

KL 变换在阳春—石录地区分散流
数据发掘中的应用

吴虹, 陈春香

(桂林工学院, 桂林 541004)

[摘要]在广东阳春—石录地区的二轮找矿中,采用数字图像处理 KL 变换技术,对 812 km²、1:50 000 比例尺的成矿元素—Cu、Pb、Zn、W、Mo、Sn、Ag、Hg 的分散流数据,开展了以隐伏矿床数据发掘为目标的地球化学可视化分析。通过对这 8 种成矿元素组合的 KL 变换的第一、第二和第三主因子图像特征的提取,结合单元彩色编码图像分析与综合地质解译,获得了过去采用常规化探数据处理方法难于获得的隐伏矿床信息,在此基础上,提出了 6 个新的找矿预测区。

[关键词]地球化学可视化分析 KL 变换 数据发掘 阳春—石录

[中图分类号]P632 **[文献标识码]**A **[文章编号]**0495-5331(2003)04-0054-06

0 概述

阳春—石录地区是我国南方重要的铜矿产区之一。自 20 世纪 50 年代以来,先后有湖南冶金地质勘探公司、广东地质矿产局、广东有色地质勘查局、宜昌地质矿产研究所、中国有色金属工业总公司矿产地质研究院和桂林工学院等单位在此开展过地质物化探找矿工作,在该区探明了一批以铜为主的多金属矿床,如石录次生堆积型铜矿床、锡山石英脉型和云英岩型钨锡多金属矿床、十二排铅锌矿床、小南山钨锡多金属矿床和鸚鵡岭钨锡多金属矿床等。其中,石录铜矿为大型矿床,其铜矿产量是石录矿业公司矿产主要来源。但是,这些矿山经过半个世纪持续不断开采,已面临着保有储量即将耗尽危机,寻找新矿床远景储量已成为十分紧迫的问题。为此,90 年代以来,在该区开展了以寻找隐伏矿床为目标的“二轮找矿”工作。

根据 1991 年桂林工学院承担的广东有色地质局科研课题计划,采用基于 KL 变换的地球化学数据数字图像处理技术,对该区分散流数据(812 km², 1:50 000 比例尺)和次生晕数据(200 km², 1:10 000 比例尺)开展了综合信息可视化分析研究,发掘提取常规化探数据处理方法难于获得的隐伏矿床异常

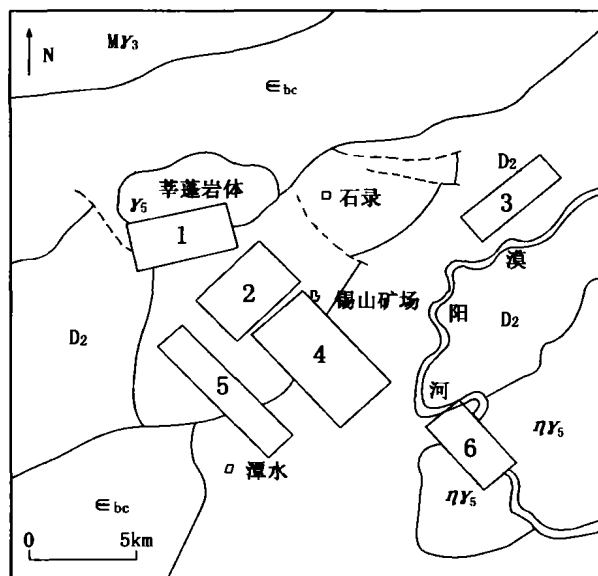


图1 阳春—石录地区分散流数据 KL 变换找矿
信息提取与找矿预测图

E_{bc}—寒武系八村群:片岩、石英岩、变质砂岩;D₂—泥盆系中统桂头组:砂砾岩、页岩;C₂—石炭系中统黄龙群:工灰岩、白云岩;γ₅—燕山期花岗岩;ηγ₅—燕山期二长花岗岩;MY₃—加里东期混合花岗岩;(矩形框为找矿预测区,框内数字为预测区编号)

信息,在此基础上,通过综合地质解释,圈定“二轮找矿”预测区。经一年工作,圈定了一批新的找矿

[收稿日期]2002-06-11;**[修订日期]**2002-08-29;**[责任编辑]**余大良。

[第一作者简介]吴虹(1947年-),男,1995年毕业于中南大学,获硕士学位,研究员,现主要从事环境与资源遥感、GIS、综合地学信息可视化分析、环境动力学模型教学与科研工作。

预测区(图1)。本文主要介绍该技术对分散流数据应用的情况。

1 依据与原理

1.1 依据

采用数字图像处理技术分析区域地球化学数据已有多年历史,最早可以追溯到1978年Stan Aronoff和Wayne Goodfellow在加拿大夸伊特湖地区对水系沉积物地球化学数据开展图像处理分析的应用^[1]。国内自80年代中期以来,在引进和自主开发这项技术上也开展了大量工作,取得了许多成果。但早期的应用一般都侧重于单元素异常增强显示和多元地学图像数据overlay分析,虽然也在运用KL变换提高多来源地学数据解释综合度上作过一些工作,但基本停留于探索阶段^[2,3]。

KL(Karhunen - Loeve)变换是一种二维酉变换。理论上,二维酉变换在图像处理中有3种主要功能:第一,从图像中提取特征;第二,图像变换编码;第三,减少计算维数。这些功能在数字图像处理和遥感多图像变换组合分析中得到了成功的应用^[4,5]。根据矿勘查数字图像处理学^[6]和数据发掘(KDD)理论,地球化学场的空间分布可看作是一种大尺度的图像,从地球化学数据中发掘找矿信息的问题,可以看作是是一种图像特征信息提取问题^[7-8]。虽然由于地质成矿作用以及地球表生作用的双重复杂性,要集中反映多种成矿元素次生地球化学场相关性的空间分布是困难的,但是,研究区具有812 km²的广阔区域,8种元素的分散流数据在区内分布基本均匀,这些条件确保了借助KL变换第一和第三种功能来实现找矿信息发掘的可能。

1.2 原理

设 $f(x, y)$ 为二维酉变换前的地球化学场图像向量(输入), $g(x, y)$ 为二维酉变换后的地球化学场图像向量(输出), (x, y) 为地球化学场图像坐标,则由 $f(x, y)$ 变换为 $g(x, y)$ 的二维酉变换表示为^[4-5]:

$$g = A \cdot f \quad (1)$$

$$f = [f_1(x, y) \ f_2(x, y) \cdots f_n(x, y)]^T \quad (2)$$

$$g = [g_1(x, y) \ g_2(x, y) \cdots g_n(x, y)]^T \quad (3)$$

$$\Sigma = \begin{bmatrix} \sum_{11} & \sum_{12} & \cdots & \sum_{1n} \\ \sum_{21} & \sum_{22} & \cdots & \sum_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{n1} & \sum_{n2} & \cdots & \sum_{nn} \end{bmatrix} \quad (4)$$

式中, A 为KL变换矩阵,由对各地化图像向量的协方差矩阵的主成分分析确定。矩阵中 $\sum_{ij}$ 为第 i, j 个地化变量的协方差(也可用标准差 s_{ij} 表示)

$$\sum_{ij} = s_{ij}^2 = \sqrt{\sum_{k=1}^n (f_{ik} - \bar{f}_i)(f_{jk} - \bar{f}_j)/(n-1)} \quad (5)$$

式中, f_{ik} 为第 i 个地球化学元素图像的第 K 个象元点的数值, f_{jk} 为第 j 个地球化学元素图像的第 K 个象元点的数值, $\bar{f}_i, \bar{f}_j$ 分别为第 i, j 个地化量图像的灰度均值, n 为 f_i 和 f_j 图像的象元点数。

为使KL变换后各图像向量间的相关性最小,必须找到协方差矩阵 Σ 的对角阵 Λ ,采用主成份分析法,有 $\Sigma A = \Lambda A$ (6)

式中 Λ 为协方差矩阵的特征值, A 为 Σ 的特征向量。通过解上式得:

$$\Lambda = \begin{bmatrix} \lambda_{11} & & & \\ & \ddots & & \\ & & \lambda_{ij} & \\ & & & \ddots \\ & & & & \lambda_{nm} \end{bmatrix} \quad (7)$$

由 Σ 矩阵及 Λ 对角矩阵可以列出以下方程

$$\begin{cases} (\sum_{11} - \lambda_1)A_1 + \cdots + \sum_{1n}A_n = 0 \\ \cdots (\sum_{22} - \lambda_2)A_2 + \cdots + \sum_{2n}A_n = 0 \\ \cdots \cdots \cdots \\ \sum_{n1}A_1 + \cdots + (\sum_{nn} - \lambda_n)A_n = 0 \end{cases} \quad (8)$$

求解此方程可得

$$\left. \begin{aligned} A_1 &= (A_{11}A_{12}\cdots A_{1n}) \\ A_2 &= (A_{21}A_{22}\cdots A_{2n}) \\ \cdots &\cdots \cdots \cdots \cdots \cdots \cdots \\ A_n &= (A_{n1}A_{n2}\cdots A_{nn}) \end{aligned} \right\} \quad (9)$$

将 $A_1, A_2, \cdots, A_n$ 代入(1)式,即可求得KL变换后的各分量的组合图像 $g_1(x, y), g_2(x, y), \cdots, g_n(x, y)$,即

$$\begin{bmatrix} g_1(x, y) \\ g_2(x, y) \\ \cdots \\ g_n(x, y) \end{bmatrix} = \begin{bmatrix} A_{11} & \cdots & A_{12} & \cdots & A_{1n} \\ A_{21} & \cdots & A_{22} & \cdots & A_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \\ \cdots \\ f_n(x, y) \end{bmatrix} \quad (10)$$

因有

$$\begin{aligned} g_1(x, y) &= A_{11}f_1(x, y) + A_{12}f_2(x, y) + \cdots + A_{1n}f_n(x, y) \\ g_2(x, y) &= A_{21}f_1(x, y) + A_{22}f_2(x, y) + \cdots + A_{2n}f_n(x, y) \end{aligned}$$

$$g_n(x, y) = A_{n1}f_1(x, y) + A_{n2}f_2(x, y) + \cdots + A_{nn}f_n(x, y) \quad (11)$$

此即 KL 变换得到的多图像线性组合结果。通常,按顺序将 $g_1(x, y), g_2(x, y), \cdots, g_n(x, y)$ 称为第一主分量图像、第二主分量图像、 $\cdots$ 、第 n 主分量图像。实际应用表明,在 KL 变换组合各分量间相关性最小的情况下, $g_1(x, y)$ 集中了最大的信息量, $g_2(x, y)$ 次之,随后的各分量集中的信息量按顺序递减。利用 KL 变换组合的这种信息压缩功能,可以对多元二维地化数据资料进行有效的综合分析解译。本课题共对 8 种不同元素图像组合作了 KL 变换分析,详见 3.2。

2 数字图像模型

地球化学元素数字图像模型包括空间量化和灰度量化两方面内容,理论基础是采样定理。由于分散流数据的离散性和空间分布不均匀特征,不可能按理论指标要求建模,因此,必须根据实际情况建立分散流数字图像模型。

2.1 空间量化预处理

采用六幅 1:50 000 的 Cu、Pb、Zn、W、Mo、Sn、Ag、Hg 元素分散流平面数据作为本次分散流图像处理的基本数据。数据区为一个边长为 28.5 km 的正方形区域,见图 1。为将非规则分散流数据转化为栅格数据,将正方形区域按 0.5 km $\times$ 0.5 km 网度划分正方形网格。网格的 $\Delta x, \Delta y$ 分别与正方形数据区的 X、Y 轴平行。共得到 58 $\times$ 58 个正交结点。这些正交结点就是空间量化插值点。采用综合插值法对这些结点插值,获得 58 $\times$ 58 个分散流网格化数据。按数字图像模型次序将这些数据输入计算机,建立起各元素的数字图像模型文件。

2.2 灰度量化预处理

灰度量化预处理的技术需求来自图像处理计算机软硬件的限制。由于本次工作以航天部五院的 S600-M75 数字图像处理系统为依托。该系统的显示存储器字长为 8 字节(bit),因此,只能模拟显示 0~255 范围内的正整型数。但实测地球化学数据值一般都超出这一范围,且为实型数,因此,必须将实测化探图像数据转换为 0~255 范围内的正整型数,即所谓的“归 255 化”。这实际上是一种实型数转换为字节(byte)型数的数型变换。采用的变换公式为:

$$\hat{f}(m, n) = \text{INT} [(f(m, n) - f_{\min}) / (f_{\max} -$$

$$f_{\min})] \times 255 \quad (12)$$

式中 $f(m, n)$ 为转换前 (m, n) 点的地化元素值, $\hat{f}(m, n)$ 为转换后 (m, n) 点的地化元素值, $f_{\max}$ 和 $f_{\min}$ 分别为图像区域内实测地化元素数据的极大、极小值, INT 表示取整运算。(12)式的运用是地球化学数字图像处理的关键。除此而外,地球化学元素数字图像建模还涉及其它一些重要的图像量化细节,因篇幅有限,在此不赘,请参阅参考文献^[9]。

3 图像处理

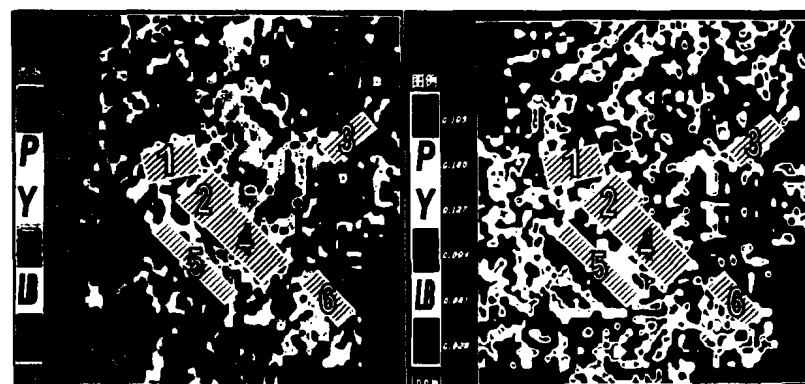
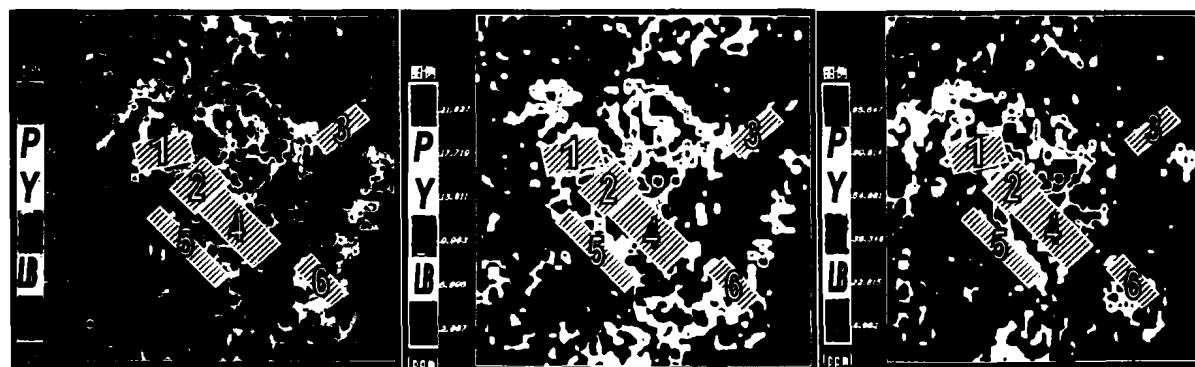
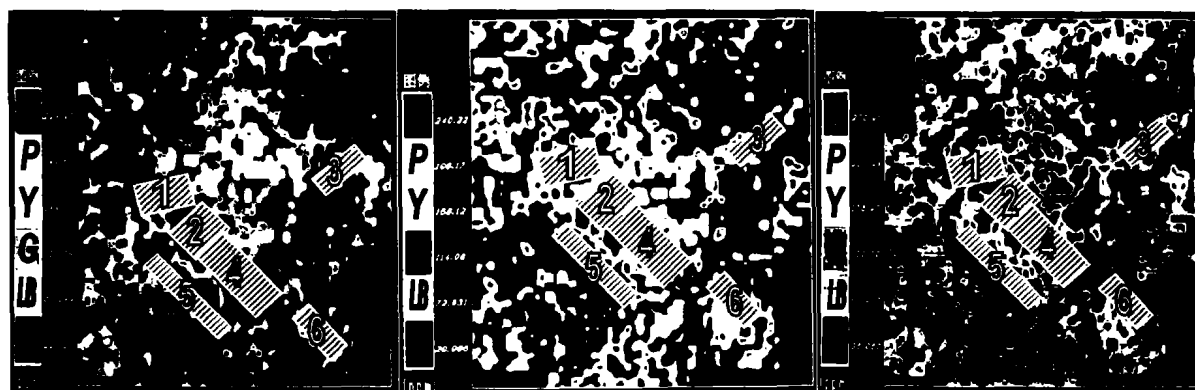
3.1 单元素图像彩色编码

目的在于为 KL 变换处理解译和效果评价提供必要的约束比较背景。由于 KL 变换的主分量图像的高度综合性, $g_1(x, y), g_2(x, y), \cdots, g_n(x, y)$ 图像反映的综合异常形态特征相对于单元素异常变化很大,以至于面目全非。为减少 KL 变换主分量图像解译的多解性,作为综合图像解译的配套措施,同时对地球化学元素的原始分布面貌作了研究。为此,对 8 种元素的分散流数据作了彩色编码增强显示处理,见图像 1~8。但需要说明,由于技术原因,非常遗憾,这些彩色编码图像在文章中只能以黑白制版。为体现原有的彩色异常分级关系,本文以大写字母 R(表示原红色编码)、P(表示原紫色编码)、Y(表示原黄色编码)、G(表示原绿色编码)、LB(表示原浅蓝色编码)和 B(表示原蓝色编码),分别记在各单变量元素彩色编码图像的相应彩色级图例位置上。由于标注于各单元素图像图例旁的分级数据仍保留着,因此,通过这种方式,可以在一定程度上弥补因不能彩色制版对异常场彩色定量分级显示所造成的不足。

3.2 KL 变换处理

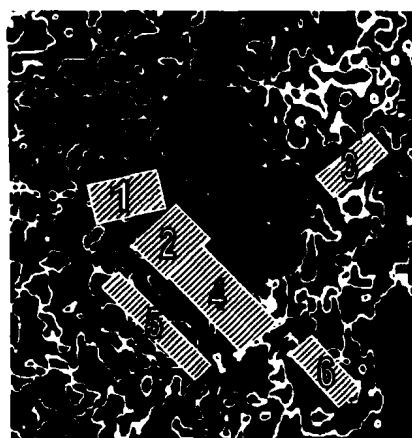
为比较 KL 变换对不同元素组合的找矿预测应用效果,分别对 5 种元素组合的分散流数据作了 KL 变换处理。这 5 种元素组合是:(Cu Pb Zn W Mo Sn Ag Hg)、(Cu Pb Zn)、(W Mo Sn)、(Pb Zn Ag)及(W Sn)。比较表明,(Cu Pb Zn W Mo Sn Ag Hg)组合的信息综合程度最高,找矿预测应用效果最好。其 KL 变换特征值和特征向量见表 1。

表 1 显示,前 3 个特征值占了综合图像信息总量的 87.7%。因此,利用其特征向量的元素线性组合,基本上可以反映 8 个元素信息综合总体分布的情况。对应于(Cu Pb Zn W Mo Sn Ag Hg)组合 KL 变换前 3 个特征值的 3 个主分量为:

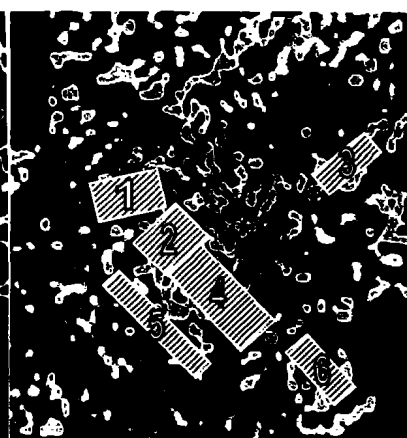


图像1-Cu	图像2-Pb	图像3-Zn
图像4-W	图像5-Mo	图像6-Sn
图像7-Ag	图像8-Hg	

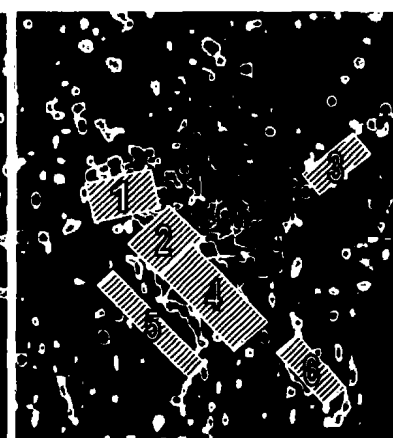
单元素分散流图。斜划线框为找矿预测区,框中数字为找矿预测区编号(下同)。



图像9 八元素分散流组合KL
变换第一主因子 g_1



图像10 八元素分散流组合KL
变换第二主因子 g_2



图像11 八元素分散流组合KL
变换第三主因子 g_3

表 1 分散流 (Cu Pb Zn W Mo Sn Ag Hg) 组合 KL 变换特征值及特征向量

特征值	特征向量 A_{ij}							
0.567	0.832	-0.390	-0.392	-0.011	-0.044	-0.006	-0.001	-0.001
0.185	-0.257	-0.896	0.337	0.119	0.060	0.005	0.000	0.001
0.125	-0.489	-0.188	-0.848	-0.074	0.010	0.007	0.001	0.002
0.085	0.018	0.096	-0.113	0.960	0.238	0.019	0.001	-0.000
0.035	-0.054	-0.016	0.002	0.243	-0.968	-0.012	-0.000	0.001
0.003	0.002	-0.001	-0.007	-0.017	-0.017	1.000	-0.003	-0.002
0.000	0.001	0.000	0.000	10.000	-0.000	0.003	0.999	0.040
0.000	0.002	0.001	0.001	0.000	0.001	0.002	-0.040	0.999

(根据广东省地质局 940 队 1981 年实测分散流数据计算)。

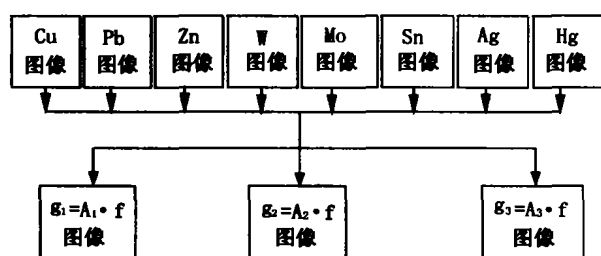


图 2 KL 变换图像处理流程框图

Cu、Pb、Zn、W、Mo、Sn、Ag、Hg 八种地球化学元素分散流异常图像 87.7% 的信息,通过 KL 变换,被综合(压缩)到三幅图像 g_1 、 g_2 和 g_3 上,见照片图

$$g_1(x, y) = 0.832\text{Cu} - 0.39\text{Pb} - 0.392\text{Zn} - 0.011\text{W} - 0.044\text{Mo} - 0.006\text{Sn} - 0.001\text{Ag} - 0.001\text{Hg},$$

$$g_2(x, y) = -0.257\text{Cu} - 0.896\text{Pb} + 0.337\text{Zn} + 0.119\text{W} + 0.060\text{Mo} + 0.005\text{Sn} + 0.000\text{Ag} + 0.001\text{Hg},$$

$$g_3(x, y) = -0.489\text{Cu} - 0.188\text{Pb} - 0.848\text{Zn} - 0.074\text{W} + 0.010\text{Mo} + 0.007\text{Sn} + 0.001\text{Ag} + 0.002\text{Hg}.$$

该 3 个主分量的彩色编码见图像 9 ~ 11。

4 找矿预测

4.1 图像分析

第一主因子图像 g_1 (图像 9): 以大片连续不规则弧形红色(黑白制版后为灰色)区域与大片连续不规则弧形蓝色区域相间分布为主要特征。大片红、蓝弧形区域互为间隔,构成一种从图区中心向四周扩散的同心环状结构,形态有点类似一锅正在煮开的稠粥的表面的花纹。石录、锡山、十二排、小南山和鸛鹑岭等已知矿床均处于同心环状结构的中心层圈地带。随着远离同心环中心,同心环状结构逐渐消失,红色异常区域范围扩大。这种特征反映出本区存在一个由已知矿床群构成的成矿中心,围绕该中心的成矿元素相关性分布具有某种空间韵律。但

是,由于成矿矿化与具有高的成矿元素相关性背景(红色)之间并无必然的可逆对应关系,因此,还不能够简单地以 g_1 红色异常作为找矿预测的指示。事实上,在同心环区域外围,还分布着更大面积的红色区域,可这些地方并无已知矿床,也不具备类似于已知矿床区的成矿地质条件。因此,实际情况也许要复杂得多。尽管如此, g_1 提供的信息还是对区域成矿规律研究和总体找矿布局提供了某种参考。

第二主因子图像 g_2 (图像 10): 以相对均匀但孤立分布的不规则红色(黑白制版后为灰色)综合异常小区为主要特征。红色小区异常分布范围不仅包括了石录、锡山、十二排、小南山和鸛鹑岭等已知矿床,而且还在图区 NE 方向有大面积超出,在图区 SW 和 NW 方向有小面积超出。根据这种情况,以及 g_2 异常规模小且零乱的特征,推测 g_2 主要是对成矿元素浅部相关性背景的反映。这种浅部相关信息含有较多不确定干扰因素,必须将其剔除后才便于利用。一种考虑是将 g_2 异常与成矿元素彩色编码增强图像作 overlay(叠加)分析,其重叠部分可以作为确定浅部矿化区的依据。

第三主因子图像 g_3 (图像 11): 以总体呈扁椭圆形的红色(黑白制版后为灰色)综合异常群为主要特征。扁椭圆区域长轴方向 NW。单个红色综合异常形态极不规则。与 g_2 图像比较, g_3 异常分布区更为集中于 Cu、Pb、Zn、W、Mo、Sn、Ag 彩色编码异常叠加区域内,并且与石录、锡山、十二排、小南山和鸛鹑岭等已知矿床存在更加密切的空间联系。此外,异常区之外的背景比较“干净”。这些图像特征表明,在 g_3 组合的信息提取水平上,浅部干扰异常信息已大部被“过滤”掉,“剩余”的 g_3 是对成矿元素深部相关性分布特征的集中反映,因此,其异常区域可以作为圈定隐伏矿床预测区的重要依据。

4.2 预测

在图像分析基础上,结合地质资料进行综合解译,共确定了6个找矿预测区,分别编号为1-6,见图1。其中,1、3、6号预测区为本次工作新确定的找矿预测区。从位置及异常规模看,由于1号与6号预测区处于侵入岩体内外接触带两侧的成矿有利地段,因此,这里有可能找到中型规模的接触变质型多金属矿床。3号预测区既远离花岗岩体,又远离已知矿床,处于泥盆纪桂头组(D_{2g})地层中,其成矿背景与花岗岩无直接关系,但可能与中低温热液活动有关,因此,推测这里可能存在中—小型的多金属矿床。至于2、4、5号预测区,均位于本区已知成矿活动最强烈的地段,因此,这里可以通过“就矿找矿”,寻找与石录铜矿床类型相近的中型规模矿床。

5 小结

通过应用,有两点值得总结:第一,不同的数量的元素变量组合对KL变换信息提取效果有很大影响,因此,在条件允许情况下,拟采用尽可能多的成矿元素变量组合;第二,单纯采用KL变换信息作找矿预测存在多解性问题,克服办法:对参与KL变换的各元素分别作彩色编码增强显示处理,以此作为KL图像解译的约束背景。

随着“地学数据仓库”^[10]和“地球化学数据库”^[11]步入实际应用阶段,如何快速有效地综合评

价大量多源地学—地球化学勘查数据资料,成为日益突出的问题。本文研究表明,基于KL变换的地球化学勘查数据发掘技术可以作为解决这一问题的有效手段。

[参考文献]

- [1] Stan Aronoff, Wayne Goodfellow. 图像处理对区域地球化学数据的有效分析[A]. 综合地学信息图像处理译文集[C]. 成都:四川高校编辑出版发行中心出版,1987,12,16~23.
- [2] 吴 虹.“三隐”预测的信息可视化提取[J]. 桂林冶金地质学院学报,1996,16(2):129~135.
- [3] 吴 虹.信息可视化提取在“二轮找矿”中的应用——以内蒙古狼山地区为例[J]. 桂林工学院学报,1997,17(1):34~39.
- [4] 李铁芳,冯均必,苏民生.遥感图像数字处理原理与应用[M]. 昆明:云南科技出版社,1987,61~166.
- [5] R C 冈萨雷斯, P 温茨. 数字图像处理[M]. 北京:科学出版社,1983,123.
- [6] 吴 虹.矿产勘查数字图像处理学初探[A]. 91北京计算机在地学中的应用讨论会论文集[C]. 北京:地质出版社,1991,790~792.
- [7] 陈春香.数据发现在地球化学数据处理中的应用[J]. 桂林工学院学报,1999,19,(2):230~233.
- [8] 史忠植.知识发现[M]. 北京:清华大学出版社,2002,1,2~4.
- [9] 吴 虹.位场数字图像模拟显示的量化技术[J]. 桂林冶金地质学院学报,1991,11(8):287~297.
- [10] 李振华,胡光道,王淑华.一个地学数据仓库的初步设计与实现[J]. 地质与勘探,2002,(5):67~70.
- [11] 史长久,张金华.全国主要大型金铜多金属矿区地球化学数据库[J]. 地质与勘探,2001,(3):41~44.

APPLICATION OF KL TRANSFORMATION OF DISPERSION FLOW DATA MINING IN THE YANGCHUN - SHILU AREA, GUANGDONG PROVINCE

WU Hong, CHEN Chun - xiang

(Guilin Institute of Technology, Guilin 541004)

Abstract: During the second prospecting round in the Yanchun - Shilu area, Guangdong province, geochemical visible analysis to interpret dispersion flow data of mineralizing elements like Cu, Pb, Zn, W, Mo, Sn, Ag and Hg of 1:50000 scale in an area of 812km² is implemented through using KL transformation technology. Through extracting display characters of the first, second and third main factor of image vectors of eight mineralizing elements and combining single element color - code images and correlative geological data, some valuable information for the hidden deposits, which is too difficult to be found them depending on the routine methods of geochemistry data processing in the past, had been got. On this base, 6 new prognosticating areas for prospecting hidden deposits had been put forward.

Key words: geochemical visible analysis, KL transformation, data mining, Yangchun - Shilu area