Mineral Prospectivity Mapping Integrating Multi-source Geology Spatial Data Sets and Logistic Regression Modelling

Cuihua Chen[#1], Hongzhang Dai[#2], Yue Liu[#3], Binbin He[*4]

[#]*College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China*

[1]chencuihua@cdut.edu.cn [2]651366362@qq.com

[3]*452672784@qq.com*

* *Institute of Geo-Spatial Information Science and Technology University of Electronic Science and Technology of China, Chengdu 611731, China*

[4]binbinhe@uestc.edu.cn

***Abstract*— A method of integrating multi-source geology spatial data sets and logistic regression modelling for mineral prospectivity mapping is described in this paper. Logistic regression model describes the relationship between a dependent variable, which is a binary variable representing the presence or absence of the mineral deposits, and *k* independent variables represent ore-controlled geological features such as faults, lithology, geochemical anomaly, which may be continuous or discrete or any combination of both types. A case study was selected located in East Kunlun region of Qinghai Province, China. Multi-source geospatial data contain geological data, geophysical data, geochemical data and remotely sensed data. The potential prospectivity map was produced by logistic regression on the resulting revised binary map patterns, same as in weights of evidence modelling, two logistic probability thresholds of 0.52 and 0.58 were used to classify the study area into three classes of high potential, moderate potential, and low potential areas, which high potential areas contain 54% of the total gold deposits, covering 7.3% of the total area. Moderate potential area contains 8% of the gold deposits, covering 8% of the total area. Low potential areas contain 38% of the gold deposits, covering 84.7% of the total area..**

***Keywords*— Mineral prospectivity mapping; logistic regression modelling; spatial analysis; GIS; Eastern Kunlun**

I. INTRODUCTION

Geographic Information System (GIS) provides the geologist with a powerful tool for mineral exploration, when used in concert with statistical and geostatistical analysis, for querying, archiving, manipulating, analysing and visualizing multi-sources geoscience data. Advances in the use of GIS software and analysis can be conjunction with traditional geoscience data sets to determine effective predictors for mineral resources assessment, from which mineral prospectivity maps can be generated to delineate potential exploration targets on a regional scale[1]-[3].

In recent years, quantitative analysis of geoscientific data to determine areas most likely to contain mineral deposits is becoming increasingly common in the mining industry, and thus, various spatial analysis techniques have been applied to evaluation of mineral resources to predict the degrees of favorability for deposit and barren. Many spatial modelling techniques have been applied in the area of mineral resource evaluation, which mainly include logistic regression model [4]-[6], weights-of-evidence model [7]-[8], fuzzy logic model [9] [10], and so on.

In this study, the logistic regression modelling as spatial statistical tools for iron mineralization potential evaluation was used in the East Kunlun Region of Qinghai Province, China. All the data sets used in this paper are derived from established multi-source geology spatial database, which mainly contains geological, geophysical, geochemical and remote sensing data. The information was captured at a scale of 1:500,000 using a Gauss-Kruger Projection for all the evidential layers. The data were used as geo-referenced raster files resampled at 1,000 m and were processed with ArcView Spatial Analyst ESRI software and a plug-in extension for multi-criteria approaches, Arc-SDM [11]. Logistic regression analysis was performed in the Statistical Package of Social Sciences (SPSS) aided by GIS techniques.

II. LOGISTIC REGRESSION MODEL

Logistic regression model describes the relationship between a dependent variable, which is a binary variable representing the presence or absence of the mineral deposits, and k independent variables represent ore-controlled geological features such as faults, lithology, geochemical anomaly, which may be continuous or discrete or any combination of both types. Parameters were estimated by likelihood maximization. The relationship between the occurrence and its dependency on several geological variables can be expressed as:

$$p = \frac{1}{1+e^{z}} \tag{1}$$

where p is the probability of an event (mineral deposit) occurring. The probability varies from 0 to 1 on an S-shaped curve and z is the linear combination, and its value varies from $-\infty$ to $+\infty$.

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

978-1-4244-8351-8/11/$26.00 ©2011 IEEE

Where b_0 is the intercept of the model, n is the number of independent variables, $bi(i=1,2,3,\ldots,n)$ is the slope coefficient of the model, and $x_i(i=1,2,3,\ldots n)$ is the independent variable (binary variable or multivariable). The linear model formed is a logistic regression representing the presence or absence of dependent (deposits) on the independent variables (ore-controlled geological features).

III. EXPERIMENTS

A case study was selected located in East Kunlun region of Qinghai Province, China. Multi-source geospatial data contain geological data, geophysical data, geochemical data and remotely sensed data. First, the weights-of-evidence method was applied to test the spatial association between known iron deposits/occurrences and various geological features (e.g., regional faults, intrusive rocks, geophysical anomalies) to determine the optimal binary geological variables derived from multi-class geological variables. And then use logistic regression modelling to examine the importance of the selected variables based on logistic parameters. Table 1 shows the regression coefficients and the significance level of the independent variables. Seven map patterns were selected for mapping iron mineral potential by logistic regression modelling. It is noted that three map patterns of aeromagnetic anomaly, hydroxyl alteration, and calcilization are missing in parts of the study area. In weights of evidence modelling, the weight calculations can deal with missing data by mask analysis. However, this procedure would lead to significant loss of information in logistic regression modelling, which can eliminate those areas where the data missing in predictive map. For this reason, the three map patterns were modified so that, in regions where the data are missing, they were treated as being "not present". Logistic regression modelling does not require that the evidence layers be conditionally independent with respect to the training points.

The potential prospectivity map was produced by logistic regression on the resulting revised binary map patterns, same as in weights of evidence modelling, two logistic probability thresholds of 0.52 and 0.58 were used to classify the study area into three classes of high potential, moderate potential, and low potential areas, which were shown in Fig.2. As viewed on Figures 1A and 1B, high potential areas contain 54% of the total gold deposits, covering 7.3% of the total area. Moderate potential area contains 8% of the gold deposits, covering 8% of the total area. Low potential areas contain 38% of the gold deposits, covering 84.7% of the total area.

IV. CONCLUSIONS

Both models, weights of evidence modelling and logistic regression modelling are applied to identify potential favourable areas in the East Kunlun region. Weights of evidence modelling have been used widely arises from its weights can be interpreted simple and understand easily, as well, the optimal thresholds derived from multi-class geological variables can be acquired objectively by proximity analysis in a GIS system. However, in most commonly, conditional independence assumption is a trouble problem for weights of evidence modelling, because of the existed correlation among the selected geological variables, sometimes, which lead some important ore-controlling variables not to be combined. In this paper, eight binary evidential maps integrated for iron metallogenic potential assessment reached a better prediction rate though conditional independence assumption was slightly violated. The study area was divided into three classes based on posterior probability thresholds, among which high potential areas occupied 11% of the total area and contained 56% of the known deposits (Fig.3 and Fig.4).

Unlike weights of evidence modelling, logistic regression modelling is based multivariable statistical analysis, which not requires the assumption of conditional independence. Logistic coefficients can be used to examine the importance of geological variables and the spatial correlation between the geological variables and known mineral deposits. In logistic regression modelling, bouguer anomaly was excepted from map patterns because its coefficient had no significant, the logistic probability map also classified into three target zones same as the posterior probability map based on seven binary evidential maps, and high potential areas accounted of 7.3% of the total area, containing 54% of the total deposits. From the comparison of both modelling, it seems that the results of logistic regression modelling are superior to the results of weights of evidence modelling, which may be that the former not requires the conditional independence assumption.

ACKNOWLEDGMENT

This study was supported by the National High-Tech R&D Program of China (Grant No. 2007AA12Z227), the National Natural Science Foundation of China (Grant No. 40701146) and Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20095122120020). We express our gratitude to Zhuang Yongcheng from the Qinghai Institute of Geology Survey for his directions.

REFERENCES

[1] G. L. Raines, "Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the Western United States", *Natural Resources Research*, vol.8, No.4, pp.257–276, 1999.

[2] E. J. M. Carranza and M. Hale, "Geologically constrained probabilistic mapping of gold potential, Baguio district, Philippines," *Natural Resources Research*, vol.9, No.3, pp.237–253, 2000.

[3] J. R. Harris, L. Wilkinson and E. C. Grunsky, "Effective use and interpretation of litho-geochemical data in regional mineral exploration programs: application of Geographic Information Systems (GIS) technology", *Ore Geology Reviews*, vol.16, No.3-4, pp.107–143, 2000.

[4] F. P. Agterberg, G. F. Bonham-Carter, Q. Cheng and D. F. Wright, "Weights of evidence modeling and weighted logistic regression in mineral potential mapping," in Computers in Geology, Davis, J. C. and H. C. Herzfeld, (Eds.), New York: Oxford University Press, 13–32, 1993.

[5] E. J. M. Carranza and M. Hale, "Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines," *Exploration and Mining Geology*, vol.10, No.3, pp.165–175, 2001.

[6] N. R. Sahoo and H. S. Pandala, "Integration of sparse geologic information in gold targeting using logistic regression analysis in the Hutti-Maski Schist Belt, Raichur, Karnataka, India—a case study," *Natural Resources Research*, vol.8, No.3, pp.233-250, 1999.

[7] G. F. Bonham-Carter, F. P. Agterberg and D. F. Wright, “Weights of evidence modeling: a new approach to mapping mineral potential”, in Statistical Applications in the Earth Sciences, Agterberg, F. P. and G. F. Bonham-Carter (Eds.), Ottawa, Canada: Geological Survey Canada 1989, Paper 89–90, 171–183.

[8] F. P. Agterberg, G. F. Bonham-Carter and D. F. Wright, “Statistical pattern integration for mineral exploration,” In Computer Applications in Resource Estimation Prediction and Assessment for Metals and Petroleum, Gaál, G. and D. F. Merriam (Eds.), Pergamon Press, Oxford-New York, 21pp, 1990.

[9] S. Choi and W.M. Moon, “Fuzzy logic fusion of W-Mo exploration data from Seobyeog-ri,” *Geosciences Journal*, vol.4, No.2, pp.43-52, 2000.

[10] X. Luo and R. Dimitrakopoulos, “Data-driven fuzzy analysis in quantitative mineral resource assessment,” *Computers & Geosciences*, vol.29, No.1, pp.3–13, 2003.

[11] L. D. Kemp, G. F. Bonham-Carter and G. L. Raines, Arc-WofE: ArcView extension for weights of evidence mapping. http: //gis.nrcan.gc.ca/software/arcview/ wofe., 1999.

TABLE Ⅰ. LOGISTIC COEFFICIENTS, WALD’S STATISTICS, AND SIGNIFICANCE LEVEL OF INDEPENDENT VARIABLES

Variable	b	Wald’s statistics	Significance level	Odds ratio
Indosinian granite	2.882	4.618	0.106	17.85
Tanjianshan Group	1.857	13.393	0.221	6.4
Dagangou Formation	2.108	13.446	0.109	8.23
Hydroxyl alteration	2.091	10.722	0.071	8.09
Regional faults	2.224	5.171	0.379	9.24
calicilization	1.502	5.768	0.016	4.49
Bouguer	-0.446	0.589	0.443	0.644
Aeromagnetic	1.234	4.227	0.041	3.43
Constant	-1.688	4.278	0.039	0.18

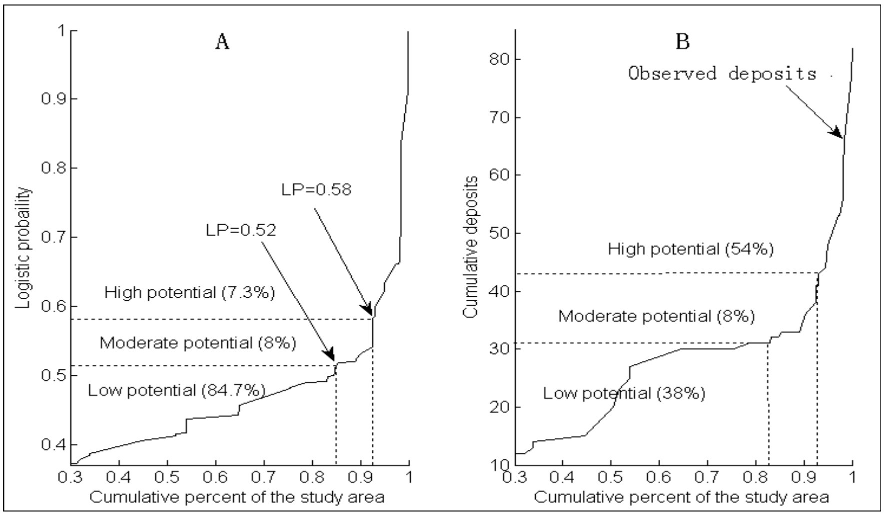

Fig. 1A. Variation of logistic probability with cumulative percent of study area based on combining seven binary evidential maps.

Fig. 1B. Variation of cumulative iron deposits with cumulative percent of study area based on seven evidential maps.

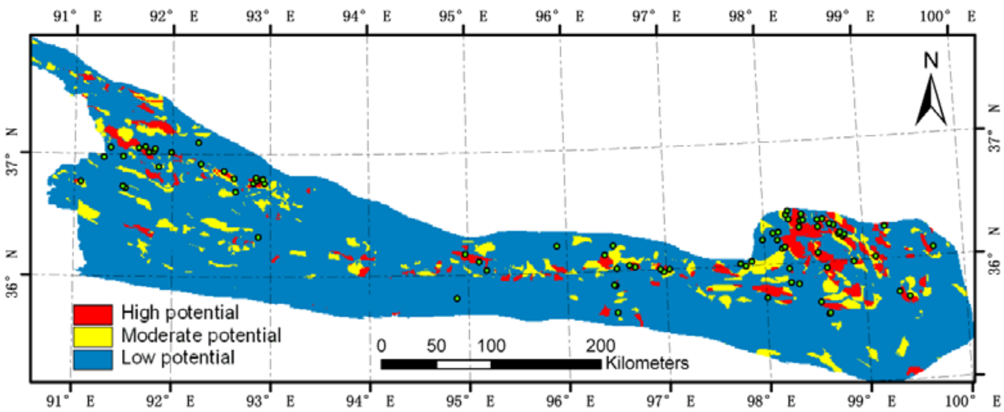

Fig. 2. Mineral potential map generated by reclassification of logistic probability based on eight binary evidential, green points are iron deposits.

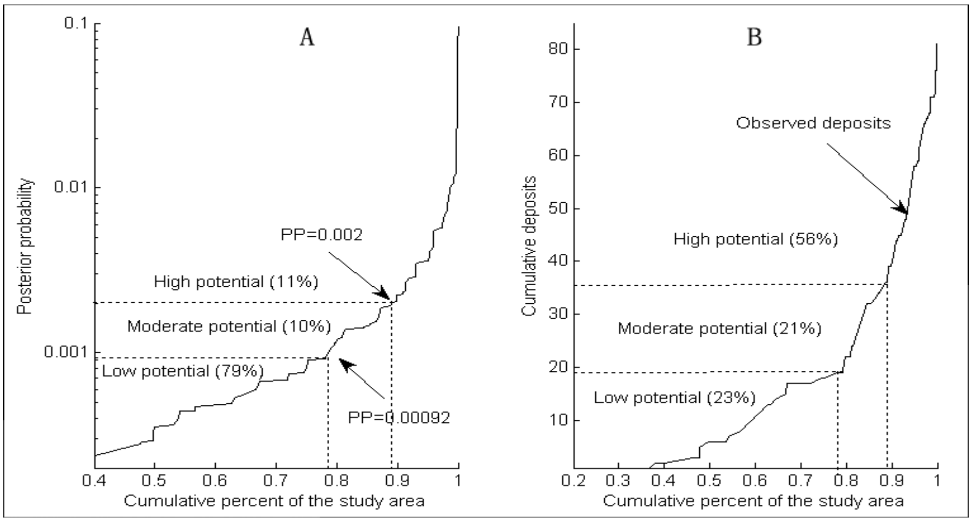

Fig. 3A. Variation of posterior probability with cumulative percent of study area based on combining eight binary evidential maps.

Fig. 3B. Variation of cumulative iron deposits with cumulative percent of study area based on eight evidential maps.

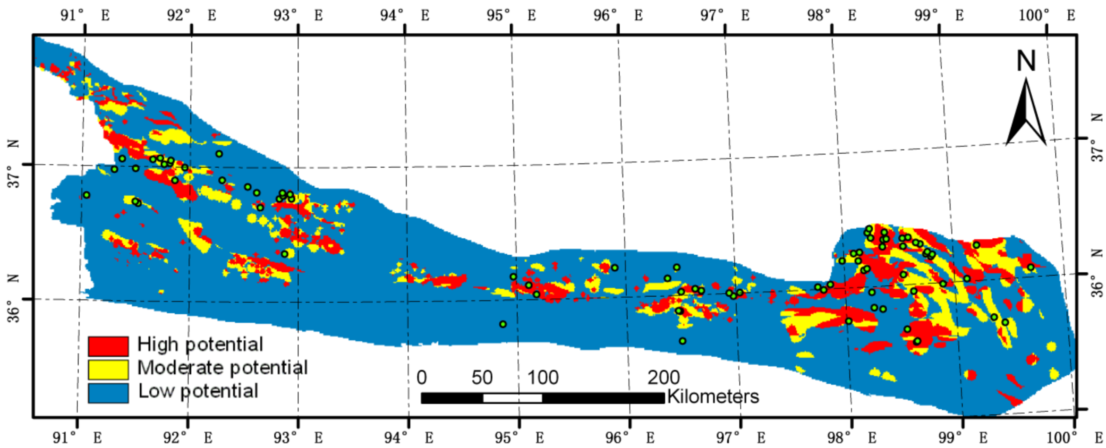

Fig. 4. Mineral potential map generated by reclassification of posterior probability based on eight binary evidential, green points are iron deposits.