

# 基于统计学特性的高密度电阻率法多排列 数据拼接及分组误差校正

姚健, 曾昭发, 黄玲, 李芳, 郑圣谈, 张代国

(吉林大学地球探测科学与技术学院, 长春 130061)

**[摘要]** 根据高密度电阻率法相邻两个排列之间有相当重叠部分这个特点, 提出将高密度各个单一排列剖面连接组成一个多排列数据拼接一起并进行误差校正。为抑制相邻排列间数据系统误差影响, 运用多排列数据拼接系统误差的校正方法, 即应用有序聚类法对多排列重叠层分组进行误差校正的一种计算方法。结合实例说明该方法的实际应用效果, 并验证了其有效性。

**[关键词]** 多排列拼接 系统误差校正 有序聚类 逐步截尾 最小二乘法

**[中图分类号]** P631.3; P628 **[文献标识码]** A **[文章编号]** 0495-5331(2007)03-0078-05

对于单排列的高密度电阻率法测量结果, 由于测量仪器、测量时间以及电极布设等一致性较好, 因而单排列高密度电阻率法容易得到较好的效果。但在多排列高密度电阻率法的数据处理与反演中, 由于不同时间、不同人员测量或仪器误差等原因, 各排列数据之间存在较大的差异。对于多排列高密度电阻率法剖面连接和误差消除, 通常采用: ①丢弃一部分重叠区的数据, 拼接成长剖面; ②取重叠区数据均值, 拼接成长剖面。运用这两种方法并未使数据差异得到控制或排除, 这不但对解释推断带来影响, 而且影响反演结果图的美观, 对背景和异常的合理划分造成困难。因而需要对数据进行分析处理, 合理地消除误差, 得到更客观的结果。文章根据高密度电阻率法相邻排列之间有相当重叠部分这个特点, 提出将高密度各个单一排列的数据连接组成一个多排列数据拼接一起并进行误差校正。这样可以方便展示一个剖面上整体的测量结果, 方便工作人员的分析 and 解释推断。同时, 为抑制多排列间数据系统误差影响, 提出多排列数据拼接系统误差的校正方法, 即应用有序聚类法<sup>[1-2]</sup>对多排列重叠层分组进行误差校正的一种计算方法, 以解决各个排列误差校正的问题。在消除不同排列间的系统误差, 解决多排列数据连接问题, 具有较好的效果。

## 1 方法基础

### 1.1 不同排列数据存在差异的原因

不同排列的高密度电阻率测量数据有时存在较大的差异, 产生这种差异的原因主要有: ①测量仪器的不稳定性; ②因测量时间的差异而造成地表电阻率的变化; ③电极接地情况的不同; ④导线敷设问题; ⑤可能存在仪器或导线漏电现象等。由于这些问题的影响, 使得测量数据因非地质因素发生很大的变化。为了避免出现这种情况, 往往采用重新测量的方式来解决, 这样不仅浪费人力、财力, 甚至耽误施工进度, 而且不一定能将这些影响完全消除。采用多排列数据拼接系统误差的校正方法能够尽量避免重复测量, 又能更准确地提供地质信息。

### 1.2 系统误差影响及消除意义

系统误差在高密度电阻率法成图中不但会影响图面的美观, 甚至往往会掩盖或扭曲地球物理特征。在电阻率剖面图中, 校正前, 拼接各个排列在边界两侧有明显的“台阶”突变, 与相邻排列形成鲜明的对照, 所以必须进行系统误差校正。

可以从几个方面来考查不同排列电阻率差异特征:

从数据角度看, 假设有系统误差的同一排列的数据与无系统误差的真实数据存在线性关系。只要

[收稿日期] 2006-03-27; [修订日期] 2006-06-27; [责任编辑] 韩进国。

[基金项目] 中国地质调查局项目(20002010002106)资助。

[作者简介] 姚健(1981年—), 男, 2005年毕业于吉林大学, 获学士学位, 在读硕士生, 现主要从事工程与环境地球物理方法技术研究工作。

能够发现系统误差的存在,那么校正它的方法就是进行线性变换。

从地球物理角度看,有系统误差的某一区域内,并没有因系统误差而改变该区地球物理分布上的相对形态,只有当多排列拼接之后系统误差影响才会在排列拼接处出现“台阶”现象。

在地球物理成图中,有系统误差的区域会造成地球物理异常的丢失或放大。系统误差校正的目的方面是为多排列拼接消灭“台阶”,另一方面就是要将该区的地球物理异常尽量恢复到原来应有的程度。

### 1.3 多排列数据的统计学特性及应用基础

根据多排列重叠层数据的地球物理特征,对重叠层倒金字塔式数据结构采用最优化分组控制策略进行分组误差校正,可以避免因重叠区域各层实测数据差异过大,直接采用整个重叠区均值与均方差进行校正,校正后相邻两排列重叠区的数据,仍存在较大差异。因此,对重叠区分组校正可以大大提高误差校正精度。

在高密度电阻率法剖面中,浅层电阻率变化大,噪声大,反映浅部地质情况,深层则相反。由此可对高密度电阻率剖面的各层进行从浅到深分组,进而进行分组校正。对相邻排列重叠层有序聚类分组,关于有序聚类法的方法原理、计算公式及其计算过程,统计学和数学地质方面的各类专著及论文已详细叙述。该法用于解决按一定顺序排列的有序层数的分组统计问题,其原理为:设有  $N$  个数据剖面层,计算每层数据的均值和均方差,应用逐步截尾法对异点数据进行剔除,对剔除后剩余的数据再求均值和均方差,这样每层剩余数据的均值与均方差构成  $N$  个有序数据列,将其按顺序分割成数组,要求每组内离差平方和尽可能小,而组间离差平方和尽可能大。实际应用中,根据数据自身的特征进行分组,也可根据实际需要分割成所希望的组数,再分析每组数据的统计学特性。

以多排列重叠层分组进行系统误差校正的最小二乘算法<sup>[3]</sup>为例讨论。该方法以某排列分组后数据的背景均值和均方差为参数,使相邻两排列重叠区的均值和均方差相等或相近,以此校正不同排列的背景差异。多排列重叠层分组进行系统误差校正的最小二乘算法在消除系统误差解决接边问题的一个隐含条件是:处于同一地质体上排列相邻区带内的数据有相同的或相近的均值与均方差。

设某一数据在某一排列  $i$  的实测值  $XZ_i$  存在系

统误差,其校正值应为  $XC_i$ ,不失一般性,实测值与真值之间的关系为:

$$XC_i = a_i XZ_i + b_i, i = 1, 2, \dots, l \quad (1)$$

其中  $a_i, b_i$  为待定值,分别称为系统误差的放大系数和基台系数。若将实测值在排列  $i$  内分组后的均方差和均值依次记为  $SZ_i$  和  $\overline{XZ_i}$ ,则其重叠层分组后校正值的均方差和均值为  $SC_i$  和  $\overline{XC_i}$ ,则:

$$SC_i = a_i SZ_i, i = 1, 2, \dots, l \quad (2)$$

$$\overline{XC_i} = a_i \overline{XZ_i} + b_i, i = 1, 2, \dots, l \quad (3)$$

以记号  $y_{k,i/j}$  表示变量  $y$  在排列  $i, j$  交界处  $i$  排列一侧重叠层第  $k$  组的校正值,简称  $y$  的重叠层分组表达式。例如  $SC_{k,i/j}$  表示排列  $i, j$  交界处  $i$  排列一侧第  $k$  组重叠层分组校正值的均方差,即此均方差的

重叠层分组表达式。除随机因素外,在相邻两排列数据的总差异中包含了由地质因素和非地质因素引起的两大差异,在消除系统误差时不应将它们一概消除,根据高密度电阻率法野外实际测量工作测线布设方式的特点,相邻两条高密度测线上有很大的重叠部分。在排列  $i, j$  重叠区域内,地质情况完全相同,其重叠的两部分数据的差异可视为完全由非地质因素引起,属于消除对象。据此显然应有:

$$SC_{k,i/j} - SC_{k,j/i} = e \quad (4)$$

$$\overline{XC_{k,i/j}} - \overline{XC_{k,j/i}} = d \quad (5)$$

相邻两排列高密度反演剖面结果图的重叠部分校正后可能仍然存在误差,其中  $e, d$  为校正后仍然可能存在的误差。由于重叠部分的地质情况完全相同,所以理论上  $e, d$  也应趋近于零。由于现实原因,不可能达到这种极限状态,因此,在高密度多排列拼接中,进行系统误差校正力求使上述  $e^2, d^2$  达到极小,以便使校正后重叠部分的地球物理参数与地质情况相吻合,达到基本相同状态。

由于实测数据中往往存在一些坏点(其值明显高于或低于剖面平均值),对整体均值和均方差影响非常大,因此必须进行坏点剔除后再进行误差校正,这样在误差校正中使用的均值和均方差才是合理的。对于异点剔除,常采用逐步截尾法。实际测量数据分布的常见情形是接近正态,但往往由于少数异点的存在出现一个长尾巴。逐步截尾法就是按一定的准则将长尾巴截去后用剩下的接近正态分布的数据来计算背景分布参数的统计方法。

## 2 基于统计学特性的多排列数据重叠层分组误差校正的最小二乘算法<sup>[4]</sup>

以相邻4个排列为例,给出多排列系统重叠层

分组误差校正的最小二乘算法。设此相邻 4 个排列依直线形式排开,依次标以 1、2、3、4 等编号,不同排列的数据存在系统误差,以排列 1 为基准进行校正(即令  $a_1 = 1, b_1 = 0$ )。对 1、2 排列,2、3 排列,3、4 排列重叠层分别进行分组,依据实际地质体情况与有序聚类法,假设对 1、2 排列重叠层分为 3 组,对 2、3 排列重叠层分为 4 组,对 3、4 排列重叠层分为 3 组,这时仿照式(4)可得方程组:

$$\begin{aligned} SC_{1,1/2} - SC_{1,2/1} &= e_1, SC_{2,1/2} - SC_{2,2/1} = e_2, \\ SC_{3,1/2} - SC_{3,2/1} &= e_3, SC_{1,2/3} - SC_{1,3/2} = e_4, \\ \dots\dots\dots, SC_{3,3/4} - SC_{3,4/3} &= e_{10} \end{aligned}$$

其中  $e_1, e_2, e_3, \dots, e_{10}$  为校正后依然存在的误差。将其他各变量参照(2)式的表达式代入,并令  $a_1 = 1$ ,经整理后可得下列矩阵形式的方程组:

$$SA - C = E \quad (6)$$

其中  $A = (a_2, a_3, a_4)^T$  为待求解向量, $S$  为 10 行 3 列的系数矩阵,其矩阵的表达式为:

$$S = \begin{pmatrix} -SC_{1,2/1} & 0 & 0 \\ -SC_{2,2/1} & 0 & 0 \\ -SC_{3,2/1} & 0 & 0 \\ SC_{1,2/3} & -SC_{1,3/2} & 0 \\ SC_{2,2/3} & -SC_{2,3/2} & 0 \\ SC_{3,2/3} & -SC_{3,3/2} & 0 \\ SC_{4,2/3} & -SC_{4,3/2} & 0 \\ 0 & SC_{1,3/4} & -SC_{1,4/3} \\ 0 & SC_{2,3/4} & -SC_{2,4/3} \\ 0 & SC_{3,3/4} & -SC_{3,4/3} \end{pmatrix},$$

$C = (-SC_{1,1/2}, -SC_{2,1/2}, -SC_{3,1/2}, 0, \dots, 0)^T$  为 10 阶常向量, $E = (e_1, e_2, e_3, \dots, e_{10})^T$ 。按最小二乘准则求这样的  $A$ ,使  $Q = E^T E = (SA - C)^T (SA - C)$  达到极小,于是由  $\partial Q / \partial A = 0$  可得方程组:

$$S^T SA = S^T C \quad (7)$$

其解为:

$$A = (S^T S)^{-1} S^T C \quad (8)$$

对于求基台系数  $b$ ,仿照(5)式可得如下方程组:

$$\begin{aligned} \overline{XC}_{1/2} - \overline{XC}_{2/1} &= d_1, \overline{XC}_{2/3} - \overline{XC}_{3/2} = d_2, \\ \overline{XC}_{3/4} - \overline{XC}_{4/3} &= d_3 \end{aligned}$$

其中  $d_1, d_2, d_3$  为校正后仍然存在的误差。将其他各变量参照(3)式的表达式代入,再将式(8)的解向量  $A$  中各分量代入,且令  $a_1 = 1, b_1 = 0$ ,经整理后可得下列矩阵形式的方程组:

$$HB - P = D \quad (9)$$

其中  $B = (b_2, b_3, b_4)^T$  为待求解向量, $H$  为 3 行 3 列的系数矩阵, $P$  为 3 阶常向量, $D = (d_1, d_2, d_3)^T$ 。同样可求这样的  $B$ ,使  $R = D^T D = (HB - P)^T (HB - P)$  达到极小,于是由  $\partial R / \partial B = 0$  可得方程组:

$$H^T HB = H^T P \quad (10)$$

其解为:

$$B = (H^T H)^{-1} H^T P \quad (11)$$

将解向量  $A, B$  中各分量代入(1)式,便可得出其他各排列相对于排列 1 的校正值。当然,还可指定其中某排列  $i$  为基准,令  $a_i = 1, b_i = 0$  求解方程组,对其他排列进行校正。上面列举了 4 个排列重叠层分组的系统误差校正的最小二乘算法。当遇到大量(多于 4 个)排列拼接时,同样可以应用这个基本原理建立方程组来进行误差校正。

### 3 应用实例和效果

以延吉某区域高密度电阻率法探测为例,探测采用温纳四级装置,电极间距为 5m,共 60 个电极,共完成 16 个隔离系数的测量。应用上述方法进行校正,具体步骤如下:

- 1) 提取各个相邻排列重叠区各层的数据,并确保提取的数据准确无误。
- 2) 以每个排列重叠部分各层为单元,运用逐步截尾法剔除其中异常点,并求取剔除异点后剩余数据的均值和均方差。
- 3) 应用有序聚类法对重叠层进行分组,此例中对 1、2 排列重叠层分为 3 组,对 2、3 排列重叠层分为 4 组,对 3、4 排列重叠层分为 3 组,注意相邻排列分组个数相同,且每组所含层数相同。当然也可根据实际地质情况进行人为分组,在遵循上述原则下,分组越多校正结果越精确,但计算量随之增大。
- 4) 对分组后的各组数据应用逐步截尾法进行异点剔除,求出剔除异点后剩余数据的均值与均方差,并依据相邻排列重叠层分组后具有相同的均值和均方差这一基本原则建立方程组。
- 5) 利用科学计算软件 MATLAB 求解方程组。在求解过程中,设定其中某一系列  $i$  为准系数  $a_i = 1, b_i = 0$ ,即以这一排列为基准进行误差校正。
- 6) 将求得的解代入(1)式,对其他排列进行校正。
- 7) 再次提取多排列校正后重叠区域数据,对相邻两个排列重叠区域对应点数据求取均值,作为重叠区域数据校正后的最终值。
- 8) 将校正后的各个排列的数据代入高密度电

阻率法数据处理软件进行成像成图。

校正前后的地电模型正演计算结果(图1,图2)对比可发现,校正前拼接的高密度多排列正演计算结果在拼接的重叠部分有着明显的“边界”,“边界”两侧有明显的突变,这样“边界”圈定的区域就

变成“异常”,给人以在拼接的地方就是异常出现分界的地方这个假象;而且剖面结构整体的背景值区域混乱,存在明显的畸变,结果图反映地质体的大小及分布都与实际情况相差甚远,这样就无法真实地反映地质信息。校正后的结果表现为整体的背景值

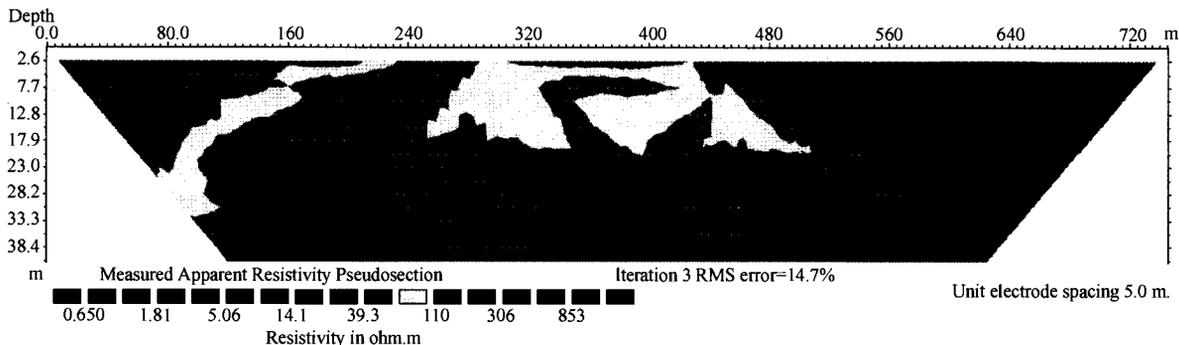


图1 校正前的多排列正演结果图

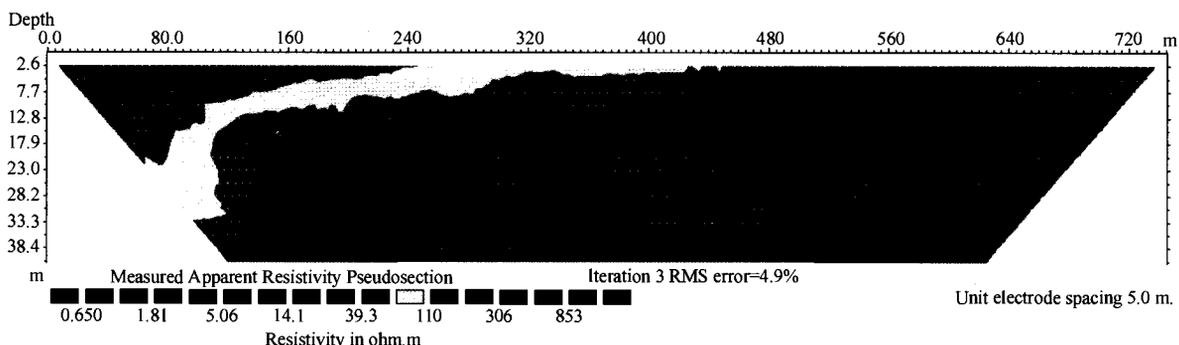


图2 以排列1为基准重叠层分组校正后的多排列正演结果图

区域近似单一,有比较圆滑的分层,高阻低阻区边界清晰。测线上位于+260m和+520m处分别布置两个平均深度为15m的勘探孔。根据钻孔取芯显示,+260m处样本自上而下为:①夹有较多的碎石、煤渣等杂物的填土,视电阻率一般为 $80 \sim 120 \Omega \cdot m$ ,在 $0 \sim 8.5m$ 处;②粉质粘土,视电阻率一般为 $40 \sim 80 \Omega \cdot m$ ,在 $8.5 \sim 12m$ 处;③淤泥质粉质粘土,视电阻率一般为 $20 \sim 40 \Omega \cdot m$ ,在 $12m$ 以下。+520m处样本取样结果为单一的泥质粉质粘土,视电阻率为

$20 \sim 35 \Omega \cdot m$ ,取样结果显示其为单一介质。多排列拼接并进行误差校正后的计算结果图与钻孔资料进行对照吻合很好,进一步证明了该方法的准确性。通过校正前后相邻两排列(排列1和排列2)重叠部分数据曲线图(图3)对比可以看出,校正前相邻两排列重叠部分两组数据差异较大,理论上这是不应该存在的,显然是由误差引起;而校正后相邻两排列重叠部分两组数据吻合较好,说明应用上述方法进行校正后能够有效地消除误差。综合上述几幅图可

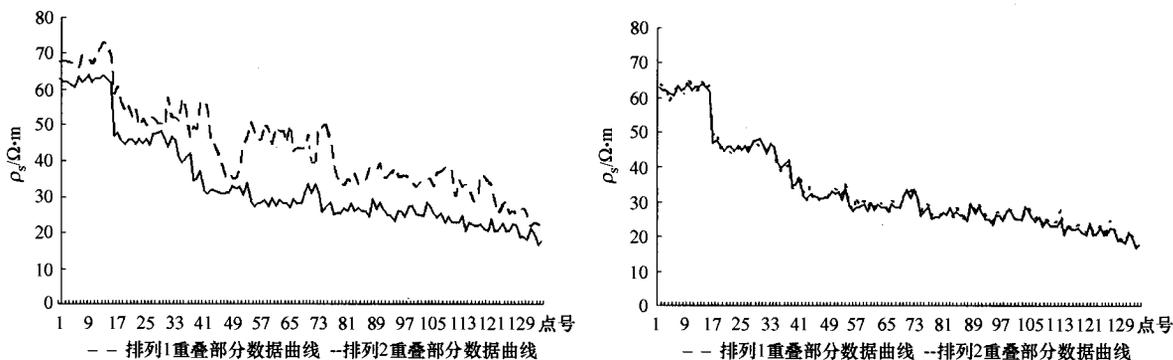


图3 相邻两排列重叠部分数据校正前后曲线对比图

直观地看出校正后排列间的系统差异明显消除,效果良好,得到校正的高密度多排列连接图能够清晰、直观、准确地反映地电断面电阻率分布状况和地质情况,使地质工作人员更为方便地进行分析和解释。

#### 4 结 语

在许多情况下,不同排列间的系统误差可以通过应用有序聚类法对多排列重叠层进行分组,并依据分组情况运用最小二乘法进行误差校正来消除。

运用重叠层分组法进行误差校正的显著优点是分组后组内数据根据实际地质情况取值比较集中,而且由于实际地质体在某一层位上视电阻率差异不大,可以避免因异常和各层导电性差异过大造成数据取值分散,给校正结果带来很大的影响,同时也大大提高了校正的精度。该方法具有较好地处理接边问题的效果,有效地避免了在接边处出现“伪异常”,得到校正的结果接近实际地质情况,提高资料解释的准确性,同时在一定程度上提高了工作效率。

在实际应用中,可以结合钻孔资料进行对照,

并根据钻孔资料来选取基准排列进行误差校正,也可具体结合实际的地质情况和数据质量,灵活地使用上述方法。

#### [参考文献]

- [1] 赵旭东. 石油数学地质概论[M]. 北京:石油工业出版社, 1992.
- [2] 韩金炎. 数学地质[M]. 北京:煤炭工业出版社, 1987.
- [3] 冯 康. 数值计算方法[M]. 北京:国防工业出版社, 1978.
- [4] 纪宏金, 时艳香, 代永刚, 等. 基于边界极限原理的多图幅系统误差的校正方法[J]. 物探化探计算技术, 2005, 27(2): 154 - 155.
- [5] 田 钢, 刘菁华, 曾昭发. 环境地球物理学教程[M]. 北京:地质出版社, 2005.
- [6] Barker R. A simple algorithm for electrical imaging of the subsurface [J]. FIRST BREAK, 1992, 10(2): 53 - 62.
- [7] Sasaki Y. Resolution of resistivity tomography inferred from numerical simulation[J]. Geophysical prospecting, 1992, 40(4): 453 - 463.
- [8] 陈永清, 纪宏金. 标准化区域地球化学图的编制方法及应用效果[J]. 长春地质学院学报, 1995, 25(2): 216 - 220.

## MULTIPLE ARRAY DATA COMBINATION AND CLUSTER ERROR CALIBRATION OF HIGH DENSITY RESISTIVITY METHOD BASED ON STATISTICS

YAO Jian, ZENG Zhao - fa, HUANG Ling, LI Fang, ZHENG Sheng - tan, ZHANG Dai - guo  
(College of Geo - Exploration Science and Technology, Jilin University, Changchun 130061)

**Abstract:** Based on overlapping characteristics of two adjacent arrays in the high density resistivity method, a method to combine arrays and remove error is proposed. To restrain system error of adjoining array data, order cluster method which can solve data combination of arrays systemic error correction based on overlapping packets is applied. This method, proved by a real example, has practice application effect and effectiveness.

**Key words:** multiple array combination, system error calibration, order cluster, stepping truncation, least square method