

加权 Logistic 回归模型在斑岩铜矿预测中的应用

——以中—哈边境扎尔玛—萨吾尔成矿带为例

努丽曼古·阿不都克力木¹⁾, 张晓帆¹⁾, 陈川¹⁾, 徐仕琪²⁾, 赵同阳²⁾

1) 新疆大学新疆中亚造山带大陆动力学与成矿预测重点实验室, 乌鲁木齐, 830047;

2) 新疆地质调查院, 乌鲁木齐, 830000

内容提要: 加权 Logistic 回归是基于 GIS 成矿预测的主要方法之一, 其模型是不同于线性模型的一种类型。它具有强大的空间分析功能、适用性强、不受任何独立条件的约束、预测结果更可靠, 因此在矿产资源评价研究中得到了很多地质学家的青睐。以矿床模型和成矿理论为基础, 加权 Logistic 回归分析模型在成矿预测中的应用主要包括三部分: 加权 Logistic 回归模型的建立及其应用、成矿有利度综合评价、成矿远景区圈定。本文以中国—哈萨克斯坦边境地区扎尔玛—萨吾尔成矿带斑岩型铜矿为例, 探讨了基于 GIS 的加权 Logistic 回归模型在成矿预测中的应用。

关键词: 扎尔玛—萨吾尔; GIS; 加权 Logistic 回归; 成矿预测

近几十年来, 人们提出了各种基于 GIS 的快速而有效的矿产资源潜力评价方法, 这些方法具有传统矿产资源评价方法无法比拟的优点 (Agterberg, 2005; 丁清峰等, 2005)。其中, 较为引人注目的是 Agterberg 等人 1992 年提出的加权 Logistic 回归法 (WLR) (Raines, 2002)。这种方法曾被 Reddy 应用于马尼托巴雪湖地区的火山成因块状硫化物矿床 (Carranza, 2001)。WLR 实质上是根据采用的不规则多边形单元数量加权的逻辑回归, 可以被用于解释非独立条件的变量。此外, 在证据权模型基础上加入 WLR 方法替代贝叶斯法则, 解除贝叶斯法则中的条件独立假设对地学应用的种种限制, 避免条件独立假设在预测结果中所造成的偏差, 而且在计算过程中不受因变量的影响 (Agterberg, 1993; Singer, 1999)。因此在成矿预测研究中得到了很多地质学家的青睐。笔者应用 WLR 法对中国—哈萨克斯坦边境地区扎尔玛—萨吾尔成矿带的斑岩铜矿进行预测研究, 快速生成矿化远景概率图, 圈定成矿远景区。

1 研究区简介

扎尔玛—萨吾尔成矿带位于中—哈边境, 构造

上处在西伯利亚板块和哈萨克斯坦—准噶尔板块汇聚带, 是中亚造山带的重要组成部分, 是研究陆壳增生机制的热点地区之一。该区域广泛发育晚古生代的火山—沉积岩并伴有丰富的多金属矿产。因此, 该地区不仅在全球的地质构造演化研究中占有重要地位, 而且也是我国重要的大型固体矿产资源基地之一。

研究区自西向东主要包括: 西南卡尔巴地区、斋桑盆地、扎尔玛地区和萨吾尔套山地区, 处于准噶尔褶皱系的西北缘, 跨东、西准噶尔两个褶皱带。区内侵入岩较为发育, 岩石类型由超基性—酸性均有, 其中以酸性为主, 而超基性、基性岩石极为少见, 研究区地质与铜矿分布见图 1。与邻国及邻区的对比分析表明, 该区与中亚晚古生代斑岩成矿带具有一些相似性, 此外, 南马克苏特—喀拉通克铜镍(中)基性成矿带也通过该区北部, 仍显示有一定的铜矿找矿前景。本文在 ArcGIS 平台上, 应用 WLR 法对各地质类因素进行成矿有利度评价分析, 圈定成矿远景区, 目的是为该区的斑岩铜矿勘查提供依据, 同时也为成矿规律的研究提供新的信息和方法。

2 加权 Logistic 回归 (WLR) 模型简介

WLR 是圈定成矿预测区的主要快捷方法之一,

注: 本文为国家“十二·五”科技支撑计划“新疆重要成矿带战略性矿产资源预测与靶区评价”(编号 2011BAB06B08)、国家重点基础研究发展计划项目“中亚造山带大陆动力学与成矿作用”(编号 2007CB411308)的成果。

收稿日期: 2011-08-03; 改回日期: 2011-12-01; 责任编辑: 章雨旭。

作者简介: 努丽曼古·阿不都克力木, 女, 1986 年生。硕士研究生, 地球探测与信息技术专业; Email: nuri315@126.com。通讯作者: 陈川, 男, 1972 年生。副教授, 硕士研究生导师。主要从事区域成矿预测与 3S 技术研究; Email: chenchuan@sina.cn。

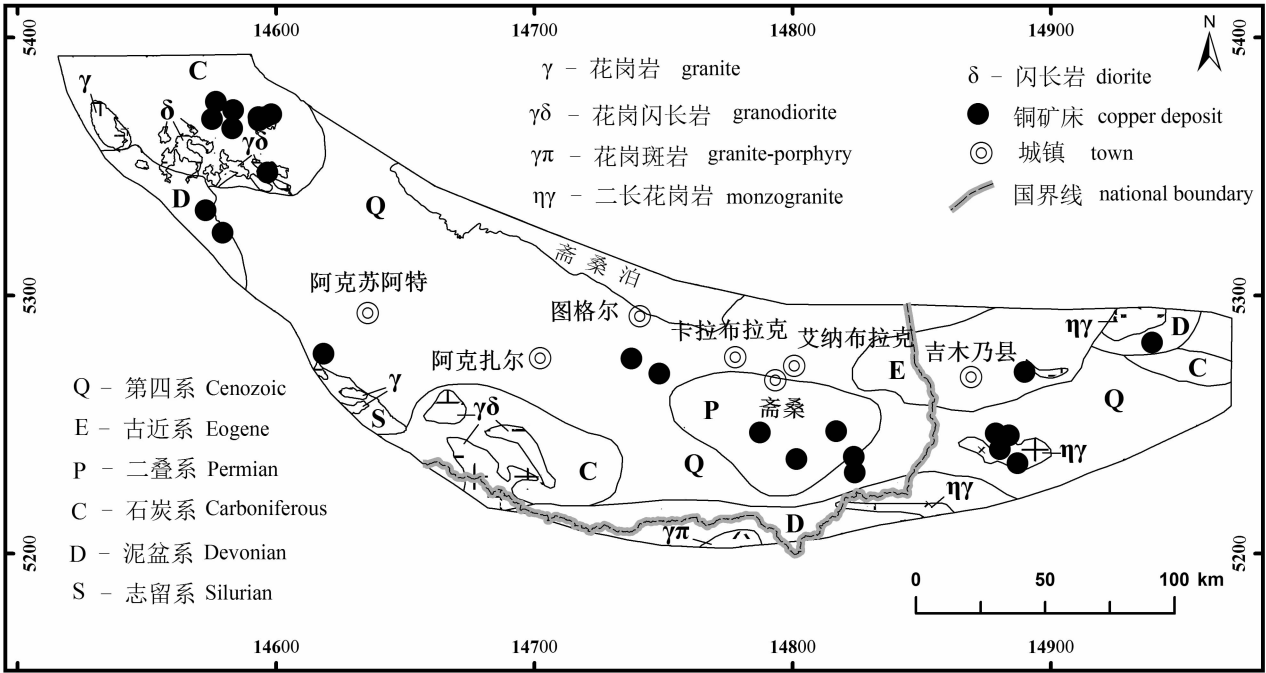


图 1 扎尔玛—萨吾尔成矿带地质矿产略图
Fig. 1 Schematic geology and mineral resources map of the Zharma—Sawur Metallogenic Belt

实质就是进行合成或综合一些数据,建立 WLR 模型,进行成矿有利度评价,主要是在一区域通过对影响某类矿产形成的一个因变量和多个自变量之间形成多元回归关系,从而根据这种多元回归关系预测任何该区域内这类矿产存在的概率(刘艺梁等, 2010;党如童, 2011)。将一研究区划分成大小相等的网格单元,并假设该区域范围内的各类影响成矿的因子也可以做同样的分割,即把单元内部看作近似均匀,而单元之间在地质构造、地层岩性等方面是变化的。这样,每一个单元就对应着多个影响因子中的具体类别(用分类代码表示)。现假定有 P 个这样的影响因子(各类影响因素数据)参与成矿有利度评价;另外还有一个响应因子(为已知矿点),它包含两类数据:① 已发生矿点的个数;② 已发生矿(化)点的二值栅格数据,栅格数据中各像元的值用两个值表现,即发现矿(化)点(=1),或未发现矿(化)点(=0)。此时,研究区中任意像元 i 中发现矿(化)点的概率 θ_i 可以用 P 个影响因子的函数表示。但应用该参数估计方法估计参数,再计算成矿概率时,往往会超出 0~1 的范围,这和概率的定义是矛盾的。为解决这一矛盾,同时克服证据权法中的条件独立问题,Cox 和 Snell 提出了如下回归方程:

$$\theta_i = \frac{e^{X'_i b}}{1 + e^{X'_i b}} \tag{1}$$

$$\text{logit}(\theta_i) = \ln\left(\frac{\theta_i}{1 - \theta_i}\right) = X'_i b \tag{2}$$

用最大似然率法估计第 i 像元回归参数 b 的值,进一步得到该像元成矿概率 θ_i ,最终就可获得研究区成矿有利度评价结果(Carranza, 2001)。

3 加权 Logistic 回归模型应用

3.1 数据组织

数据组织主要包括进行加权逻辑回归分析中所运用的各类数据组合。本研究中将已经建立的扎尔玛—萨吾尔成矿带空间数据库作为研究区成矿预测的数据基础。

3.2 成矿有利因素的提取

利用空间统计方法进行成矿潜力定量评价的本质是度量影响因子(断层、岩石组合、化探异常、物探异常等)和响应因子矿化点之间的空间关系,最后给出所有影响因子综合作用结果,从而得到成矿有利地区评价结果。进行合成或综合时往往采用 WLR 模型进行成矿有利度评价。

3.3 影响因子的确定

在 WLR 中,证据层的选择依然采用证据权模型的方法,即用 W^+ 、 W^- 以及 C 值定量评价单一影响

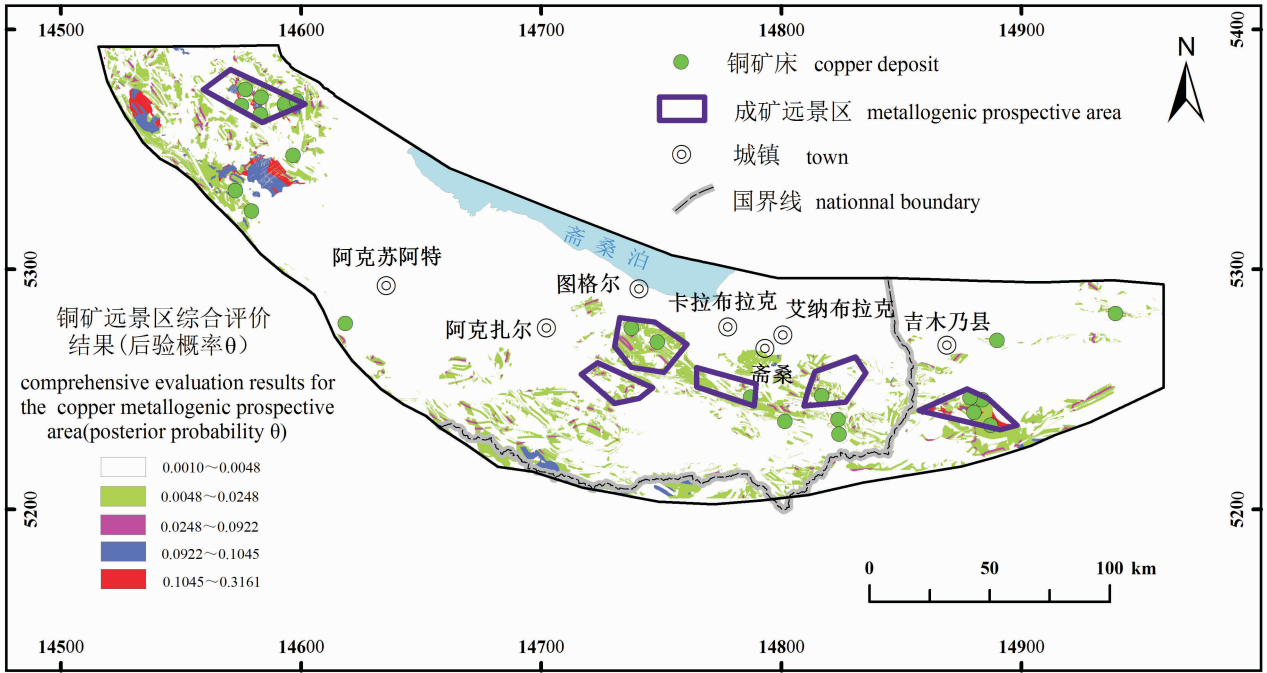


图 2 扎尔玛—萨吾尔成矿带铜矿综合评价成矿远景区图

Fig. 2 comprehensive evaluation results for the copper metallogenetic prospective area in the Zharma—Sawur Metallogenic Belt

因子对成矿的影响;各证据层的正权重 (W^+) 和负权重 (W^-) 按以下公式计算:

$$W^+ = \ln \left[\frac{P(B|D)}{P(\bar{B}|\bar{D})} \right] \tag{3}$$

$$W^- = \ln \left[\frac{P(\bar{B}|D)}{P(B|\bar{D})} \right] \tag{4}$$

第 i 图层与矿点同时存在定义为正权,有矿点而无第 i 图层存在定义为负权。反差 ($Contrast$),以 C 表示, $C = W^+ - W^-$ 。上述权值和反差的计算结果,还需对其显著性水平进行定量评价,评价公式如下:

$$Studentized(Contrast) = \frac{S^2(Contrast)}{\sqrt{S^2(W^+) + S^2(W^-)}} \tag{5}$$

上式中, $S^2(W^+)$ 、 $S^2(W^-)$ 、 $S^2(Contrast)$ 表示 W^+ 、 W^- 、 $Contrast$ 的方差, $Studentized(Contrast)$ 是对反差 $Contrast$ (即 C 值) 做学生化处理 (杨茂森等, 2005; 孙艳霞等, 2010; 李以科等, 2009)。根据成矿模型与有利区提取模型的研究成果,选择可能影响成矿的因素,建立证据权分析过程中影响因素图层,即证据图层。

在扎尔玛—萨吾尔构造带地

区,以斑岩型铜矿为主攻目标进行成矿预测。下面通过采用上述模型中 W^+ 、 W^- 、 C 值以及 C 值的学生化公式,计算扎尔玛—萨吾尔地区各影响因素与斑岩型铜矿床 (包括矿 (化) 点) 的空间关系部分结果如表 1。

表 1 中—哈边境扎尔玛—萨吾尔地区斑岩型铜矿床与各影响因素关系定量评价表

Table 1 Evaluation parameters for the porphyry copper deposit of the Zharma—Sawur Metallogenic Belt, China—Kazakhstan border area

影响因素	正权值 (W^+)	负权值 (W^-)	学生化反差 ($S(C)$)	综合权重
断层距离	0.7341	-0.4188	2.3031	0.734100
侵入岩	5.1828	-0.2863	3.7890	5.182800
沉积岩时代地层	5.2330	-0.4037	4.4741	5.233000

表 2 单因素因子对中—哈边境扎尔玛—萨吾尔地区斑岩型铜矿成矿有利度影响程度分类表

Table 2 Classification of influencing factors on porphyry copper mineralization in the Zharma—Sawur Metallogenic Belt, China—Kazakhstan border area

主要影响因素	侵入岩岩性组合,侵入岩影响距离,火山岩岩石组合
参与评价的次要影响因素	断层密度,断层距离沉积岩地层
不参与评价的次要影响因素	沉积岩岩石组合
非影响因素	断层方向,断层空间分配,变质岩建造

表 3 中—哈边境扎尔玛—萨吾尔地区斑岩型铜矿数据驱动评价图层与矿点权值计算结果
Table 3 Weight distribution calculated for porphyry copper mineralization evaluation in Zharma—Sawur Metallogenic Belt, China—Kazakhstan border area

级别	成矿后验概率 区间(θ)		面积(km ²)	单元 数	矿床 数	正权值 (W ⁺)	负权值 (W ⁻)	反差(C)	学生化 反差[S(C)]	综合权重
1	0.0008	0.0200	28173.00	7043	3	-1.3847	1.1740	-2.5587	-3.9920	-1.384700
2	0.0200	0.0437	6161.24	1540	3	0.1369	-0.0291	0.1660	0.2589	0.136881
3	0.0437	0.0787	3052.59	763	4	1.1302	-0.2035	1.3336	2.3050	1.130200
4	0.0787	0.3931	284.02	71	6	3.9934	-0.4631	4.4564	8.3896	3.993400

依据影响因素对矿床或矿(化)点影响程度定量判定原则,根据各影响因素单因素计算结果,各种影响因素按对矿(化)点的影响程度分类如表 2。

4 成矿有利度综合评价

随着专家系统技术的发展,在 GIS 矿产资源评价系统中引入专家系统,利用高水平专家的知识建立地质计算机专家系统。专家系统的引入将会智能化地、更快地进行矿产资源评价及各种专题图的制作(孙宝生等,2002)。WLR 模型对基于数据驱动模型评价结果和基于专家驱动模型评价结果进行综合,获得最终定量评价结果(后验概率 θ)见表 3。

将成矿后验概率图层按后验概率进行重分类得到连续性栅格数据,表中后验概率最大的区间,即表中第 4 级区间,概率在 0.0787~0.3931,落入该区间 71 个单元内的矿床或矿(化)点数有 6 个,占有所有矿床或矿(化)点数的 37.5%,其学生化反差为 8.39,综合权值达到 3.99。显然,后验概率最大的区间即为矿床产出最有利地区。

5 成矿远景区圈定

在采用 WLR 模型进行计算过程中,仅依据地质条件,即按断层、侵入岩、沉积岩、火山岩、变质岩等要素对研究区铜矿进行了有利度评价,得到了研究区综合评价结果。但在成矿远景区圈定过程中,必须结合一定的物化探数据进行验证,仅依据 WLR 模型得到的综合评价后验概率图是不够的,即使综合评价置信度区间均高于 1.96(即可信度概率大于 96%),所以仍然要借助已知物探、化探等数据等进行验证(本文不详细讨论)。圈定结果如图 2 所示:

本次成矿远景区分类采用成矿有利性分类方案,即根据成矿有利度、成矿概率,结合具体的地质情况对预测区按三类(第Ⅰ类、第Ⅱ类、第Ⅲ类)划分,所圈定的 6 个成矿远景区内第Ⅰ类有 3 处,第Ⅱ类有 1 处,第Ⅲ类有 2 处。

6 结论

在 GIS 的支持下,结合加权 Logistic 回归法对斑岩型铜矿床进行成矿预测,由于避免条件独立假设在预测结果中所造成的偏差,在证据权重模型的基础上加入 WLR 替代贝叶斯法则,解除贝叶斯法则中的条件独立假设对地学应用的种种限制。加权 Logistic 回归法提高了成矿预测的自动化程度、预测效率和水平,简化了预测过程,节约了时间,缩短了研究周期。通过建立数学模型,挖掘了多源数据的潜在信息,大大提高了数据的综合利用程度,提高了矿产预测的智能化程度。但是,在研究区的所有解释模式没有完全已知或者缺失值的假设不存在的情况下,WLR 不能被应用。此外,如果存在多重共线性,标准差的回归系数会增大,使模型不合理(Agterberg, 1993)。

本文以已知矿体与地质要素的空间相关性作为预测依据,在栅格数据环境下(张会琼等,2010),利用加权 Logistic 回归模型,实现了成矿有利度定量评价,从而更加准确、真实地反映了研究区内成矿潜力,对中—哈边境地区扎尔玛—萨吾尔成矿带的铜成矿潜力进行了评价,并结合其它相关知识,为研究区内进一步铜矿勘探提供理论依据、技术支持及数据支撑,以服务于研究区铜矿勘查部署工作。

参 考 文 献 / References

党如童. 2011. 运用逻辑回归模型预测泥沙灾害危险系数. 水土保持应用技术, 1:11~12.
丁清峰, 孙丰月. 2005. 专家证据权重法及其在东昆仑地区的应用. 地质与勘探, 41(4):88~94.
李以科, 孙刚. 2009. 证据权模型在赣东北地区铜、金成矿预测中的应用. 黄金科学技术, 17(6):29~33.
刘艺梁, 殷坤龙, 刘斌. 2010. 逻辑回归和人工神经网络模型在滑坡灾害空间预测中的应用. 水文地质工程地质, 37(5):92~96.
孙宝生, 张晓帆, 陈川, 焦旭东, 杨向荣. 2002. RS 和 GIS 技术在新疆地质和矿产资源评价中的应用. 新疆大学学报, 19(3):333~341.
孙艳霞, 张达, 王长明, 岳铮生, 孙卫志. 2010. 证据权重法在新矿

床类型成矿预测中的应用. 金属矿山, 9;79 ~ 83.

杨茂森, 黎清华, 杨海巍. 2005. GIS 支持下的证据权法在胶东金矿集中区预测中的应用. 地球学报, 26(5):487 ~ 491.

张会琼, 张寿庭, 韦昌山, 王京彬, 王玉往. 2010. 在 ArcGIS 栅格数据环境下的综合信息成矿预测. 矿产勘查, 1(2):165 ~ 172.

Agterberg F P, Bonham - Charter G F, Cheng Q, Wright D F. 1993. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. Computers in Geology, Oxford University Press, 13 ~ 32.

Agterberg F P, Bonham - Charter G F. 2005. Measuring the performance of mineral—potential maps. Natural Resources Research, 14(1):1

~ 17.

Carranza E J M, Hale M. 2001. Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines. Explor. Mining Geol, Canada, 10(3):165 ~ 175.

Raines G L, Mihalasky M J. 2002. A reconnaissance method for delineation of tracts for regional-scale mineral -resource assessment based on geologic-map data. Natural Resources Research, 11(4): 241 ~ 248

Singer D A, Kouda R. 1999. A comparison of the weights of evidence method and probabilistic neural networks, Natural Resources Research, 8(4): 287 ~ 298.

The Application of Weighted Logistic Regression Model in Prediction of Porphyry Copper Deposit

——take Zharma—Sawur metallogenic belt, China—Kazakhstan border area, as an example

Nulimangu ABUDUKELIMU¹⁾, ZHANG Xiaofan¹⁾, CHEN Chuan¹⁾, XU Shiqi ²⁾, ZHAO Tongyang²⁾

1) *Xinjiang Key Laboratory for Geodynamic Processes and Metallogenic Prognosis of the Central Asian Orogenic Belt, Xinjiang University, Urumqi, 830047;*

2) *Xinjiang Geological Survey Academy, Urumqi, 830000*

Abstract: Weighted Logistic Regression is one of the main methods of mineral potential mapping. It is different from linear model. Because of its powerful spatial analysis function, strong adaptability, unconstrained by independent conditions, and more reliable prediction results, Weighted Logistic Regression is widely used by many geologists in mineral resources assessment. Based on the mineral deposit model and theory, Weighted Logistic Regression is consists of three parts: (1) Establishment of weighted logistic regression model for mineral potential mapping; (2) comprehensive evaluation of favorable degrees; (3) mineral potential mapping of study area. By the Weighted Logistic Regression model for mineral potential mapping, Zharma—Sawur Metallogic Belt which across border region of China and Kazakhstan is studied and mineral prospecting area of porphyry copper deposit is mapped. At the end, the availability of Weighted Logistic Regression Model for mineral potential mapping is discussed.

Key words: Zharma—Sawur Metallogic Belt; GIS ; weighted logistic regression model; mineral potential mapping