

利用样本排序方法比较化探异常识别模型的效果

陈志军^{1,2}, 成秋明^{1,2,3}, 陈建国^{1,2}

1. 中国地质大学地质过程与矿产资源国家重点实验室, 湖北武汉 430074

2. 中国地质大学资源学院, 湖北武汉 430074

3. 加拿大约克大学地球空间科学与工程系, 加拿大多伦多 M3J1P3

摘要: 地球化学异常的有效识别是化探找矿成败的关键环节。利用样本排序方法对各种化探异常识别模型的处理效果和优劣性进行了比较。以云南个旧及其周边地区铜元素水系沉积物为例, 应用元素含量、异常衬度、局部奇异性分析三大类方法对数据进行了处理, 对化探异常指示变量的排序值进行了3个方面的对比: (1) 在高背景区和低背景区样品的排序特征; (2) 有矿出现位置处样品的样本排序特征; (3) 累积面积(等效于上侧样本排序值)不同分组所圈定的异常范围和矿床(点)之间的空间相关性。结果表明, 局部奇异性分析方法较按含量高低的全局性方法对弱缓异常识别更为有效, 也相对优于滑动衬值。按奇异性指数基于证据权法圈定的异常远景区更具预测意义, 显著优于用元素含量值所圈定的异常范围。局部奇异性方法原理清晰、方法简便、可操作性强, 在地球化学异常识别中可以用其替代滑动衬值方法。

关键词: 局部奇异性指数; 滑动衬值; 样本排序; 数据探查分析; 证据权方法; 个旧地区。

中图分类号: P628

文章编号: 1000-2383(2009)02-0353-12

收稿日期: 2008-12-26

Comparison of Different Models for Anomaly Recognition of Geochemical Data by Using Sample Ranking Method

CHEN Zhi-jun^{1,2}, CHENG Qiu-ming^{1,2,3}, CHEN Jian-guo^{1,2}

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

2. Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China

3. Department of Earth and Space Science and Engineering, York University, Toronto M3J1P3, Canada

Abstract: The geochemical anomaly recognition is the key to geochemical prospecting. Many new models are brought forward to identify and extract the geochemical weak anomalies from the complex background. How to compare the differences in aspect of the anomaly recognition effect of these different models? The authors advance the sample ranking method to discuss this problem. The gliding anomaly contrast and the local singularity analysis are applied to the Cu element data of the stream sediment samples from Gejiu area, Yunnan Province, China. The ranks for the raw data, contrast value and the local singularity exponents, denoted by $RANK(Raw)$, $RANK(CV)$ and $RANK(\Delta\alpha)$ respectively, are calculated by ordering the samples from the high anomaly to the low anomaly. Three ways are employed to compare the $RANK(CV)$ and $RANK(\Delta\alpha)$ with the $RANK(Raw)$: (1) the characteristics of ranks for samples with the strong background and the weak background; (2) the characteristics of ranks for samples where the Cu deposits occur; and (3) the spatial correlation between the locations of Cu deposits and the cumulative area with the same thresholds of the upper rank values. The results demonstrate that the local singularity analysis is a useful model for the weak geochemical anomaly recognition, whose effect corresponds with the gliding anomaly contrast model or even better. The prospective areas delineated by means of weights of evidence method on the basis of local singularity exponents can provide new information and may be significant for the prediction of the undiscovered mineral deposits, which is significantly superior to the results on the basis of raw concentration data. The local singularity analysis has the

基金项目: 国家自然科学基金项目(Nos. 40638041, 40802081); 国家 863 计划(No. 2006AA06Z115); 教育部创新团队基金(No. IRT0755); 国际数学地球科学学生奖励基金; GPMR 国家重点实验室开放课题资助(GPMR200615); 地质过程与矿产资源国家重点实验室科技部专项经费资助。

作者简介: 陈志军(1978—), 男, 博士, 主要从事数学地质的科研和教学工作。E-mail: zjchenes@gmail.com

advantage of the perspicuous principle, convenience and effective performance, and we can substitute it for the gliding contrast value method for the anomaly recognition of geochemical data.

Key words: local singularity exponent; gliding contrast value; sample rankings; exploratory data analysis (EDA); weights of evidence method; Gejiu area.

地球化学异常的识别与提取一直是地球化学勘探的主要目标(Reimann *et al.*, 2005; Grunsky, 2007). 计算机技术尤其是 GIS 技术在化探界的普及, 给区域化探异常数据处理、解释与制图带来了勃勃生机(Grunsky, 2007). 我国实施以水系沉积物为采样对象的全国区域化探扫面计划近 30 年, 积累了丰富的高质量的区域化探扫面数据, 这些数据为近年来大量新矿床的发现提供了有益线索. 近几年, 同行学者致力于资料的二次开发, 问题的焦点又回到了背景和异常识别上, 特别是弱缓异常的识别问题. 勘查地球化学数据具有极其复杂的不规则性, 蕴含着地球化学场的丰富信息, 如: 空间结构性(各向异性)、尺度不变性等特征(鲍征宇等, 1999; Cheng, 1999a). 在地质情况复杂的区域内, 采用统一的异常下限值圈定异常的不合理性已经成为勘查地球化学家的共识, 并提出了一些确定异常下限的新方法, 主要有异常衬度法、趋势面法、子区中位数衬度滤波法(史长义等, 1999)、空间最优 U 统计量法(Cheng *et al.*, 1996; Cheng, 1999a)等. 近些年来, 一些学者还探索用分形方法确定地球化学背景及异常(Cheng *et al.*, 1994, 2000; Li *et al.*, 2003; 韩东昱等, 2004), 其中局部奇异性分析方法是一种日益受到国际关注的新模型(Cheng, 1999b, 2006a, 2006b; Chen *et al.*, 2007; Cheng and Agterberg, 2008), 它用于指示异常的指标来自模型的幂指数 α , 称为奇异性指数, 记 $\Delta\alpha = E - \alpha$, 这里 E 表示空间维数(对二维地球化学图 $E = 2$), 于是, $\Delta\alpha$ 越大于 0 (即 α 越小于 2), 则指示(正)异常强度越大. α 和 $\Delta\alpha$ 是由局部邻域上的多尺度计算而来, 可用来刻画物质(或能量)的相对富集或亏损(成秋明, 2008). 应用不同的化探数据处理模型, 可以得到不同的异常指示变量. 这些变量可能服从不同的分布, 还可能具有不同的量纲, 如何对比这些异常指示变量对圈定异常效果的优劣?

人们最为熟知的异常下限确定方法是由背景平均值加两(或三)倍标准方差来确定, 从另一角度来看, 这种方法可视为对数据进行了标准化变换, 得到一个无量纲的标准得分(记为 k), k 越大于 0, 则(正)异常强度越大. 异常衬度(记为 CV)是指某一

元素所形成的异常含量平均值与异常所在区域的背景平均值的比值, 它作为异常清晰度的量度而广泛使用. CV 越大于 1, 则(正)异常强度越大. 元素的含量值、 CV 值、 $\Delta\alpha$ 值等异常指示变量, 由于它们的物理意义不同、统计分布特征不同, 难以直接比较. 笔者注意到, 除元素的含量值、 CV 值、 $\Delta\alpha$ 值自身之外, 还有一个与分布特征无关的量, 这就是它们的样本排序值. 样品观测值经某些数学变换, 如: 标准化变换、正规化变换、平方根变换、对数变换后, 可以改变分布的类型、分布的形态, 有助于增强异常, 但是其共同特点之一是这些变换不改变样品的相对排序. 异常衬度法、趋势面法等异常识别模型, 则会改变样品排序. 对同一个采样样品来说, 按不同的异常指示变量进行排序, 它所处的序次可能相对稳定、也可能发生较大变化, 这可在一定程度上体现出不同方法异常识别效果的趋同性或差异性.

当化探取样范围覆盖全区且每个取样数据所代表的空间范围相同时(对非规则采样, 可经空间插值实现), 由样品的有序排列可容易换算为空间范围, 例如排序前 2.5% 的有序样品与研究区 2.5% 的面积是相当的. 对不同异常指示变量进行排序后, 无论是对逐个样品的比较, 还是对某一组有序样品的比较, 它们都可视为在等面积意义下的比较. 在尚未有矿产发现、研究程度较低的地区开展化探异常识别工作, 利用样本排序方法可以简便地确定最有潜力的地区, 如用排序取值前 2.5% 的样品来圈定研究区 2.5% 的范围, 并通过与地质构造背景分析来比较不同方法之间异常识别效果的异同. 在已有较多矿产发现的地区, 还可借助证据权重法模型(Agterberg, 1989; Agterberg *et al.*, 1993; Bonham-Carter, 1994), 通过化探异常所圈定的空间范围与其所覆盖或遗漏矿产两者之间的空间相关关系来评价异常识别效果. 采用样本排序值可克服分类标准不统一的困难, 使不同模型的异常识别结果都转化成为等累积面积下的比较.

因此, 本文拟利用顺序统计量思想, 选取元素含量、异常衬度、奇异性指数作为 3 种代表性的异常指示变量, 先将其转换成对样品的排序序次, 再按排序值来进行异常识别效果的比较. 不同异常指示变量

的样本排序, 在高背景区、低背景区和有矿产出现位置处样品各自有何特点? 用不同的异常识别模型来圈定相等的异常面积(例如占研究区 2.5% 面积), 它们所确定的那些样品各自在空间上是如何分布的? 与矿产分布的空间相关性如何? 本文将分别用排序值的 X - Y plot 图和基于证据权法的学生 t 统计量进行比较分析。 CV 和 $\Delta\alpha$ 圈定异常的更多异同之处(包括计算方法原理、多尺度识别效果等), 限于篇幅另文阐述。

1 样本排序方法

顺序统计量在数学上通常是指样本值由小到大排序而得到的统计量, 它有着广泛的应用, 有些性质不依赖于母体分布并且计算量很小, 使用起来比较方便。本文所指的样本排序方法约定为: 按变量指示异常性能从优到劣进行排序。以样品数为 n 的元素含量 Z 的一组观测值($z_1, z_2, \dots, z_n$) 为例, 按观测值从大到小排序 $z_1 \geq z_2 \geq \dots \geq z_n$, 记排序序列 $RANK(Z) = (1, 2, 3, \dots, n)$, 最佳异常样本的排序取值 $rank = 1$, 最差异常样本的 $rank = n$ (大写 $RANK$ 表示变量, 小写 $rank$ 表示排序取值)。若其中有两个样品的取值相同, 按上述约定其先后顺序是任意的。为了避免对 $RANK(Z)$ 取值分类时, 相同的观测值 z 处于不同的分级中, 进一步约定: 将那些重复的观测值所对应的多个有序排序值($rank_s, rank_s + 1, \dots, rank_t - 1, rank_t$) ($n \geq t > s \geq 1$), 计算其排序的中位数并取整作为这一组相同观测值共同的 $rank$ 值, 计算公式为 $rank = \lceil (rank_s + rank_t) / 2 \rceil$ ($\lceil$ 为取整符号)。对 CV 和 $\Delta\alpha$ 同样是按降序排序得到对应的顺序值 $RANK(CV)$ 和 $RANK(\Delta\alpha)$, 若对奇异性指数 α , 则需按升序确定。由于受样品分析精度的限制, 元素含量值相等的样品可能占有一定比例, 而 $CV, \Delta\alpha$ 则一般无重复值出现。

本文在异常识别效果对比中采用的是样本的排序序次(也即样本统计量的下标索引), 而没有用分位数。对异常指示变量按其分位数进行地球化学制图时, 尽管分类更加直接, 但在不同方法的对比中则相对繁琐。分位数与异常指示变量具有相同的量纲, 分位数 $q(u)$ 通常是按变量从小到大排序计算, 且一般约定 $0 < u < 1$, 对 CV 和 $\Delta\alpha$ 来说, 需换算成上侧分位数 $q(1-u)$ 才对应于高异常样品, 分位数很少用来与样本观测值建立一一对应关系。分位数作为

一种统计量, 常用 $q(0.5)$ 表示中位数(Q_2), 用 $q(0.25)$ 和 $q(0.75)$ 分别表示第 1 四分位数(Q_1) 和第 3 四分位数(Q_3), 区间 $[Q_1, Q_3]$ 包含了 50% 的不受特异值影响的数据, 内四分位数极差 $IQR = Q_3 - Q_1$ 。排序序次尽管从某种程度上损失了一定信息, 但更易于不同方法之间的对比, 它直接表明了不同方法处理所获取的异常的相对强弱, 不受量纲和分布特征的影响, 与样品一一对应, 上侧 $rank$ 值与累积面积成正比。地球化学数据作为一种典型的空间数据, 在地理信息系统(GIS)支持下, 可通过“某变量在某样品的 $rank$ 值 $\rightarrow$ 某样品空间位置(x, y) $\rightarrow$ 变量在某样品的取值”对应关系来获取异常指示变量的取值, 变量取值及 $rank$ 值的组合也即等效于确定了顺序统计量取值和上侧分位数, 其中 $u = (rank - 0.5) / n$ (n 为样本数) (Martinez and Martinez, 2005)。

2 数据来源和统计分布特征

研究区位于我国云南东南部的个旧及其周边地区, 地理范围为东经 $102^\circ 00' \sim 103^\circ 30'$, 北纬 $23^\circ 00' \sim 24^\circ 00'$, 面积约为 $18\,096\text{ km}^2$ (图 1)。研究区的大地构造位置位于滨太平洋构造域与特提斯构造域交界部位, 欧亚板块、太平洋板块和印度板块三者复合的碰撞挤压带, 属于扬子克拉通(分区 I)、华南褶皱系(分区 II)、兰坪—思茅褶皱系(分区 III) 3 大主要构造单元的交汇部位, 其北西以师宗—弥勒深大断裂为界(终止于红河断裂), 其南以红河断裂为界(图 1) (王宝禄等, 2004)。该区在地质历史中经历了复杂的地质演化过程, 区内多次构造运动和强烈的岩浆活动为该区多种矿产的形成提供了有利的地质条件(庄永秋等, 1996)。本区除有世界著名的以个旧锡矿为代表的滇东南锡矿带之外, 铜矿资源也非常丰富。不同的成矿作用在不同的地质时代于不同地质构造环境中孕育了多样化的铜矿资源。

笔者系统收集了全国区域化探扫面计划在本区的水系沉积物分析数据, 样品数 4 524 个, 按 $2\text{ km} \times 2\text{ km}$ 网格密度等间距分布(58 行 $\times$ 78 列)。EDA (exploratory data analysis) (Tukey, 1977; Martinez and Martinez, 2005) 提供了优秀的描述性统计和科学数据可视化分析技术, 它的数据分析过程无需基于假设和模型。采用 EDA 技术对 4 524 个 Cu 元素分析数据的统计分析见图 2, 其上方图形为直方

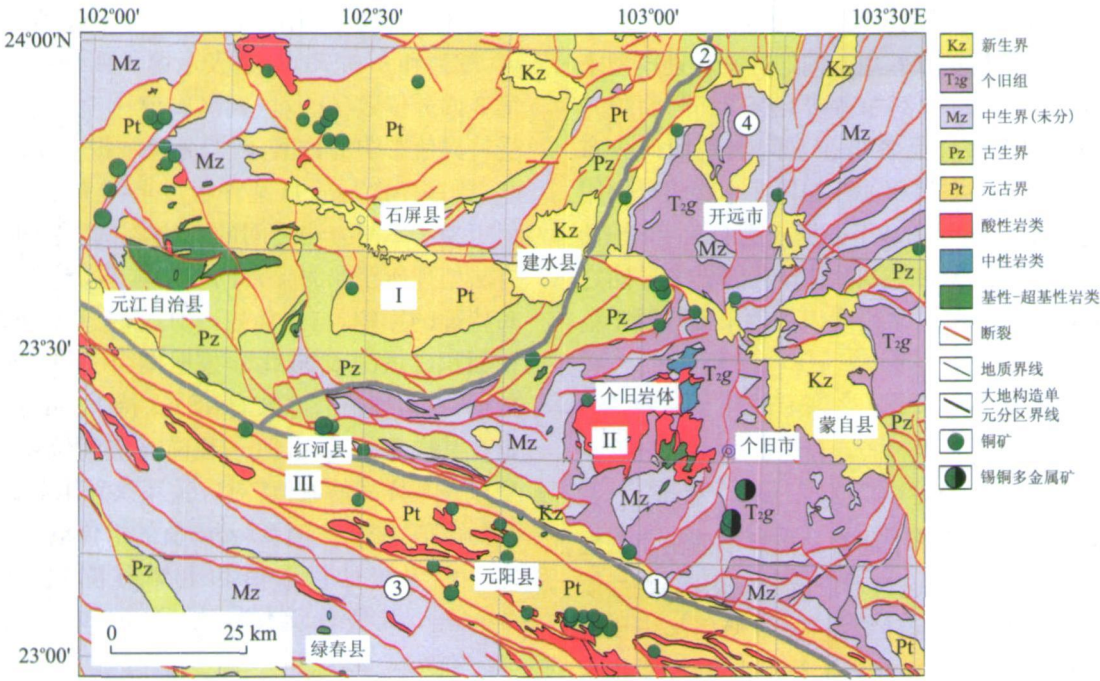


图 1 云南个旧及其周边地区地质简图

Fig. 1 Simplified geological map of the study area in Gejiu and its surrounding areas, Yunnan Province, China
I. 扬子克拉通; II. 华南褶皱系; III. 兰坪—思茅褶皱系; ①红河断裂带; ②建水—弥勒—师宗断裂带; ③哀牢山断裂带; ④小江断裂

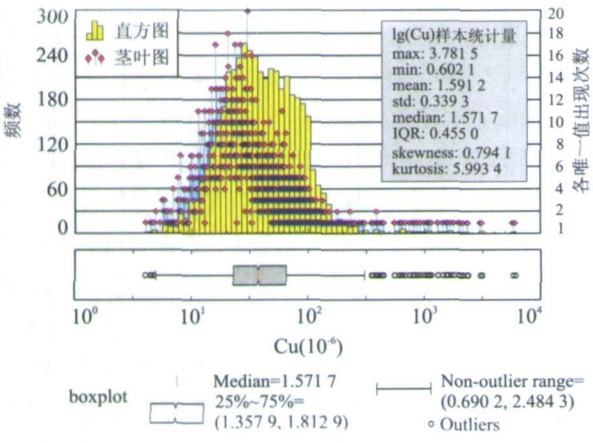


图 2 4 524 个 Cu 元素水系沉积物样品的统计分布特征
Fig. 2 Statistical distribution features of 4 524 Cu element data of the stream sediment samples

上方图形为直方图和茎叶图的叠合显示,下方为箱线图,直方图和箱线图取 10 为底的对数进行绘制。直方图统计的频数见上方图形左侧纵轴,茎叶图按含量数据唯一值进行统计,各唯一值重复出现次数见上方图形右侧纵轴

图和茎叶图的叠合显示,下方为箱线图,在上侧分布尾具有明显的特异值。受测试分析精度所限,本研究区 Cu 元素含量值相等的样品占有较大数量,含量值为 29.8 的样品出现频数最多(即为众数,其常用对数值为 1.471 2),达 20 次,含量取值中仅有 395 个数据未有重复值,占总数的 8.73%(图 2)。按照本

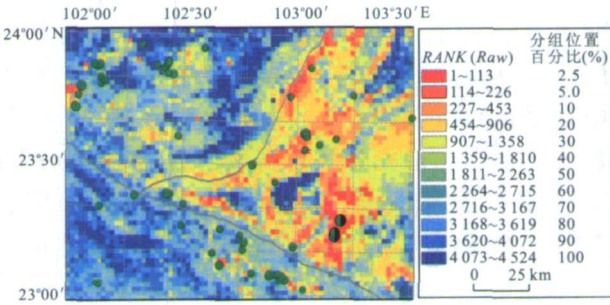


图 3 Cu 元素含量 RANK 值色块图

Fig. 3 Geochemical mapping for Cu concentrations
图中圆形符号为铜矿产,各单元色块代表面积 4 km²

文约定方法,对 Cu 元素含量数据按从高到低进行排序,得到顺序值的序列记为 $RANK(Raw)$,称为原始排序。图 3 展示了 Cu 元素以 $RANK(Raw)$ 绘制的地球化学图,可见 Cu 元素在分区 II 具有大量高值,相比之下,分区 I 和 II 的化探异常显示较弱。作者按三大构造分区和地层(新生界、中生界、古生界、元古界,岩浆岩按其侵入时代归入对应时代的地层中)将全区化探数据分成 11 类,绘制了 multi-boxplot 图,并将区内出现矿产位置处的样品提取出来,将其含量值投到对应分组的 boxplot 图上以反映铜矿资源容矿的主要层位(图 4)。由图 4 可见,多数分组的 Cu 元素含量具有特异值,不同分组含量

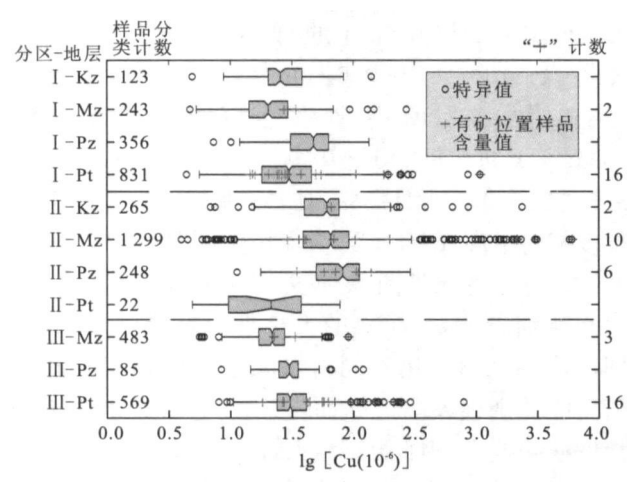


图 4 4 524 个 Cu 样品按构造分区和地层分类统计的 multi-boxplot 图

Fig.4 Multi-boxplots of 4 524 Cu samples classified by the structural zones and strata

Cu 含量数据取常用对数进行统计, 在各个 boxplot 图形上叠加了“+”符号(红色), 表示有矿位置出样品 Cu 元素含量取值, 纵轴右侧表示了这些样品的个数

的集中和离散程度具有较大差异. Cu 元素极值分布最明显、容矿数量最多的分组出现在 II-Mz(及 II-Kz)、III-Pt 和 I-Pt, 这三大类分属不同的构造分区, Cu 元素含量的样本统计量(如中位数、极差、IQR 等)及特异值分布均表现出明显的差异, 铜矿产不仅出现在含量高值区, 在 IQR 区间内也出现较多数量, 甚至在低于 Q_1 的含量区间上也占有总数的 7/55. Cu 元素的空间分布和统计特征表明, 研究区地球化学场的分布复杂, 需要挖掘空间结构、尺度特征等隐含在数据内部的深层次信息才能获得好的异常识别效果, 弱缓异常的识别是本区化探异常找矿的关键. 笔者以异常衬度法和局部奇异性分析法为例, 介绍它们在本地区的应用效果.

3 滑动衬值和局部奇异性指数计算方法

3.1 滑动衬值

异常衬度又称衬值(contrast value), 其计算公式为:

$$CV(x) = z_a(x)/z_b(x), \quad (1)$$

其中, $z_a(x)$ 为 x 位置处元素在异常范围内异常的平均值, $z_b(x)$ 为 x 位置处元素的背景平均值. 对于网格化数据, 经常采用滑动衬值进行地球化学异常制图, 背景值计算具体可分为常规滑动平均和逐步剔除特异值(按三倍标准方差逐步截尾) 2 种方法

(熊光楚等, 1997). 计算窗口的选择具有一定人为性, 本文选取的异常范围窗口大小为 1×1 (一个单元窗口大小为 $2 \text{ km} \times 2 \text{ km}$, 下同), 背景窗口大小为 11×11 . 背景平均值共分 4 种情形进行计算: 原始数据常规滑动平均计算衬值、原始数据逐步剔除特异值计算衬值、原始数据取常用对数后的常规滑动平均计算衬值、原始数据取常用对数后的逐步剔除特异值计算衬值, 以上滑动衬值计算结果分别表示为: $CV1$ 、 $CV2$ 、 $CV1_{(lg)}$ 、 $CV2_{(lg)}$. 对其排序结果分别记为 $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{(lg)})$ 、 $RANK(CV2_{(lg)})$.

3.2 局部奇异性分析

局部奇异性度量模型可表示为 (Cheng, 2006b; 陈志军, 2007):

$$\langle z(x, l) \rangle = c(x) (l/L)^{\alpha(x)-E}, \quad (2)$$

其中, $\langle \rangle$ 表示统计期望, E 为空间维数, L 为固定的最大尺度, $z(x, l)$ 为以 x 为中心在 l^E 空间尺度范围上某种元素的含量, 指数 $\alpha(x)$ 表示在位置 x 处的奇异性指数, 因子 $c(x)$ 是在 $\alpha(x)$ 维空间中的“分形密度” (成秋明, 2008).

对于网格化数据, 局部奇异性可按如下步骤计算: (a) 由小到大确定滑动平均窗口: $l_1 \times l_1$, $l_2 \times l_2$, $\dots$, $l_m \times l_m$, 得到 m 幅滑动平均图, 记为 Z_1 , Z_2 , $\dots$, Z_m . (b) 取 $L = \max\{l_i\} = l_n$, 其中 $i = 1, 2, \dots, m$, 以 $\log(Z_i \times (l_i/L)^E)$ 为因变量, $\log(l_i/L)$ 为自变量, 遍历所有空间位置进行最小二乘拟合, 回归直线的斜率即该位置处的 α 值, 令 $l = L$, 则回归直线拟合值即为 $\log(c)$ 值.

本文选取的窗口大小序列为 1×1 , 2×2 , $\dots$, 11×11 (最大窗口的范围与计算滑动衬值时背景窗口范围大小相等, 共 6 个), $L = 11$ 进行局部奇异性指数的计算, 并转换成 $\Delta\alpha = E - \alpha$. 分两种情形进行计算: 原始数据直接计算和原始数据取常用对数后计算, 计算结果分别表示为 $\Delta\alpha$ 和 $\Delta\alpha_{(lg)}$. 若元素含量值有小于 1 的情形, 可将含量值扩大 10^n 倍 (可视含量单位的变化, 如由 10^{-6} 换算成 10^{-4}), 使得最小值恰能满足大于 1, 然后再取对数进行计算. 对 $\Delta\alpha$ 和 $\Delta\alpha_{(lg)}$ 进行排序, 序次分别记为 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha_{(lg)})$.

3.3 滑动衬值和局部奇异性指数的统计分布特征

对不同的异常指示变量进行排序, 样品可具有多种 rank 值. 对任一样品, 也即具有以下空间位置信息和数据取值:

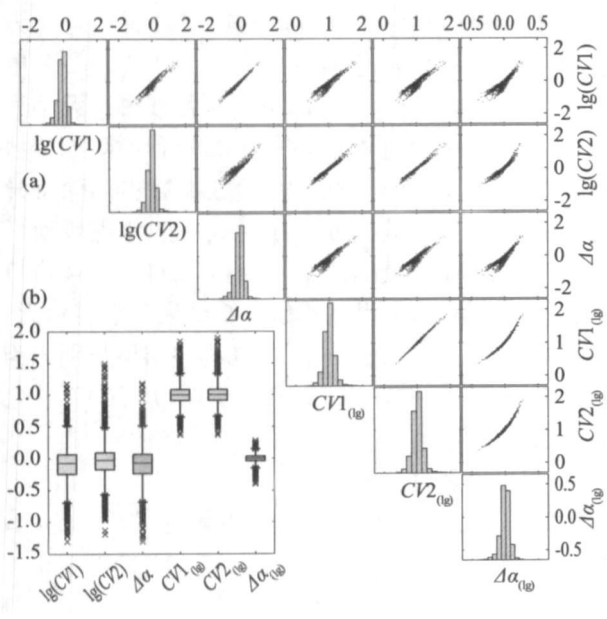


图 5 $CV1$ 、 $CV2$ 、 $CV1_{lg}$ 、 $CV2_{lg}$ 、 $\Delta\alpha$ 和 $\Delta\alpha_{lg}$ 的 marix plot 和 multi-boxplot 图

Fig. 5 Matrix plot and multi-boxplot of the $CV1$, $CV2$, $CV1_{lg}$, $CV2_{lg}$, $\Delta\alpha$, $\Delta\alpha_{lg}$ mapping for the gliding contrast values and local singularity exponents of Cu element

a. matrix plot 图, 按对称性省略了左下角的 X - Y plot 图形, 对角线上为各个变量的统计直方图; b. multi-boxplot 图, “x” 符号表示特异值。以上变量均无量纲, $CV1$ 、 $CV2$ 在图 a、b 中均先取常用对数再统计作图

$\{x, y; z(\text{含量值}), CV1, CV2, CV1_{lg}, CV2_{lg}, \Delta\alpha, \Delta\alpha_{lg}; RANK(Raw), RANK(CV1), RANK(CV2), RANK(CV1_{lg}), RANK(CV2_{lg}), RANK(\Delta\alpha), RANK(\Delta\alpha_{lg})\}$ 。

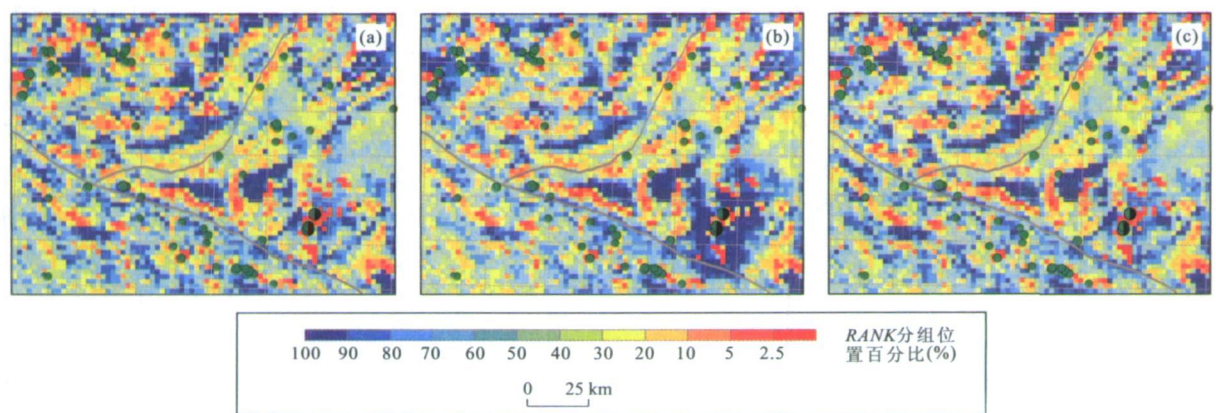


图 6 Cu 元素滑动衬值和局部奇异性指数的 $RANK$ 值色块图

Fig. 6 Geochemical mapping for the gliding contrast values and local singularity exponents of Cu element a. $RANK(CV2_{lg})$; b. $RANK(\Delta\alpha)$; c. $RANK(\Delta\alpha_{lg})$. 各色块图按相同的分类标准共分 12 类, 该分类标准与图 3 相同

$\Delta\alpha$ 计算中 R^2 拟合度的样本统计量: mean = 0.995 1, std = 0.009 6, min = 0.871 7, max = 1.000 0; $\Delta\alpha_{lg}$ 计算中 R^2 拟合度的样本统计量: mean = 0.999 8, std = 0.000 4, min = 0.991 4, max = 1.000 0. 可见 $\Delta\alpha_{lg}$ 比 $\Delta\alpha$ 在估值中回归拟合优度相对更高. $CV1$, $CV2$, $CV1_{lg}$, $CV2_{lg}$, $\Delta\alpha$, $\Delta\alpha_{lg}$ 的统计分布特征见图 5. 本文不详细阐述 CV 和 $\Delta\alpha$ 之间差别. 实际上, CV 的大小与计算窗口大小(也即尺度)有关, 采用不同窗口大小来计算 CV , 其取值可能差别很大. 对 $\Delta\alpha$ 而言, 由于计算中采用了多尺度回归技术, 因此理论上 $\Delta\alpha = E - \alpha$ 是与尺度无关的. 若 α 的估值具有较高的拟合度, 则 α 和 CV 存在如下关系: $\alpha \propto \frac{\log(z(l_m)/z(l_h))}{\log(l_m/l_h)} \propto \log(CV)$ (拟合度 $R^2 \rightarrow 1$ 时), 这种近似关系可反映在图 5a 所示的 X - Y plot 图中. 当滑动衬值和局部奇异性指数的计算考虑更多因素, 例如借助空间 U 统计量法在不同尺度上用各向异性的窗口(椭圆状)代替方形窗口进行局部奇异性指数的计算, 此时的局部奇异性指数和滑动衬值就可能存在较大的差异, 异常识别的效果也将有所不同(陈志军, 2007). 在图 5a 所示的 X - Y plot 图中, 各个横轴变量的高值部分和低值部分相关程度均有所差异. 从异常识别角度, 主要关注这些变量的高值部分.

各种异常指示变量采用不同的分类标准与图面着色方案可以制作多种形式的地球化学图, 为便于比较, 笔者应用与 $RANK(Raw)$ 相同的分类标准和颜色渲染方案对以上滑动衬值和局部奇异性指数的排序值分别进行了地球化学制图(图 6), $RANK(Raw)$ 、 $RANK(CV1)$ 、 $RANK(CV2)$ 、

$RANK(CV1_{(lg)})$ 、 $RANK(CV2_{(lg)})$ 、 $RANK(\Delta\alpha)$ 、 $RANK(\Delta\alpha_{lg})$ 的地球化学图总体特征大体类似,但在局部位置具有一定差异性.图 6 是 $CV1_{(lg)}$ 、 $\Delta\alpha$ 和 $\Delta\alpha_{lg}$ 关于 $rank$ 值分类的地球化学色块图,从中可见它们均显著降低了分区 II 中的高背景,同时还突出了分区 I 和 III 的异常显示.对 CV 和 $\Delta\alpha$ 之间异常识别效果的细节差异在排序值的地球化学图中尚不能清晰辨别,下面笔者就 3 个方面来比较不同处理方法之间的异常识别效果.

4 讨论

4.1 高背景区和低背景区不同异常指示变量的排序特征

笔者取 $RANK(Raw)$ 前 1.25% 的样品(56 个),这些样品主要分布在图 1 的分区 II,图 7a 所示的 X - Y plot 图展示了在高背景区 $RANK(Raw)$ 、 $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{(lg)})$ 、 $RANK(CV2_{(lg)})$ 、 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha_{lg})$ 之间的异同.从图 7 中可见,这些原始排序相对靠前的样品,仅有较少的样品在 CV 和 $\Delta\alpha$ 的样本排序中仍然靠前,较多样品的排序都被下调.下调的幅度以 $RANK(\Delta\alpha)$ 和 $RANK(CV1)$ 为最.这 56 个高背景样品中有部分样品的 $\Delta\alpha$ 值小于 0,从局部奇异性角度被识别为负异常.笔者由 $lg(Cu)$ 的第 1 四分位数 1.357 9 向低值区也取 56 个样品,这些样品主要分布在分区 I 和 III,从图 7b 的 X - Y Plot 图可见,其对原始排序的改变情形和图 7a 恰好相反,这些原始排序相对靠后的样品,较多样品的 CV 和 $\Delta\alpha$ 的样本排序被上调,部分 $\Delta\alpha$ 值大于 0、 CV 值大于 1 可被筛选为异常.从图 7 可见,滑动衬值方法和局部奇异性分析方法具有压制高背景、突出弱缓异常的效果.

4.2 有矿位置处化探样品不同异常指示变量的排序特征

笔者采用与图 7 一样的比较策略,将研究区中已知矿产地理位置处所对应的 55 处样品提取出来,进而比较各类方法的排序值(图 8).从图 8a 可见,在原始排序靠前的 10 个样品(top 3%),滑动衬值和局部奇异性指数的排序值略有不同,改变幅度不大, CV 值和 $\Delta\alpha$ 值仍在异常可被显著识别的区间内;而后面 45 个样品($RANK(Raw)=100$ 之后),衬值、局部奇异性指数对原始排序的改变幅度都很大.从表 1 可见,在相对低背景区,不同异常指示变量总体

表 1 不同异常指示变量在原始排序后 45 个有矿位置处化探样品的排序升降情况

Table 1 Rank changes of different anomaly indexes at the sites of 45 Cu deposits occurring with low $RANK(Raw)$ values

	<i>rank</i> 降低	<i>rank</i> 相等	<i>rank</i> 升高
$RANK(CV1)$	20	0	25
$RANK(CV2)$	18	0	27
$RANK(CV1_{(lg)})$	17	0	28
$RANK(CV2_{(lg)})$	17	0	28
$RANK(\Delta\alpha)$	17	0	28
$RANK(\Delta\alpha_{lg})$	15	1	29

上排序值升高者比降低者明显多,其中以 $\Delta\alpha_{lg}$ 的 $rank$ 值升高的样品个数最多,约占 64%,这使得在低背景中的某些样品有可能被识别出来.图 8b 是按 $RANK(\Delta\alpha)$ 升序排列比较,在 $\Delta\alpha>0$ 的有矿出现位置处的 32 个样品中($CV2_{(lg)}$ 值也都大于 1),Cu 元素含量值变化范围较大,最高者达 $6\,046.3\times 10^{-6}$,而低者仅为 33.5×10^{-6} ,这反映衬值和局部奇异性指数对弱缓异常识别的有效性.有部分异常样品用衬值和局部奇异性指数仍然难于识别,这可能与水系沉积物所获的异常与真实异常源之间具有不同程度的位移有关,也可能由于 CV 和 $\Delta\alpha$ 计算中需要考虑各向异性、地质构造背景的控制作用等因素,当然还受异常下限确定方法的影响.

4.3 不同异常指示变量等面积含矿率比较与最佳异常下限确定

对已有较多已知矿产发现的地区进行矿产资源定量预测工作中,常将最大含矿率及最小漏矿率作为圈定远景区或靶区的准则之一.这里采用证据权法来对各类异常识别模型圈定矿床(点)的效果进行比较.笔者对 $RANK(Raw)$ 、 $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{(lg)})$ 、 $RANK(CV2_{(lg)})$ 、 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha_{lg})$ 都采用相同的分类标准,高 $rank$ 区间具有更小的分组间距,以更精细地反映高异常区的变化性,共分 68 组不同的面积累积范围(对于原始排序,有些样品具有相同 $RANK$ 值,约定取中位数),从图 2 所示茎叶图可见最多的重复出现 20 次,用 $rank/4\,524\times 100\%$ 来计算累积面积百分比(实际累积面积的误差可忽略不计).在每一分类区间,计算 t 统计量,其公式为: $t=C/s(C)$,其中 $C=W^+-W^-$, C 称为对比度,反映证据图层同矿产的空间相关性大小, W^+ 和 W^- 分别称为证据正权和证据负权, $s(C)$ 为对比度 C 的标准方差. t 值越大,

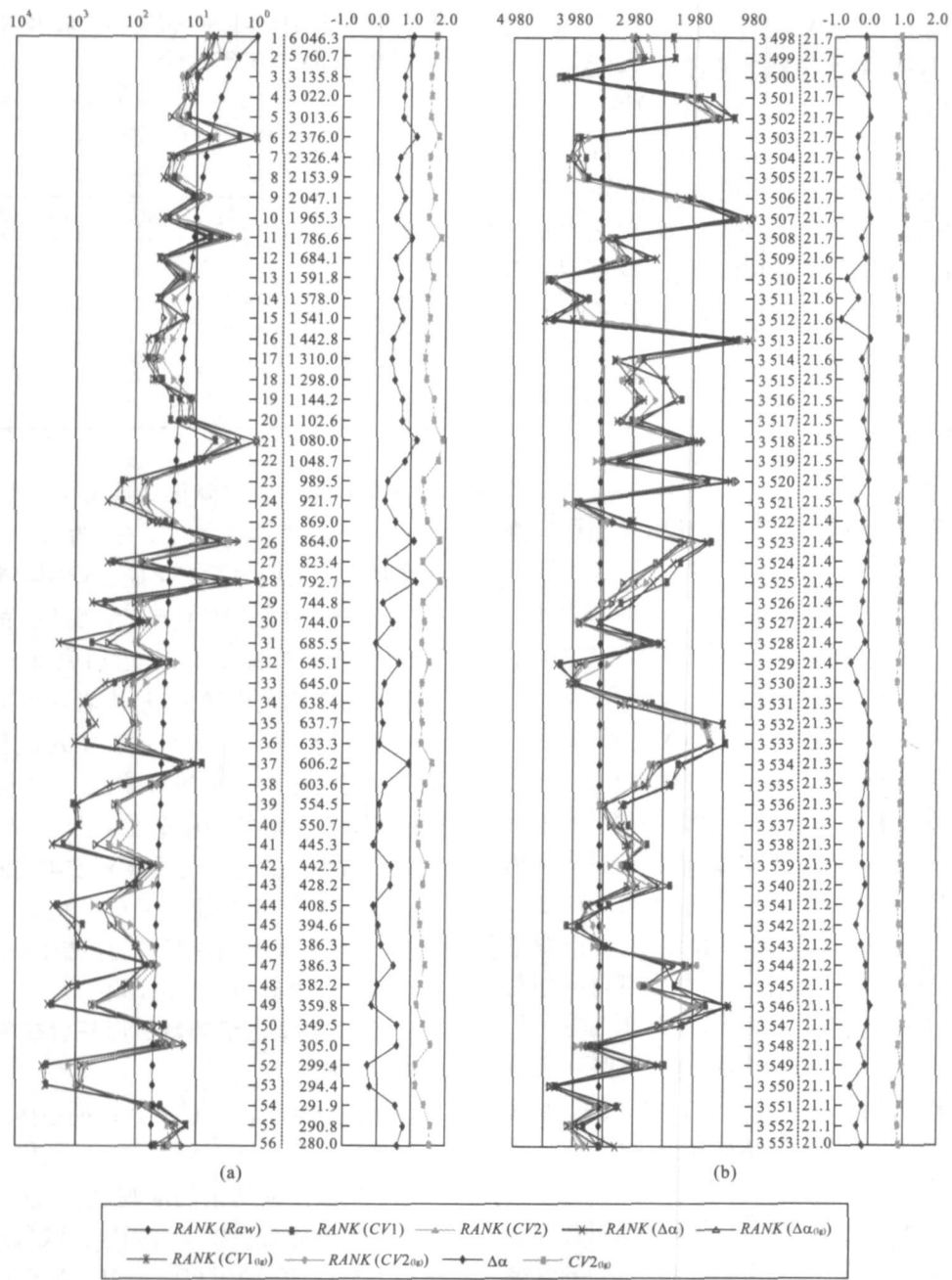


图 7 高背景区和低背景区不同异常指示变量的排序比较

Fig. 7 Comparison of sample ranking of the different geochemical anomalies in the areas with the high and low back-grounds respectively

a. 各类方法对 $RANK(Raw) \leq 56$ (top 1.25%) 样品处理后排序值的变化; b. 各类方法对 $3498 \leq RANK(Raw) \leq 3553$ 样品(自 $\lg(Cu)$ 的第 1 四分位数 1.3579 向低值区取 56 个) 处理后排序值的变化. a, b 两图纵轴为分类轴, 标注格式为“Cu 元素的 rank 值| Cu 元素含量值”, 自上而下按 Cu 元素的 rank 值升序排列; 左图横轴表示 rank 值, 右图横轴表示与 rank 值对应样品计算所得的 $\Delta\alpha$ 值和 $CV2_{0p}$ 值, 右图对分类轴各样品的属性作辅助说明

说明在该累积面积上化探异常与矿化关系空间相关的显著性程度越高. 从图 9 所示的累积面积 - t 统计量关系图可见, 在前 50% 累积面积的那些分组中, $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{1p})$ 、 $RANK(CV2_{1p})$ 、 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha_{1p})$ 的 t

曲线几乎都在 $RANK(Raw)$ 的 t 曲线上方, 也即无论按多少异常面积来进行比较, 衬值和局部奇异性指数的矿化异常识别效果将明显优于元素含量值的效果. 在 1.25%、21% 累积面积附近, $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{1p})$ 、 $RANK$

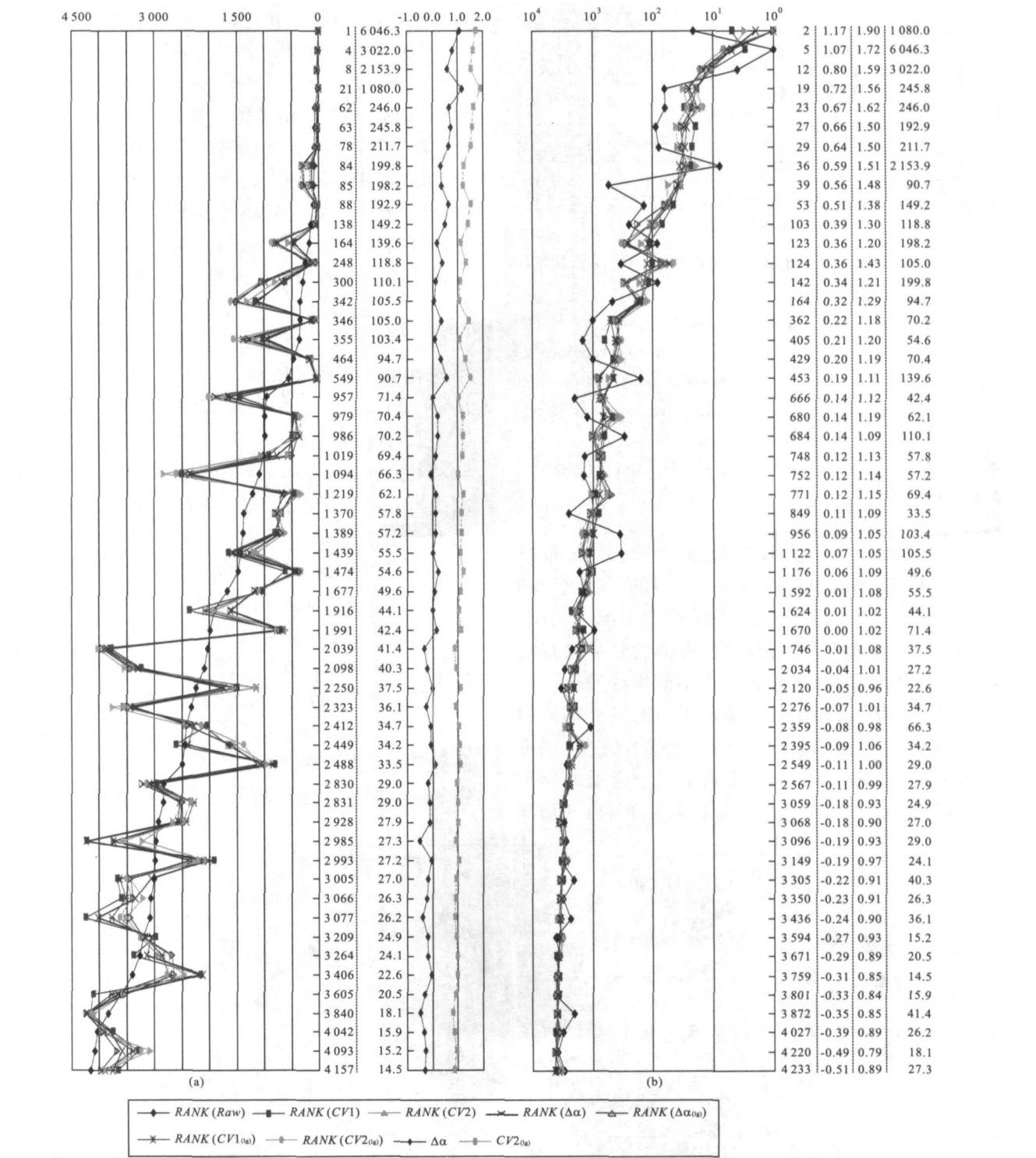


图 8 不同异常指示变量在 55 个铜矿位置处化探样品的排序比较

Fig. 8 Comparison of the sample ranking of the different geochemical anomaly indexes at the sites of 55 Cu deposits occurring

a. 按 $RANK(Raw)$ 升序排列比较。纵轴为分类轴, 标注格式为“Cu 元素的 rank 值|Cu 元素含量值”, 自上而下按 Cu 元素的 rank 值降序排列; 左图横轴表示 rank 值, 右图横轴表示与 rank 值对应样品计算所得的 $\Delta\alpha$ 值和 $CV2_{lg}$ 值, 右图对分类轴各样品的属性作辅助说明。b. 按 $RANK(\Delta\alpha)$ 升序排列比较。纵轴为分类轴, 标注格式为“ $\Delta\alpha$ 的 rank 值| $\Delta\alpha$ 值| $CV2_{lg}$ 值|Cu 元素含量值”, 自上而下按 $\Delta\alpha$ 的 rank 值升序排列; 横轴表示 rank 值

($CV2_{lg}$)、 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha_{lg})$ 的 t 曲线呈现上凸形状, 而 $RANK(Raw)$ 的 t 曲线则反呈下凹形状, 在 3% 累积面积附近, $RANK(\Delta\alpha)$ 和 $RANK(CV1)$ 的 t 曲线形状呈现为上凸而

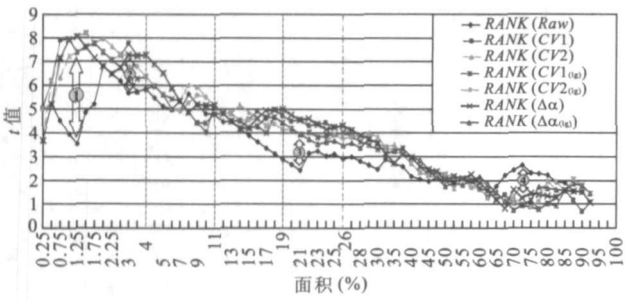


图 9 不同模型异常指示变量应用证据权法的化探异常累积面积— t 统计量图

Fig.9 Student's t statistics calculated by weights of evidence method for measuring spatial correlation between locations of 55 Cu deposits and the areas with geochemical anomaly indexes employing different models for anomaly recognition

箭头及其数字 1, 2, 3, 4 标明了 $RANK(Raw)$ 的 t 曲线与其他 t 曲线具有明显差异的位置

$RANK(Raw)$ 的 t 曲线则反呈下凹形状;而在 72.5% 累积面积附近, $RANK(Raw)$ 的 t 曲线与其他的 t 曲线与前 50% 累积面积的凸凹情况则相反. 这说明应用元素含量圈定异常, 在高背景区进行空间统计分析时, 面积的显著增加不能保证发现矿产数量的同步增长;而在低背景处, 则可能出现较多的矿床(点), 这使得 t 曲线的变化趋势复杂多变, 很难根据 $RANK(Raw)$ 的 t 曲线确定一个合适的异常范围, 即异常下限难于确定. 而对于衬度异常和局部

奇异性指数来说, 其排序的 t 曲线的总体变化趋势均为: 随着累积面积的增大, t 曲线不断向上攀升并达到顶峰(明显高于 1.96), 然后再不断下降. 由于本研究区中成矿环境的复杂性, $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1(lg))$ 、 $RANK(CV2(lg))$ 、 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha(lg))$ 的 t 曲线表现出一定的局部波动, 以 $RANK(\Delta\alpha)$ 的 t 曲线为例, 在 1.25%、4%、11%、19%、26% 累积面积处($rank=56, 180, 497, 859, 1176$) t 曲线分别达到局部极大值, 在其面积内包含的矿床(点)与发现矿产占总数比值依次为 10/55、15/55、19/55、26/55 和 29/55. 尽管在某些位置, $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1(lg))$ 、 $RANK(CV2(lg))$ 的 t 值略高于 $RANK(\Delta\alpha)$ 和 $RANK(\Delta\alpha(lg))$, 但总体说来, $RANK(\Delta\alpha)$ 相对更佳, 其 t 曲线在前 30% 面积内相对其他曲线始终保持较高的 t 值. 在 $RANK(\Delta\alpha)$ 的 t 曲线图上确定以累积面积 4%、11%、19% 作为分级阈值, 在 $RANK(\Delta\alpha)$ 地球化学色块图上经双线性插值生成光滑平面图, 以 $rank=859$ (累积面积百分比 19%, $\Delta\alpha=0.054$) 作为异常下限, 圈定了化探异常远景区(图 10). 从图 10 可见, 化探异常空间分布与多数已知矿床(点)分布的空间吻合程度较高, 在 3 个不同分区化探异常都有较好的显示, 对比图 3 的高异常区基本集中分布在分区 II, $RANK(\Delta\alpha)$ 所圈定的异常图很好地抑制了三叠系的高背景影响, 分

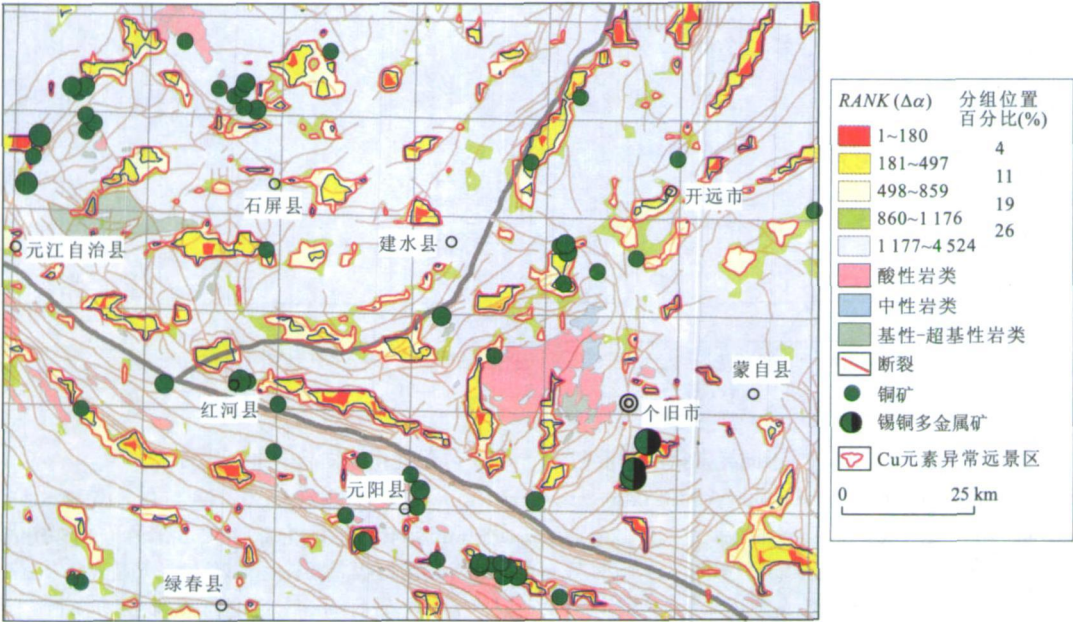


图 10 用 $RANK(\Delta\alpha)$ 圈定 Cu 元素化探异常远景区
Fig. 10 Prospective areas of Cu element delineated by $RANK(\Delta\alpha)$

区I和分区II中一些低缓异常被较好圈定,对未知矿床的预测提供了元素含量值所不具备的有用信息,比用元素含量值圈定异常远景区更具预测意义。用 $RANK(CV1)$ 、 $RANK(CV2)$ 、 $RANK(CV1_{(g)})$ 、 $RANK(CV2_{(g)})$ 和 $RANK(\Delta\alpha_{(g)})$ 所确定的异常远景区与 $RANK(\Delta\alpha)$ 所圈定的在总体上是一致的,这也从一个侧面反映了利用 $\Delta\alpha$ 圈定异常的可靠性。

5 结论

对元素含量、滑动衬值、局部奇异性指数这些不同的异常指示变量获取其样本排序取值,可以消除分布特征和量纲的影响,在等面积条件下进行相对异常强度的比较,对个旧及其周边地区铜元素的水系沉积物应用结果表明,利用样本排序方法比较异常识别模型的效果是一种可行的方法,这种方法可推广应用到其他化探异常识别模型的效果对比中。

异常识别模型压制高背景、突出低缓异常的能力可从它相对样品原始排序的优化调整情况中反映出来。在高背景区, CV 和 $\Delta\alpha$ 对于样品原始排序具有显著的降低作用;在低背景区, CV 和 $\Delta\alpha$ 对于样品原始排序具有提升排序的能力。在圈定异常范围时, CV 和 $\Delta\alpha$ 由于降低了高背景区样品的排序,避免了非矿异常面积的大量增加,而代之以其他空间位置处弱缓异常面积的增加,从而保持其异常范围与矿床(点)分布具有较高分度的空间相关性,因此, CV 和 $\Delta\alpha$ 这两类方法对识别弱缓异常识别具有较好的能力。在本研究区,通过 $\Delta\alpha$ 圈定了化探异常远景区,并被滑动衬值方法的结果所佐证,具有较高的可信度,在本区三个截然不同的地质构造环境分区中都确定了一定面积的异常远景区,比用元素含量值来圈定更具预测意义,可为该区的铜矿资源找矿工作提供参考。局部奇异性分析方法是一种非常有效的弱缓异常识别方法。 $\Delta\alpha$ 可以通过原始数据直接计算,也可取对数后计算,在本研究区 $\Delta\alpha$ 通过原始数据直接计算为更佳。总体而言,局部奇异性方法原理清晰、方法简便、可操作性强,在地球化学异常识别中完全可以用其替代滑动衬值方法。

此外,相对直方图、Q-Q图等简单的统计图形,利用当今先进的计算机技术,EDA分析技术(box plot, matrix plot等)可展现更丰富的数据信息,例如离群值分布,在化探数据分析处理中值得广泛应用。

致谢: 本文在写作过程中和张生元教授、徐德义教授、夏庆霖副教授进行了有益的讨论,在此表示衷心的感谢。

References

- Agterberg, F. P., 1989. Computer programs for mineral exploration. *Science*, 245: 76—81.
- Agterberg, F. P., Bonham-Carter, G. F., Cheng, Q. M., 1993. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. In: Davis, J. C., Herzfeld, U. C., eds., *Computers in geology*. Oxford University Press, New York, 13—32.
- Bao, Z. Y., Li, F. L., Jia, X. Q., 1999. Methodology of temporal spatial structure of geochemical fields. *Earth Science—Journal of China University of Geosciences*, 24 (3): 282—286 (in Chinese with English abstract).
- Bonham-Carter, G. F., 1994. Geographic information system for geosciences: Modelling with GIS. Pergamon Press, Oxford, 398.
- Chen, Z. J., 2007. Multifractal theory based local singularity analysis method and its application in spatial information extraction for mineral exploration (Dissertation). China University of Geosciences, Wuhan (in Chinese with English abstract).
- Chen, Z. J., Cheng, Q. M., Chen, J. G., et al., 2007. A novel iterative approach for mapping local singularities from geochemical data. *Nonlin. Processes Geophys.*, 14: 317—324.
- Cheng, Q. M., 1999a. Spatial and scaling modeling for geochemical anomaly separation. *Journal of Geochemical Exploration*, 65: 175—194.
- Cheng, Q. M., 1999b. Multifractality and spatial statistics. *Computers & Geosciences*, 25(9): 949—961.
- Cheng, Q. M., 2006a. GIS-based fractal/ multifractal anomaly analysis for modeling and prediction of mineralization and mineral deposits. In: Harris, J. R., ed., *GIS application in the earth sciences—GAC special paper*. Geological Association of Canada Special Book, 285—296.
- Cheng, Q. M., 2006b. 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. M., 2008. Singularity of mineralization and multifractal distribution of mineral deposits. *Bulletin of Mineralogy, Petrology and Geochemistry*, 27(3): 299—305 (in Chinese with English abstract).
- Cheng, Q. M., Agterberg, F. P., 2008. Singularity analysis of

- ore-mineral and toxic trace elements in stream sediments. *Computers & Geosciences*, DOI: 10.1016/j.cageo.2008.02.034.
- Cheng, Q. M., Agterberg, F. P., Ballantyne, S. B., 1994. The separation of geochemical anomalies from background by fractal methods. *Journal of Geochemical Exploration*, 51(2): 109—130.
- Cheng, Q. M., Agterberg, F. P., Bonham-Carter, G. F., 1996. Spatial analysis method for geochemical anomaly separation. *Journal of Geochemical Exploration*, 56(3): 183—195.
- Cheng, Q. M., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. *Natural Resources Research*, 9(1): 43—51.
- Grunsky, E. C., 2007. The interpretation of regional geochemical survey data. In: Gubins, A. G., ed., *Proceedings of Exploration 2007: Fifth Decennial International Conference on Mineral Exploration*, 139—182.
- Han, D. Y., Gong, Q. J., Xiang, Y. C., 2004. Some new fractal methods for regional geochemical survey data processing. *Geological Bulletin of China*, 23(7): 714—719 (in Chinese with English abstract).
- Li, C. J., Ma, T. H., Shi, J. F., 2003. Application of a fractal method relating concentrations and distances for separation of geochemical anomalies from background. *Journal of Geochemical Exploration*, 77: 167—175.
- Martinez, W. L., Martinez, A. R., 2005. Exploratory data analysis with MATLAB. Chapman & Hall/CRC, New York, 405.
- Reimann, C., Filzmoser, P., Garrett, R. G., 2005. Background and threshold: Critical comparison of methods of determination. *Science of the Total Environment*, 346: 1—16.
- Shi, C. Y., Zhang, J. H., Huang, X. M., 1999. Subregion median contrast filtering method and recognition of weak anomalies. *Geophysical & Geochemical Exploration*, 23(4): 250—257 (in Chinese with English abstract).
- Tukey, J. W., 1997. Exploratory data analysis. Addison-Wesley, Reading.
- Wang, B. L., Lü, S. K., Hu, J. G., 2004. A tentative description of the Chuan-Dian-Qian rhombic massif. *Yunnan Geology*, 23(2): 140—153 (in Chinese).
- Xiong, G. C., Xie, D. S., Zhang, W. B., et al., 1997. The rapid exploration method and technique systems for metallic minerals in Xinjiang Province. Geological Publishing House, Beijing (in Chinese).
- Zhuang, Y., Q., Wang, R. Z., Yang, S. P., et al., 1996. Geology of Gejiu tin-copper polymetallic deposit. China Seismic Publishing House, Beijing (in Chinese).

附中文参考文献

- 鲍征宇, 李方林, 贾先巧, 1999. 地球化学场时空结构分析的方法体系. 地球科学——中国地质大学学报, 24(3): 282—286.
- 陈志军, 2007. 多重分形局部奇异性分析方法及其在矿产资源信息提取中的应用(博士学位论文). 武汉: 中国地质大学.
- 成秋明, 2008. 成矿过程奇异性与矿床多重分形分布. 矿物岩石地球化学通报, 27(3): 299—305.
- 韩东昱, 龚庆杰, 向运川, 2004. 区域化探数据处理的几种分形方法. 地质通报, 23(7): 714—719.
- 史长义, 张金华, 黄笑梅, 1999. 子区中位数衬值滤波法及弱小异常识别. 物探与化探, 23(4): 250—257.
- 王宝禄, 吕世琨, 胡居贵, 2004. 试论川滇黔菱形地块. 云南地质, 23(2): 140—153.
- 熊光楚, 谢德顺, 张文斌, 等, 1997. 新疆金属矿产快速勘查方法技术系统. 北京: 地质出版社.
- 庄永秋, 王任重, 杨树培, 等, 1996. 云南个旧锡铜多金属矿床. 北京: 地震出版社.