第9卷 第4期 1987年12月 物化探计算技术 Vol.9 No.4 December 1987
COMPUTING TECHNIQUES FOR GEOPHYSICAL AND GEOCHEMICAL EXPLORATION

化探数据中特高值的确定及其处理方法

廉龙洙

（黑龙江省地质三队）

一、二项系数加权游动平均值的计算方法及原则

本文将提出，在化探数据处理中运用二项系数加权游动平均法确定特高值及其处理的新方法，仅供讨论。

1．对于等间距分布的一维数列，每取相邻的三个点值为基本游动区间，以1，2，1的系数把下式作为基本公式，计算对应于中间点X_i值的加权游动平均值：

$$\hat{X}_i = (X_{i-1} + 2X_i + X_{i+1})/4 \qquad (1)$$

2．当计算对应于数列首尾两个端点值的加权游动平均值时，须把端点值X_i分别赋给公式（1）中X_{i-1}或X_{i+1}计算。

3．当游动区间的中间点为空白时，把区间内三个值全部赋给零，使其$\hat{X}_i=0$，并仍给予空白代码；如果区间内i－1或i＋1点为空白，则把中间的X_i值分别赋给X_{i-1}或X_{i+1}计算。

4．如此计算的加权游动平均值$\hat{X}_i$，须与原始数据X_i一一对应；且二者均值要相等：

$$\overline{X} = \overline{\hat{X}} \qquad (2)$$

5．对于成方格网分布的二维数阵，仍用公式（1）先按列计算一维加权游动平均值，在其基础上再按行计算一维加权游动平均值，就得相当于每取正方形九个网点为游动区间的二维加权游动平均值。

二、特高值的确定及处理方法

下面将通过实例说明化探数据的统计特征及特高值的确定及处理方法、步骤。

1．原始数据的统计特征

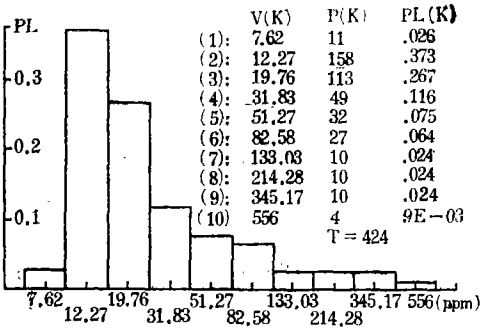
以某地土壤中砷元素的观测值为例子。观测点成22×20网点布列，点距100×100米，其中有效观测值424个，空白点16个。含量变化6～556ppm。

任何化探数据都是包含规律性基本变化和随机性变化的复杂的复合变量。它的概率曲线不一定就符合什么型式的分布函数。

收稿日期：1987年6月

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图1表示，上述424个原始数据的对数频率统计及其直方图。从图可知，并不符合通常所认为的对数正态分布，不能用正态分布函数来处理。



V(K)—组中值(真数ppm)；P(K)—频数；PL(K)—频率

图1．砷元素对数频率直方图

对于这类复合变量的变化大小，用其总方差表示。观测值 X_i 对样本均值 $\overline{X}$ 的离差 $\Delta_i=X_i-\overline{X}$，叫总离差。总离差的平方的平均值叫总方差，用 D_X 表示：

$$D_X=\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N} \qquad D_X=\overline{X^2}-\overline{X}^2 \qquad (3)$$

用（3）式计算总方差：

$$D_X=8009-44.49^2=6030$$

总方差的平方根就是标准差。这个标准差达均值的1.75倍，足以说明原始数据变化很大，有可能存在特高值和异常值。

2．二项系数加权游动平均值的统计特征

对于原始数据，按上述方法进行二项系数加权游动平均，可得新的与原始数据一一对应的加权游动平均值 $\hat{X}_i$。这个游动平均值，代表以九个网点的区间为单位滤去随机性变化而显现出来的规律性变化。这种规律性变化，用其基本方差表示。

加权游动平均值 $\hat{X}_i$ 对其均值 $\overline{\hat{X}}$ 的离差 $d_i=\hat{X}_i-\overline{\hat{X}}$，叫基本离差。基本离差的平方的平均值，称为基本方差，用G表示：

$$G=\frac{\sum_{i=1}^{N}(\hat{X}_i-\overline{X})^2}{N} \qquad G=\overline{\hat{X}^2}-\overline{X}^2 \qquad (4)$$

根据（4）式算得基本方差：

$$G=5167-44.49^2=3188$$

这个基本方差与总方差的比值 G/D_X，表明原始数据总变化中53%作为规律性变化呈现出来，其余47%的变化为随机性变化被滤掉。被滤掉的随机性变化如此之大，正是原始数据中特高值或异常值存在的反映。

3．随机离差的统计特征

原始观测值与二项系数加权游动平均值之差 $\delta_i=X_i-\hat{X}_i$，叫随机离差。

图2表示，424个随机离差值的频率分布情况，从图2可知，随机离差的绝大部分基本上服从正态分布，且少数高值呈现异常。

随机离差的变化大小，用其简单随机方差表示。随机离差的平方的平均值，叫简单随机方差，用 $D\delta$ 表示：

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$D\delta=\frac{\sum_{i=1}^{N}(X_i-\hat{X}_i)^2}{N} \tag{5}$$

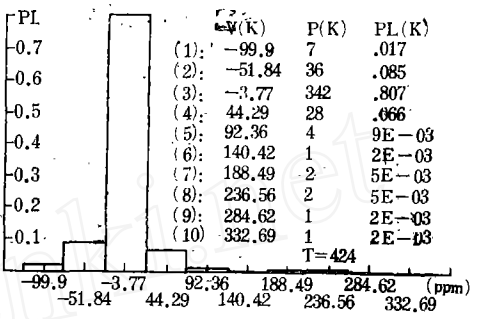
根据（5）式算得简单随机方差

$$D\delta=1556$$

因$\overline{X}=\overline{\hat{X}_i}$，随机离差的均值必定是：

$$\overline{\delta}=0 \tag{6}$$

对于原始数据中特高值或异常值的存在与否，最敏感的是这个随机离差。因此，其简单随机方差可作为确定特高值的主要指标。



V(K)—组中值；P(K)—频数；PL(K)—频率

图2　砷元素随机离差频率直方图

4．特高值的确定

上述随机离差δ_i，是均值为零且服从正态分布的变量。因此用正态分布函数可确定随机离差的异常值。对于正态分布变量中的异常，通常用“均值加三倍标准差”确定。为了和一般异常相区别，笔者认为用六倍标准差确定特高异常为宜。

简单随机方差$D\delta$的平方根，就是随机离差的标准差：

$$S=\sqrt{D\delta} \tag{7}$$

因$D\delta=1556$，所以$S=39.45$

六倍标准差　$P=6S=236.68$

整个424个随机离差中，大于236.68者只有333和296两个，也就是图2的最后两组。这是随机离差的特高异常。

与这个随机离差的特高异常值相对应的原始数据是532和413，这就是所要确定的原始化探数据的特高值。

5．特高值的处理

特高值的处理，应尊重它比相邻样品值具有略高的信息。为此笔者主张用代表该点的规律性变化的二项系数加权游动平均值，代替特高值的办法处理为合理。

对应于特高值532和413的加权游动平均值分别为199和117，这就是特高值的校正值。

特高值的确定和处理，可用如下模型表示：

随机离差＝观测值－游动平均值

$$\delta_i = X_i - \hat{X}_i$$

大于$6\sqrt{D\delta}$则特高值 ← 校正值

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

三、计算程序框图

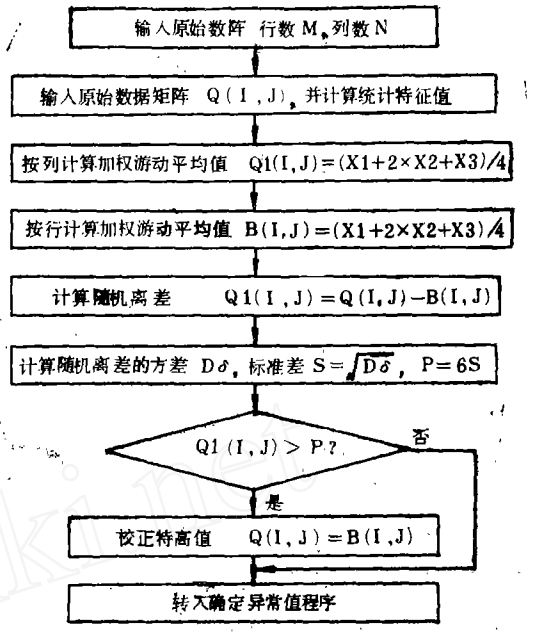


图3　程序框图

四、结　论

1．用本方法确定的特高值，不仅在整个样本中相对的较高，更重要的是以其随机离差最大为特征。

2．特高值的校正值，代表以该点为中心的小区间的规律性变化，且高于相邻样品值，被减少的仅仅是随机性变化，这样处理是合情合理的。

3．样品中最高者不一定都是特高值。如本例中最高含量是556ppm，其随机离差达标准差的五倍，但该点的游动平均值却低于相邻样品值。556对相邻样品值来说并不很高，所以不属特高值。

4．本例中所确定两个特高值经校正后，使整个样本的均值降2%，总方差降13.5%。可见特高值对样本统计特征的影响之大。

5．运用二项系数加权游动平均法，是以相邻样品之间具有较密切的相关性为前提。本例中原始数据与加权游动平均值的相关系数为0.87，它间接的反映相邻样品之间的相关程度。

对于点距较大的中小比例尺观测值来说，因观测点之间地质条件有可能不同，相关性不明显，所以不宜处理特高值，应直接圈定异常。

6．二项系数加权游动平均法，对下一步确定异常值也是比较理想的数学工具。只是把确定特高值的指标“六倍标准差”改为“均值＋3倍标准差”，校正特高值的过程改作剔出异常值，如此反复进行到再没有异常值为止。

参 考 文 献

〔1〕杨善慈，矿床地质变量统计理论和扬赤中滤波与推估，中南矿冶学院采矿系、矿床地质变量统计学科研组编，1984年6月。

THE METHOD OF TREATTING SUPER-VALUE POINTS IN GEOCHEMICAL PROSPECTING DATA

Lian Longzhu

(No.3 Geology Team of Heilongjiang Province)

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net