

文章编号: 1001-1986(2006)02-0058-04

## 系统聚类逐步判别法对皖北矿区突水水源的分析

殷晓曦<sup>1</sup>, 许光泉<sup>1</sup>, 桂和荣<sup>1</sup>, 陈陆望<sup>2</sup>

(1. 安徽理工大学资源与环境工程系, 安徽 淮南 232001;

2. 中国科学院武汉岩土力学研究所, 湖北 武汉 430071)

**摘要:**以  $\text{Ca}^{2+}$ 、 $\text{Mg}^{2+}$ 、 $\text{K}^{+} + \text{Na}^{+}$ 、 $\text{Cl}^{-}$ 、 $\text{SO}_4^{2-}$ 、 $\text{CO}_3^{2-}$ 、 $\text{HCO}_3^{-}$ 、pH 值与 TDS 的测试结果为指标, 对皖北矿区四舍、煤系、太灰、奥灰 4 个主要突水含水层系统分别进行聚类分析。在此基础上, 筛选出能代表各含水层水化学特征的水样进行逐步判别分析, 并从中选出了  $\text{SO}_4^{2-}$  与 TDS 两种标型组分, 建立了皖北矿区突水水源的判别模型, 该模型具有较好的判别效果。

**关键词:**水化学; 系统聚类; 逐步判别; 突水水源

**中图分类号:** P641.461 **文献标识码:** A

### Analyzing for sources of inrush-water in Wanbei Mining Area by systemic clustering and stepwise distinguishing

YIN Xiao-xi<sup>1</sup>, XU Guang-quan<sup>1</sup>, GUI He-rong<sup>1</sup>, CHEN Lu-wang<sup>2</sup>

(1. Department of Resources and Environment Engineering, Anhui University of Science and Technology, Huainan 232001, China; 2. Institute of Rock and Soil Mechanics, The Chinese Academy of Sciences, Wuhan 430071, China)

**Abstract:** Four main aquifers of Wanbei Mining Area, such as the fourth aquifer of the Quarternary, the coal aquifer, the Taiyuan limestone aquifer and the Ordovician limestone aquifer, are analyzed by systemic clustering with the indexes of  $\text{Ca}^{2+}$ 、 $\text{Mg}^{2+}$ 、 $\text{K}^{+} + \text{Na}^{+}$ 、 $\text{Cl}^{-}$ 、 $\text{SO}_4^{2-}$ 、 $\text{CO}_3^{2-}$ 、 $\text{HCO}_3^{-}$ 、pH and TDS. From the clustering result, the water samples representing the characteristics of every aquifer are picked and then are analyzed by stepwise distinguishing. Therefore, a distinguishing model is established by the two representative elements of  $\text{SO}_4^{2-}$  and TDS, which has a good effect on distinguishing sources of inrush-water of Wanbei Mining Area.

**Key words:** hydrochemistry; systemic clustering; stepwise distinguishing; sources of inrush-water

## 1 引言

皖北矿区位于淮北平原的北部, 是华东地区重要的煤炭基地。皖北矿区分为宿县、濉萧和临涣 3 个分矿区。矿区内开采煤层的突水水源主要为第四系第四含水层(简称“四舍”)、煤系砂岩含水层(简称

“煤系”)、石炭系太原组灰岩含水层(简称“太灰”)和奥陶系灰岩含水层(简称“奥灰”)。这些含水层对矿区内受地下水严重威胁的矿井尤为重要。

突水水源判别是矿井防治水工作的基础, 许多专家与学者对皖北矿区进行过水文地球化学研究, 但这些研究只限于局部地区。成春奇等<sup>[1]</sup>通过分析

收稿日期: 2005-04-20

作者简介: 殷晓曦(1973—), 女, 安徽淮南人, 安徽理工大学在读硕士生, 现从事水化学及环境方面的研究。

b. 涂抹区范围的大小与挤土程度密切相关, 对塑料排水板和袋装砂井等不同竖井的涂抹区半径取值不加区别有欠妥当。

c. 涂抹区半径与竖井半径的比值不仅与深度相关, 也与土质、竖井类型, 乃至施工方法等密切相关。宜根据不同土质、竖井类型和施工方法进行分类细化, 以便更合理地选择。

## 参考文献

[1] 地基处理手册编写委员会. 地基处理手册. 第2版[M]. 北京:

中国建筑工业出版社, 2000: 55-141.

- [2] Vesic A S. Expansion of cavities in an infinite soil mass [J]. Journal of Soil Mechanics and Foundation. ASCE, 1972, 98(3): 265-291.
- [3] 樊良本, 朱国元. 桩周土应力状态的圆柱孔扩张理论试验研究 [J]. 浙江大学学报(工学版), 1998, 32(2): 228-235.
- [4] 陈文. 饱和粘土中静压桩沉桩机理及挤土效应研究 [D]. 南京: 河海大学岩土所, 1999.
- [5] 王幼青. 挤土桩水平向挤土位移分析 [J]. 哈尔滨工业大学学报, 2003, 35(4): 472-475.
- [6] 刘吉福, 杨春林. 珠江三角洲地区竖向排水体施工扰动初探 [J]. 岩石力学与工程学报, 2003, 22(1): 142-147.

比较临涣分矿区的四含、煤系、灰岩含水层3个地下水子系统的水化学特征,得出了子系统的常规离子水化学标型组分,为区域矿井水源判别提供了依据。成春奇等<sup>[2]</sup>对百善矿地下水系统的四含、煤系和灰岩含水层的常规水化学特征进行了系统研究,提出了各含水层的水质模型,给出了突水水源的化学判别模式。李定龙等<sup>[3]</sup>分析了临涣分矿区开采前、后四含的常规离子含量特征及其形成作用,提出了四含水化学成因模式。孙本魁等<sup>[4]</sup>根据任楼矿主要突水含水层常规离子的含量,找出了一种快速、有效、准确的水源判别方法。董春江等<sup>[5]</sup>对芦岭矿提出:某一出水点的水质类型如果与某两个含水层的水质类型均相似,则可选出含水层的标型组分建立判别函数,进而确定突水水源。

皖北矿区主要突水含水层水质常规指标不是孤立存在的,它们之间存在某些必然的联系。所以,使用水质常规指标通过系统聚类和逐步判别分析方法,建立矿区突水水源判别模型,对整个皖北、淮北矿区乃至永城矿区的防治水工作都具有一定意义。

## 2 水样及分析指标

利用矿区内各生产矿井长观孔、水源井和井下放水孔及井下出水点取样,力求各含水层取样点在矿区范围内均匀分布。把取自同一矿井同一含水层并且各测试指标几乎相同的一组水样取其平均值,当作一个采样点所得到的水样。矿区主要突水含水层水样点的测试指标和测试结果见表1。

表1 皖北矿区主要突水含水层水化学数据

Table 1 Hydrochemic data of the key intrushing aquifers in Wanbei Mining Area

水样 编号	取样 层位	采样 地点	Ca <sup>2+</sup> /mg·L <sup>-1</sup>	Mg <sup>2+</sup> /mg·L <sup>-1</sup>	K <sup>+</sup> + Na <sup>+</sup> /mg·L <sup>-1</sup>	Cl <sup>-</sup> /mg·L <sup>-1</sup>	SO <sub>4</sub> <sup>2-</sup> /mg·L <sup>-1</sup>	CO <sub>3</sub> <sup>2-</sup> /mg·L <sup>-1</sup>	HCO <sub>3</sub> <sup>-</sup> /mg·L <sup>-1</sup>	pH 值	TDS /mg·L <sup>-1</sup>
1		刘二	60.0	74.8	529.2	145.0	719.5	0.0	778.4	7.91	2 385
2*		刘一	25.3	24.1	50.4	20.5	748.0	0.0	15.4	7.92	1 970
3*	四	任楼	117.4	24.0	471.5	905.8	23.6	0.0	39.7	7.74	1 562
4*		海孜	110.2	59.6	53.8	124.1	181.6	15.6	289.9	8.18	690
5*	含	董亭	80.0	72.4	175.6	124.5	540.0	0.0	107.0	7.97	1 113
6*		桃园	186.8	74.9	136.0	301.3	212.3	0.0	488.2	8.12	1 155
7*		朱仙庄	19.6	15.8	120.5	31.9	43.2	0.0	323.4	8.23	393
8		芦岭	27.3	26.7	158.0	28.4	103.7	27.6	366.1	8.48	555
9*		刘二	58.1	18.0	332.2	230.4	175.8	0.0	457.7	8.36	1 043
10*		临涣	17.2	5.1	949.0	126.5	1 779.8	0.0	288.6	8.20	3 180
11		百善	168.3	95.1	397.1	200.3	196.9	0.0	1 363.8	7.88	1 740
12*	煤	董亭	117.2	62.7	856.6	195.9	1 843.6	0.0	267.5	7.90	3 223
13		海孜	103.3	51.1	118.5	161.3	213.3	12.0	286.8	8.32	803
14*	系	任楼	8.4	4.9	860.3	496.3	11.5	69.6	1 139.2	8.42	2 016
15*		朱仙庄	193.2	61.8	357.5	378.5	525.2	0.0	532.1	7.80	2 083
16		芦岭	237.3	188.7	388.7	650.0	1 833.0	0.0	194.6	7.64	3 200
17*		祁南	2.81	2.2	460.0	109.9	1.0	126.0	695.6	8.92	1 050
18*		朔里	5.0	13.1	314.4	55.0	13.8	36.0	163.5	9.04	314
19*		张庄	80.6	32.3	20.3	12.4	25.0	0	404.6	7.94	373
20*		杨庄	28.7	27.9	179.3	58.5	118.4	36.0	340.5	8.42	619
21*		刘二	331.6	120.6	326.2	200.0	1 447.2	0.0	307.8	7.80	2 796
22*	太	海孜	46.9	50.1	323.8	239.3	210.4	0.0	505.3	7.98	1 123
23		董亭	8.3	18.9	743.71	149.6	775.5	43.2	776.2	8.3	2 835
24	灰	临涣	297.4	122.5	598.1	276.5	215.2	24.0	305.1	7.82	1 839
25		任楼	235.1	98.8	460.0	996.2	225.0	0.0	334.4	8.06	2 182
26*		桃园	123.9	83.0	135.3	269.4	172.5	0.0	441.2	8.16	1 005
27		芦岭	47.5	33.7	166.3	85.1	163.8	0.0	397.9	7.6	806
28		祁南	190.8	93.8	27.3	251.7	220.0	0.0	405.8	7.45	986
29		朱庄	27.1	23.7	15.0	8.9	9.8	0.0	210.5	8.34	190
30*		杨庄	13.1	9.6	177.0	1.8	30.9	0.0	477.8	8.36	471
31*		勘探队	105.4	24.3	8.8	21.3	35.0	0.0	383.2	7.92	386
32*	奥	一电厂	106.6	26.7	13.0	30.1	53.2	0.0	371.0	7.91	415
33*		二电厂	94.6	29.9	18.3	14.2	66.3	0.0	374.1	8.26	410
34	灰	董亭	6.8	8.5	154.8	171.9	29.8	0.0	107.4	9.82	426
35		任楼	14.8	3.7	6.5	1.8	9.8	0.0	64.1	8.64	69
36		芦岭	83.05	27.2	131.1	86.5	154.33	0.0	392.2	7.60	760
37		朱仙庄	6.7	3.1	94.0	42.5	2.1	0.0	189.2	7.72	243

注:在水样编号右上角标“\*”表示同一矿井同一含水层各测试指标几乎相同的一组水样平均值。

### 3 系统聚类分析模型

皖北矿区某些含水层的水质较为相似,仅靠水质类型难以区分含水层水质之间的细微差别,难以判别突水水源。因此,有必要对四含、煤系、太灰和奥灰分别作聚类分析。通过聚类分析,除去异常水样,然后,重新建立水样集合,进行逐步判别,分析各含水层水化学特征。

对含  $m$  个(本文为 9 个)测试指标的每一水样样本,可定义为  $m$  维空间的点,在  $m$  维空间中的任意两点,其相似性可用“距离”度量,定义为“ $d_{ij}$ ”。若将任一样本看作一类,其类间相似性可用欧氏距离  $D_E$  表示,则:

$$D_E = d_{ij} = \left[ \sum_{l=1}^m (x_{il} - x_{jl})^2 \right]^{\frac{1}{2}}, \quad (1)$$

式中  $l$ ——样本指标数,  $l=1, 2, \dots, m$ ;  $i, j$ ——样本序号;  $x_{il}, x_{jl}$ ——样本各指标。

系统聚类法即对  $n$  个样本计算出两两间距离  $d_{ij}$ , 从中找出距离最小的两类  $G_p$  与  $G_q$ , 合并成一个新的类  $G_r$ ; 重新计算新类与其他各类间的距离, 再将欧氏距离最小的两类合并; 重复以上过程至所有样本聚为一类为止。矿区水样系统聚类时, 选取欧氏距离  $D_E$  小于某一规定值  $P$  的矿区突水含水层, 所有水样(设有  $n$  个)参与逐步判别分析。对  $n$  个具有  $m$  项指标的水样样本分为 4 类, 逐步判别分析可在其  $m$  项指标中挑选出对于 4 类判别作用显著的指标(如:  $X_1, X_2, \dots, X_l$ ) 构成判别函数:

$$Y_{(A_n)} = C_{0(A_n)} + C_{1(A_n)} X_1 + C_{2(A_n)} X_2 + \dots + C_{l(A_n)} X_l, \quad (2)$$

式中  $Y_{(A_n)}$ ——第  $A_n$  类的判别函数;  $A_n = 1, 2, 3, 4$ ;  $C_{0(A_n)}, C_{1(A_n)}, C_{2(A_n)}, \dots, C_{l(A_n)}$ ——第  $A_n$  类的判别系数。

对任一样本进行判别时, 可将相应特征数据代入每一类含水层的判别函数中, 计算各类的  $Y$  值, 取最大  $Y$  值, 将样本归入该类。

### 4 含水层水化学系统聚类分析

根据检测指标, 对表 1 中的四含水、煤系水、太灰水与奥灰水分别作系统聚类分析, 确定类间欧氏距离不大于 15 作为选取对象标准, 聚类分析结果见图 1。从图 1(A) 中可以看出, 任楼四含与其他生产矿井相差较大, 类间欧氏距离为 25, 说明任楼四含水化学特征与其他四含相比有较大差异, 排除 3 号水样, 选取矿区四含水 1、2、4、5、6、7、8 号水样参与逐步判别分析。根据图 1(B), 芦岭矿煤系水与其他

煤系类间欧氏距离为 25, 临涣煤系和童亭煤系所构成的一类与祁南煤系、任楼煤系这两类类间欧氏距离大于 20, 故排除 10、12、14、16、17 号水样点。所以, 可选取煤系水 9、11、13、15 号水样点参与逐步判别分析。用同样的方法, 可以排除图 1(C) 中 21、23 号水样, 选取太灰 18、19、20、22、24、25、26、27、28 号水样参与逐步判别分析; 排除图 1(D) 中 36 号水样, 选取奥灰 29、30、31、32、33、34、35、37 号水样参与逐步判别分析。

## 5 含水层逐步判别分析

### 5.1 判别分析结果

根据矿区水文地质条件和水样所处层位, 选取四含、煤系、太灰和奥灰 4 个主要突水含水层为判别归类对象, 即建立四含、煤系、太灰和奥灰 4 类判别函数。根据上述系统聚类分析的结果, 在表 1 所示的 37 个水样中选取了 28 个水样, 每个水样包括 9 个水质指标, 以 1、2、3、4 作分类变量, 分别代表四含、煤系、太灰和奥灰 4 个含水层, 应用逐步线性判别分析, 得出判别函数为:

$$Y(1) = 0.0084[SO_4^{2-}] + 0.0017(TDS) - 3.9341; \quad (3)$$

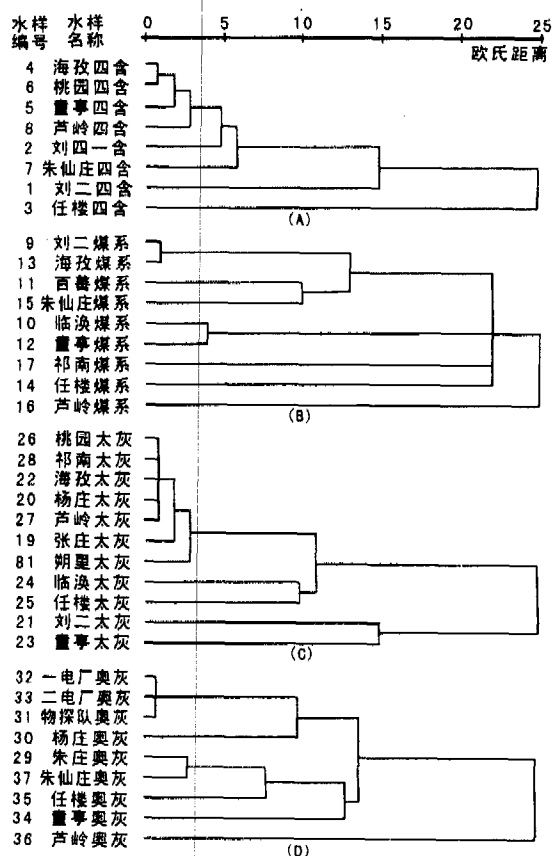


图 1 矿区含水层水化学系统聚类分析

Fig. 1 Systemic clustering of hydrochemistry in the aquifers of the mining area

$$Y(2) = -0.0049[\text{SO}_4^{2-}] + 0.0056(\text{TDS}) - 4.6475; (4)$$

$$Y(3) = -0.0083[\text{SO}_4^{2-}] + 0.0051(\text{TDS}) - 3.3946; (5)$$

$$Y(4) = -0.0044[\text{SO}_4^{2-}] + 0.0020(\text{TDS}) - 1.6536. (6)$$

式中  $[\text{SO}_4^{2-}]$ ——硫酸根离子浓度; (TDS)——总溶解固体浓度。

可见,当所有样本分为4类时,通过逐步判别分析,将每个样本原有的9个指标优选为2个。需要进行判别分类时,只需将样本的  $\text{SO}_4^{2-}$  和 TDS 浓度代入判别函数式(3)、(4)、(5)、(6)中,分别计算各类的  $Y$  值,取最大  $Y$  值,将样本归入该类即可。

## 5.2 判别效果检验

### 5.2.1 $F$ 检验

若检验 1,2 两类间的判别效果,可采用  $F$  检验值:

$$F_{1,2} = \frac{n_1 n_2 (n_1 + n_2 - 1 - p)}{(n_1 + n_2)(n_1 + n_2 - 2)p} D_{1,2}^2, \quad (7)$$

式中  $n_1, n_2$ ——第 1,2 类样品个数;  $p$ ——判别变量的个数;  $D_{1,2}^2$ ——广义马氏距离,其值可表示为:  $D_{1,2}^2 = (\bar{X}_{(1)} - \bar{X}_{(2)}) S^{-1}$ , 其中  $S = (S_{ij})_{p \times p}$  表示变量间的协方差矩阵;  $\bar{X}_{(1)}, \bar{X}_{(2)}$  分别表示第 1,2 类的子样均值向量。

在均值相等的假设下,  $F_{1,2}$  服从自由度为  $p$  和  $n_1 + n_2 - 1 - p$  的  $F$ -分布,在给定显著性水平  $\alpha$  下,若  $F_{1,2} > F_\alpha$  ( $F_\alpha$  是在显著性水平为  $\alpha$  的  $F$  临界值),则该两类判别效果显著。对于多类判别的情况,可把各个类别两两配对,逐对计算  $F$  值,用以辩明各对的判别效果。当计算所得  $F$  值大于相应临界值时,两类间有显著性差异,且  $F$  值越大,差异越显著,判别效果越好。

矿区含水层逐步判别  $F$  检验结果见表 2。由表 2 可知,在检验水平  $\alpha = 0.05$ , 各类间  $F$  值大于  $F_{0.05}$ , 类间差异显著,说明选入  $\text{SO}_4^{2-}$  和 TDS 两变量的判别能力显著,效果较好,能有效地判别矿区 4 个主要突水含水层。

表 2 矿区水样判别效果的  $F$  检验

Table 2  $F$ -verification of the distinguishing effect of water samples in the mining area

类别	类间 $F$ 值	$F_{0.05}$
2 与 1	52.369 8	4.46
3 与 1	30.236 5	3.81
3 与 2	62.312 3	4.10
4 与 1	231.365 3	3.89
4 与 2	73.265 3	4.26
4 与 3	163.235 1	3.74

### 5.2.2 回判检验

将 28 个水样点数据分别代入判别函数进行回判分析检验,回判结果见表 3。回判结果归属概率较高(大于 0.700)的,可以将水样划入该类;归属概率较低(小于 0.700)的,说明有其他判别水源混入。上述 28 个水样中,有 3 个水样误判,水源判别正确率为 89.3%。

表 3 矿区水样判别分析回判检验

Table 3 Retrospective verification of distinguishing results of water samples in the mining area

水样 编号	原归 类	判别 归类	归属 概率	水样 编号	原归 类	判别 归类	归属 概率
1	1	1	0.916	22	3	3	0.75
2	1	1	0.992	24	3	3	0.85
4	1	3	0.534	25	3	3	0.87
5	1	1	0.954	26	3	3	0.74
6	1	3	0.662	27	3	3	0.70
7	1	1	0.943	28	3	3	0.94
8	1	1	0.721	29	4	4	0.95
9	2	2	0.750	30	4	4	0.90
11	2	2	0.850	31	4	4	0.95
13	2	1	0.438	32	4	4	0.90
15	2	2	0.803	33	4	4	0.74
18	3	3	0.952	34	4	4	0.93
19	3	3	0.953	35	4	4	0.86
20	3	3	0.734	37	4	4	0.94

## 6 结语

以取自皖北矿区四含、煤系、太灰和奥灰 4 个主要突水含水层水样的 9 个水质常规指标 ( $\text{Ca}^{2+}$ 、 $\text{Mg}^{2+}$ 、 $\text{K}^+ + \text{Na}^+$ 、 $\text{Cl}^-$ 、 $\text{SO}_4^{2-}$ 、 $\text{CO}_3^{2-}$ 、 $\text{HCO}_3^-$ 、pH 值与 TDS) 进行聚类分析,排除了在各含水层中水化学特征异常的水样后,再联立系统聚类分析过程后各含水层水化学特征相近的水样,进行整体逐步判别分析,选出了  $\text{SO}_4^{2-}$  与 TDS 两种标型组分,建立了皖北矿区突水水源的判别模型。该模型判别效果显著。

## 参考文献

- [1] 成春奇,徐 龙. 临涣矿区水化学特征及在矿井水源判别中的意义[J]. 煤田地质与勘探, 1995, 23(1): 43-46.
- [2] 成春奇,葛晓光,王登榜. 百善矿区水文地球化学特征及其在矿井涌出水源判别中的应用[J]. 淮南矿业学院学报, 1994, 14(1): 25-32.
- [3] 李定龙,周治安. 临涣矿区底含水化学特征及其形成作用探讨[J]. 煤田地质与勘探, 1994, 22(4): 37-41.
- [4] 孙本魁,段中稳,金洪元. 任楼矿水源判别模型的自动识别[J]. 煤田地质与勘探, 1999, 27(3): 39-43.
- [5] 董春江,宋三胜. 判别分析方法在芦岭矿井出水水源判别中的应用[J]. 矿业科学技术, 2000(3): 1-4.