specific predictive model.



**Figure 4.** Comparison of ROC of the anomaly map based on all the 39 elements and the anomaly map based on the selected five elements.

## RESULTS AND DISCUSSION

### Geochemical Anomaly Patterns Related to Known Fe Polymetallic Mineralization

All the data were preprocessed using LSA with the support of MATLAB-based software developed and made freely available by H. Wang and R. Zuo (2015) for processing exploration geochemical data by means of fractal/multifractal modeling. The enrichment and depletion characteristics of the studied 39 elements for each of the known skarn-type Fe polymetallic deposits are shown in Figure 2. It can be seen that (1) the geochemical anomaly pattern around each of the known mineral deposits is different from each other, and (2) most of the known Fe deposits are enriched in As, B, Be, Bi, Ti, Y, Cao, Co, Cr, Sb, MgO, Mn, Mo, Ni, Pb, P, Th, F, Ag, Cd, Cu, Sn, W, Fe<sub>2</sub>O<sub>3</sub>, Zn, and V, but depleted in Hg, K<sub>2</sub>O, SiO<sub>2</sub>, Au, Ba, La, Nb, Zr, Li, U, Al<sub>2</sub>O<sub>3</sub>, and Sr (Fig. 2). These suggest that geochemical anomaly patterns associated with skarn-type Fe polymetallic mineralization in the study area are complex. Therefore, the typical elemental association Fe<sub>2</sub>O<sub>3</sub>–Cu–Pb–Zn–Mn considered for such mineralization (e.g., Zuo et al. 2015b; Wang et al. 2015a, b; Xiong and Zuo 2016, 2017) does not completely represent the sought geochemical signature of mineralization, leading to uncertainty in geochemical prospecting.

The typical elemental association of Fe<sub>2</sub>O<sub>3</sub>–Cu–Pb–Zn–Mn considered by previous geological studies and field observations was based on reasoning of geological causal relationship between known mineralization and geochemical signatures. However, the formation of magmatic-hydrothermal mineralization is a complex process and involves a number of elements participating in water-rock interaction, in addition to weathering and denudation, and the masking effects of covers, thereby leading to complex geochemical anomaly patterns in sediments. The typical elemental association of Fe<sub>2</sub>O<sub>3</sub>–Cu–Pb–Zn–Mn only refers to five elements related to skarn-type Fe polymetallic mineralization in the study area, but does not portray other elements such as the main rock-forming elements, which are of significance for indicating the geological environment of mineralization. In addition, the elemental association is only relevant to positive anomalies, but ignores negative geochemical anomalies, which are rarely used in the search for ore deposits (Rose et al. 1979). Increasing attentions have been paid to negative geochemical anomalies (e.g., Ma et al. 2013; Liu et al. 2016), which can offer critical information related to basic geology, geochemical anomaly interpretation, and the establishment of models (Shi and Wang 1995). For instance, Liu et al. (2016) reported that the spatial distribution of the negative Na<sub>2</sub>O anomaly could represent the fluid-rock interaction zones and can be used to guide mineral

exploration. Therefore, it is important to consider all the elements included in the data to properly explore the statistical correlations between the whole data and all instances of mineralization.

### Identifying Geochemical Anomalies by Big Data Analytics

The isometric logratio (*ilr*) transformation (Egozcue et al. 2003) was applied to preprocess the raw geochemical data to address the closure problem with compositional data. For a convenient comparison, the DAN was used to analyze the spatial correlations between all 39 geochemical variables and 19 known mineral deposits. The DAN was previously applied for identifying geochemical anomalies of skarn-type Fe polymetallic mineralization in the same study area by Xiong and Zuo (2016), only considering  $\text{Fe}_2\text{O}_3$ , Cu, Pb, Zn, and Mn. A four-layer autoencoder network at the encoder phase was created. The number of hidden units for each layer is 128–64–32–16, and the number of iteration over the training dataset is 200. The reconstruction error of each geochemical sample was calculated by DAN, and then these reconstruction errors were interpolated by the inverse distance weighting method. The generated geochemical anomaly map exhibits a strong spatial association with the known mineral deposits (Fig. 3).

To demonstrate the performance of big data analytics for identifying geochemical anomalies, the anomaly map obtained using the elemental association of  $\text{Fe}_2\text{O}_3$ –Cu–Pb–Zn–Mn (Xiong and Zuo 2016) was used as a reference. It can be observed from Figure 4 that the AUC of the anomaly map based on all geochemical variables (Fig. 3) and  $\text{Fe}_2\text{O}_3$ –Cu–Pb–Zn–Mn (cf. Fig. 8 in Xiong and Zuo 2016) are 0.900 and 0.853, respectively. The larger AUC of the anomaly map derived by big data analytics suggests that it is a more powerful tool for proper extraction of significant multivariate geochemical anomalies.

### CONCLUSIONS

In this study, a geochemical anomaly map related to skarn-type Fe polymetallic mineralization was obtained through big data analytics of regional geochemical stream sediment data. Analysis of geochemical enrichment and depletion characteris-

tics of known mineral deposits reveals the complexity of geochemical anomaly patterns in sediments around mineralized areas. The AUC of the anomaly map obtained in this study indicates a better performance of big data analytics compared to that of the traditional method. These results suggest that big data analytics, which consider all available data of geochemical variables and here supported by ML, is a robust method for identification of significant multivariate geochemical signatures associated with mineralization. Big data analytics of geochemical anomalies in complete multielement compositions is a new research direction in geochemical prospecting. However, data redundancy and computational complexity together impose problems for big data analytics and should be studied in future.

### ACKNOWLEDGMENTS

We thank John Carranza (Editor-in-Chief) for his edits and comments and three reviewers for their comments and suggestions, which helped to improve this study. This research benefited from the joint financial support from the National Key Research and Development Program of China (2016YFC0600508), the National Natural Science Foundation of China (41772344 and 41522206), the Natural Science Foundation of Hubei Province (2017CFA053), and the MOST Special Fund from the State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences (MSFGPMR03-3).

### REFERENCES

- Bergmann, R., Ludbrook, J., & Spooren, W. P. J. M. (2000). Different outcomes of the Wilcoxon–Mann–Whitney test from different statistics packages. *The American Statistician*, 54, 72–77.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15, 662–679.
- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An overview of machine learning. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 3–23). Berlin: Springer.
- Chen, Y., Lu, L., & Li, X. (2014). Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *Journal of Geochemical Exploration*, 140, 56–63.

- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19, 171–209.
- Chen, Y., & Wu, W. (2016). A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve analysis. *Ore Geology Reviews*, 74, 26–38.
- Chen, Y., & Wu, W. (2017). Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochemistry: Exploration, Environment, Analysis*, 17, 231–238.
- 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. (2012). Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. *Journal of Geochemical Exploration*, 122, 55–70.
- Egozcue, J. J., Pawłowsky-Glahn, V., Mateu-Figueras, G., & BarceloVidal, C. (2003). Isometric logratiotrans formations for compositional data analysis. *Mathematic Geology*, 35, 279–300.
- Ekbja, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., et al. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66, 1523–1545.
- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1, 293–314.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874.
- Finlayson, A. E. T. (2011). Dealing with data: Fostering fidelity. *Science*, 331, 1515.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC Iview*, 1142, 1–12.
- Ge, C., Han, F., Zhou, T., & Chen, D. (1981). Geological characteristics of the Mafikeng iron deposit of marine volcano-sedimentary origin. *Acta Geologica Sinica*, 3, 47–69. **(in Chinese with English abstract)**.
- Gonbadi, A. B., Tabatabaei, S. H., & Carranza, E. J. M. (2015). Supervised geochemical anomaly detection by pattern recognition. *Journal of Geochemical Exploration*, 157, 81–91.
- Guo, H., Wang, L., Chen, F., & Liang, D. (2014). Scientific big data and digital earth. *Chinese Science Bulletin*, 59, 5066–5073.
- Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, 504–507.
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., et al. (2008). Big data: The future of biocuration. *Nature*, 455, 47–50.
- Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2, 652–687.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255–260.
- Kirkwood, C., Cave, M., Beamish, D., Grebby, S., & Ferreira, A. (2016). A machine learning approach to geochemical mapping. *Journal of Geochemical Exploration*, 167, 49–61.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31, 249–268.
- Liu, Y., Ma, S., Zhu, L., Sadeghi, M., Doherty, A. L., Cao, D., et al. (2016). The multi-attribute anomaly structure model: An exploration tool for the Zhaojiku epithermal Pb–Zn deposit, China. *Journal of Geochemical Exploration*, 169, 50–59.
- Ma, S., Zhu, L., & Liu, C. (2013). Anomaly models of spatial structures for copper–molybdenum ore deposits and their application. *Acta Geologica Sinica (English Edition)*, 3, 843–857.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. New York: Houghton Mifflin Harcourt Publishing Company.
- Nykänen, V., Niiranen, T., Molnár, F., Lahti, I., Korhonen, K., Cook, N., et al. (2017). Optimizing a knowledge-driven prospectivity model for gold deposits within the Kallakumpu Belt, Northern Finland. *Natural Resources Research*, 26, 571–584.
- O'Brien, J. J., Spry, P. G., Nettleton, D., Xu, R., & Teale, G. S. (2015). Using random forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb–Zn–Ag deposits in the Broken Hill domain, Australia. *Journal of Geochemical Exploration*, 49, 74–86.
- Parsa, M., Maghsoudi, A., & Yousefi, M. (2017). A receiver operating characteristics-based geochemical data fusion technique for targeting undiscovered mineral deposits. *Natural Resources Research*. doi:10.1007/s11053-017-9351-6.
- Rose, A. W., Hawkes, H. E., & Webb, J. S. (1979). *Geochemistry in mineral exploration* (p. 657). New York, NY: Academic Press.
- Shi, C., & Wang, C. (1995). Regional geochemical secondary negative anomalies and their significance. *Journal of Geochemical Exploration*, 55, 11–23.
- Twarakavi, N. K. C., Misra, D., & Bandopadhyay, S. (2006). Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data. *Natural Resources Research*, 15, 15–26.
- Wang, J., Chen, W., Zhang, Q., Jin, W., Jiao, S., Wang, Y., et al. (2017). MORB data mining: Reflection of basalt discrimination diagram. *Geotectonica et Metallogenia*, 41, 420–431. **(in Chinese with English abstract)**.
- Wang, H., Cheng, Q., & Zuo, R. (2015a). Quantifying the spatial characteristics of geochemical patterns via GIS-based geographically weighted statistics. *Journal of Geochemical Exploration*, 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). *Journal of Geochemical Exploration*, 148, 259–269.
- Wang, X., Xie, X., Zhang, B., & Hou, Q. (2011). Geochemical probe into China's continental crust. *Acta Geoscientia Sinica*, 32, 65–83. **(in Chinese with English abstract)**.
- Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of big data. *Information Sciences*, 367, 747–765.
- Wang, H., & Zuo, R. (2015). A comparative study of trend surface analysis and spectrum-area multifractal model to identify geochemical anomalies. *Journal of Geochemical Exploration*, 155, 84–90.
- Wang, J., & Zuo, R. (2015). A MATLAB-based program for processing geochemical data using fractal/multifractal modeling. *Earth Science Informatics*, 8, 937–947.
- Wang, Z., Zuo, R., & Zhang, Z. (2015). Spatial analysis of Fe deposits in Fujian Province, China: Implications for mineral exploration. *Journal of Earth Science*, 26, 813–820.
- Xie, X., Mu, X., & Ren, T. (1997). Geochemical mapping in China. *Journal of Geochemical Exploration*, 60, 99–113.
- Xie, X., Wang, X., Zhang, Q., Zhou, G., Cheng, H., Liu, D., et al. (2008). Multi-scale geochemical mapping in China. *Geochemistry: Exploration Environment, Analysis*, 8, 333–341.



## Big Data Analytics of Identifying Geochemical Anomalies

- Xiong, Y., & Zuo, R. (2016). Recognition of geochemical anomalies using a deep autoencoder network. *Computers & Geosciences*, 86, 75–82.
- Xiong, Y., & Zuo, R. (2017). Effects of misclassification costs on mapping mineral prospectivity. *Ore Geology Reviews*, 82, 1–9.
- Xiong, Y., Zuo, R., Wang, K., & Wang, J. (2017). Identification of geochemical anomalies via local RX anomaly detector. *Journal of Geochemical Exploration*. doi:10.1016/j.gexplo.2017.06.021.
- Zhai, M. (2017). Granites: Leading study issue for continental evolution. *Acta Petrologica Sinica*, 33, 1369–1380. **(in Chinese with English abstract)**.
- Zhang, Q., Jiang, S., Li, C., & Chen, W. (2017). Granite and continental tectonics, magma thermal field and metallogenesis. *Acta Petrologica Sinica*, 33, 1524–1540. **(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 Geology Reviews*, 57, 53–60.
- Zhang, Z., & Zuo, R. (2015). Tectonic evolution of southwestern Fujian Province and spatial-temporal distribution regularity of mineral deposits. *Acta Petrologica Sinica*, 31, 217–229. **(in Chinese with English abstract)**.
- 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. *International Journal of Earth Sciences*, 104, 663–682.
- Zhang, Z., Zuo, R., & Cheng, Q. (2015b). Geological features and formation processes of the Makeng Fe deposit, China. *Resource Geology*, 65, 266–284.
- Zhang, Z., Zuo, R., & Xiong, Y. (2016). A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, 59, 556–572.
- Zhao, J., Chen, S., & Zuo, R. (2016). Identifying geochemical anomalies associated with Au–Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China. *Journal of Geochemical Exploration*, 164, 54–64.
- Zhou, L., Pan, S., Wang, J., & Vasilakos, V. A. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350–361.
- Zuo, R. (2016). A nonlinear controlling function of geological features on magmatic-hydrothermal mineralization. *Scientific Reports*, 6, 27127.
- Zuo, R. (2017). Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural Resources Research*, 26, 457–464.
- Zuo, R., Carranza, E. J. M., & Wang, J. (2016). Spatial analysis and visualization of exploration geochemical data. *Earth-Science Reviews*, 158, 9–18.
- Zuo, R., & Cheng, Q. (2008). Mapping singularities: A technique to identify potential Cu mineral deposits using sediment geochemical data, an example for Tibet, west China. *Mineralogical Magazine*, 72, 531–534.
- 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, 225–235.
- Zuo, R., & Wang, J. (2016). Fractal/multifractal modeling of geochemical data: A review. *Journal of Geochemical Exploration*, 164, 33–41.
- Zuo, R., Wang, J., Chen, G., & Yang, M. (2015a). Identification of weak anomalies: A multifractal perspective. *Journal of Geochemical Exploration*, 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. *Applied Geochemistry*, 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 Geology Reviews*, 71, 502–515.