

矿区地下水系统水质分类判别的多元统计分析

江晓益¹,成春奇²

(1. 浙江省水利河口研究院,杭州 310020;2. 安徽理工大学地球与环境学院,淮南 232001)

摘要:以某矿区地下水系统为例,对该矿区地下水水化学资料进行了多元统计分析方法耦合应用研究,主要包括利用因子分析对存在相关关系的离子变量进行空间降维处理,找出能够反映众多离子信息的基础变量(正交因子),以其作为系统聚类变量;运用系统聚类法获取能代表各地下水子系统水化学特征的典型水样;使用贝叶斯逐步线性判别分析建立地下水各子系统水化学判别模型(判别函数),并对随机检验样品进行判别归属和判别模型统计检验。结果表明:这是一种稳定性较好且切实有效的、适用于矿区地下水系统水化学分类及水源水化学判别的方法。

关键词:地下水系统;水化学分类;因子分析;聚类分析;判别分析

中图分类号: P641.2

文献标识码: A

文章编号: 1000-3665(2009)03-0016-05

地下水系统的水化学分类及判别研究是水文地质学的重要内容之一。矿区地下水系统水化学分类判别广泛采用舒卡列夫定名分类法,此法忽略各标量组分之间的相关性,属于简单定性分类。近年来,一些不确定性数学方法,逐渐应用到矿区地下水系统水化学分类和水源水化学判别^[1-3],其优点在于考虑了地下水系统水化学界线的模糊性和灰色性,结果具有一定的合理性,但每种分析方法的主观性都较强,如模糊数学分析时权矩阵和隶属函数的选择、灰色系统分析时权矩阵和白化函数的选择,仍需要靠经验去确定,从而影响了结果的客观性和唯一性。正确的水化学分类判别模型,应通过对水化学资料进行多元统计研究,寻找该地区水化学的内在规律或统计规律,在此基础上来确定反映客观地质现象的水化学分类界线,进行归类判别^[4]。本文以安徽淮南某矿区为例,对该矿区地下水化学资料进行了多种统计分析方法耦合应用研究,主要包括因子分析、聚类分析和判别分析^[5],最终建立地下水各子系统水化学判别模型(判别函数)。

1 某矿区地下水系统简介

某矿区位于淮南煤田南缘中段,属推覆体下缓倾斜单斜构造单元。矿区地质剖面示意如图1所示。井田自上而下发育有新生界松散层孔隙含水组(Q)、下元古界片麻岩裂隙含水组(P₁₁)、寒武系灰岩溶隙含水

组(O₁₊₂)、“夹片”溶(裂)隙含水组(O-P)、煤系砂岩裂隙含水组(I-13)、太原组灰岩溶隙含水组(C₃)及奥陶系灰岩溶隙含水组(O₁₊₂)等7个含水组。新生界松散层孔隙含水组划分为3个含水层和3个隔水层,富水性中等,与地表水和大气降水联系较密切,其中部分地段形成三含或二含直接与寒武系接触的“天窗”。下元古界片麻岩裂隙含水组呈狭长条带状覆盖于井田南部的“夹片”地层和煤系地层之上,富水性弱且不均一。寒武系灰岩溶隙含水组呈条带状覆盖于煤系地层之上,富水性在平面和垂向上具明显的差异性。“夹片”溶(裂)隙含水组呈带状透镜体直接覆盖于各煤层隐伏露头之上,富水性较弱且不均一,局部地段富含水。煤系砂岩裂隙含水组富水性较弱,多以静储量为主。太原组灰岩溶隙含水组富水性中等但不均一,为区域主要含水层。奥陶系灰岩溶隙含水组富水性中等—丰富,为区域强含水层,是太原组灰岩的直接补给水源。

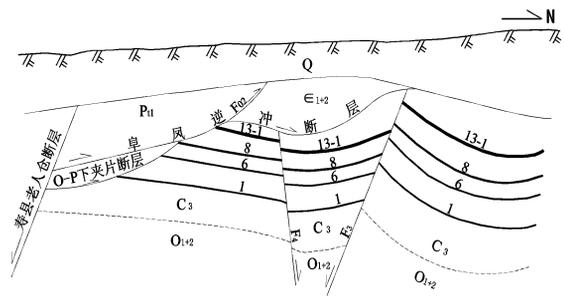


图1 矿区地质剖面示意图

Fig.1 Diagram showing mine geological section

收稿日期: 2008-07-28; 修订日期: 2008-10-17

作者简介: 江晓益(1973-),男,工程师,硕士,从事水文地质及工程地质研究工作。

E-mail: xyjianghz@sohu.com

2 地下水系统水质分类判别多元统计分析

本文收集了该矿区 102 个水样水化学成分资料(表 1),水样采集于水源井、长观孔及采掘工作面,时

间主要为 1992~2005 年,在此期间矿区总体开采规模不大,地下各含水组未出现较大水位动态变化,故本次研究仅从静态角度对水质资料进行多元统计分析。

表 1 矿区地下含水层部分水样数据表(单位:mg/L)

Table 1 Water samples of mine groundwater(mg/L)

编号	Na ⁺ + K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻	CO ₃ ²⁻	编号	Na ⁺ + K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻	CO ₃ ²⁻
新生界松散层孔隙含水组								52	766.0	19.6	8.5	811.8	172.9	296.6	28.8
1	74.8	48.5	15.6	14.3	4.1	372.7	3.0	54	449.3	44.7	33.5	597.3	11.0	316.1	21.0
2	45.6	51.6	22.5	24.4	0.8	338.5	0	56	542.8	7.4	3.4	430.7	143.6	439.8	0
3	73.8	53.7	16.4	17.6	1.8	391.5	0	57	422.6	66.6	44.9	623.4	89.3	274.8	0
4	56.0	55.6	18.0	20.7	4.8	356.5	0	58	240.0	24.8	15.7	185.4	20.8	348.3	22.8
5	62.0	51.6	16.4	15.5	0.1	360.5	0	59	442.8	39.8	34.5	565.0	3.5	414.7	0
6	68.8	52.8	16.1	16.5	2.0	383.5	0	60	453.0	34.4	30.6	547.9	2.3	377.2	20.3
7	60.5	52.9	16.7	14.5	3.9	349.0	13.6	61	575.0	4.2	1.1	472.2	2.1	606.5	0
9	33.4	62.3	12.7	7.6	5.0	237.7	0	63	839.8	71.9	63.2	1146.8	172.3	258.7	0
10	36.5	59.6	11.5	7.0	3.5	300.4	12.8	64	976.8	14.4	6.4	1084.8	144.3	368.8	22.6
12	46.7	79.4	7.3	17.2	22.6	279.6	26.5	65	463.5	81.9	50.4	619.1	178.3	341.4	0
13	50.0	67.5	8.0	14.2	3.5	339.3	0	67	767.3	20.1	8.9	798.3	19.2	500.4	51.0
14	93.3	53.7	11.5	31.2	11.5	380.8	0	68	1010.0	25.3	10.4	1083.4	220.5	390.5	18.0
16	489.5	43.3	22.7	583.2	132.7	242.2	12.9	69	813.0	9.9	5.5	357.0	219.5	1029.4	58.8
18	621.3	71.8	26.8	822.5	65.6	368.6	0.6	70	689.3	21.4	7.9	756.6	10.3	432.0	19.8
19	370.5	61.3	24.7	423.3	96.7	324.6	19.2	72	427.4	28.6	23.8	476.5	9.9	348.5	38.4
寒武系灰岩溶隙含水组								73	635.0	33.1	21.6	617.6	116.5	524.1	11.7
20	358.8	67.3	26.0	407.6	122.2	335.1	9.6	74	596.0	38.1	13.7	692.4	125.7	252.0	17.7
21	431.5	27.3	21.7	413.4	128.8	283.0	0	75	636.8	16.8	5.7	610.5	8.6	572.3	0
22	309.3	75.2	34.9	399.0	129.6	307.5	0	76	690.5	21.6	9.0	794.8	3.5	424.1	0
24	538.4	19.2	19.0	400.6	419.7	165.2	42.3	77	613.5	7.9	3.2	598.8	1.9	413.8	46.7
25	314.3	60.0	51.3	400.9	127.6	356.2	0	78	854.5	71.9	70.5	1272.7	170.1	252.3	0
26	469.3	43.3	1.5	487.5	66.3	235.9	62.8	79	529.0	45.0	39.3	644.1	101.4	328.9	29.4
27	402.0	53.5	27.4	452.0	121.5	349.3	0	80	772.5	11.2	20.4	645.2	40.0	640.7	109.0
29	326.3	79.0	51.4	459.8	103.5	371.6	0	82	459.0	10.4	3.9	268.0	1.7	708.4	0
30	142.5	124.3	1.7	89.5	47.7	522.0	0	83	634.5	12.6	7.9	638.5	0.6	527.2	0
32	240.5	21.0	15.4	128.7	73.9	330.6	42.0	84	594.0	4.8	2.4	468.7	49.2	607.7	0
33	556.7	74.7	39.4	714.0	169.5	342.9	0	85	641.5	47.7	42.8	769.3	207.8	303.3	16.8
下元古界片麻岩裂隙含水组								86	959.0	4.8	2.3	841.0	1.7	834.7	40.8
34	1062.5	110.4	81.4	1530.4	329.0	211.6	36.7	87	764.5	2.8	3.2	583.9	60.9	735.9	35.4
35	958.0	179.2	97.2	1696.8	204.6	180.1	23.7	太原组灰岩溶隙含水组							
36	686.5	84.2	45.9	905.7	132.6	435.4	0	88	767.1	4.4	9.3	811.1	15.6	421.8	47.7
37	706.0	100.4	67.5	1050.0	343.4	187.3	0	89	96.7	67.3	16.0	66.9	41.9	338.7	0
38	1406.0	229.7	108.2	2451.7	229.4	205.4	0	90	368.1	11.2	0.7	260.6	129.2	264.4	33.9
“夹片”溶(裂)隙含水组								92	280.0	113.0	46.3	459.1	117.8	271.5	24.0
39	934.5	149.7	100.2	1367.7	416.9	255.7	20.1	93	230.8	87.8	44.5	388.2	80.1	289.9	0
40	830.5	131.5	80.7	1209.9	490.4	273.2	0	94	254.8	54.4	22.2	276.2	86.4	328.5	0
41	651.9	91.4	47.9	925.1	304.9	107.8	11.3	奥陶系灰岩溶隙含水组							
42	483.0	112.2	60.8	714.2	267.5	233.2	11.8	95	149.3	60.5	19.2	164.8	49.6	298.4	0
43	948.8	92.7	54.5	1290.2	226.4	309.9	25.8	96	298.0	72.6	27.7	349.6	104.6	353.9	0
煤系砂岩裂隙含水组								断层带							
44	373.5	39.4	32.3	391.3	1.8	464.0	0	97	576.3	8.5	5.7	108.5	4.2	926.7	170.0
46	388.3	29.7	13.2	362.7	2.1	472.5	0	98	896.0	146.0	96.2	1296.0	545.6	203.4	14.3
47	365.5	4.6	3.1	234.4	1.3	434.5	40.8	99	856.8	2.4	2.2	170.2	7.7	1634.5	85.0
49	721.8	13.2	4.3	481.6	7.2	858.3	57.2	100	962.5	2.0	1.5	141.8	7.7	1610.9	244.8
50	878.8	40.1	9.7	792.3	194.0	345.6	177.0	101	936.5	2.4	1.5	138.3	3.8	1449.5	299.0
51	900.8	39.0	10.5	885.4	134.3	399.6	138.9	102	770.7	152.3	38.9	1256.3	10.7	364.8	0

2.1 地下水化学成分的因子分析

地下水中 8 种常量离子总量可占水中 TDS 的 95%~99%^[6]。这些离子属地下水中的标型组分,其差异是地下水化学分类的主要依据。一般各离子并非完全独立地存在于地下水系统中,它们之间总有些相关性,离子之间的相关意味着其反映的地质信息重合。如果不作处理就进行聚类分析,相当于给这些离子变量进行了加权,必然夸大某些地质信息因素的作用,使

计算结果发生畸变。为克服这种现象,必须对离子进行因子分析,找出基本的数据结构,用新生成的正交因子作为新变量进行聚类分析。

利用 SPSS 软件对 7 个离子变量进行相关分析表明:Na⁺ + K⁺ 和 Cl⁻、Ca²⁺ 和 Mg²⁺ 之间存在较强的相关关系,其它离子变量之间也存在一定相关关系(表 2),需要对原始变量数据进行因子分析。

表 2 相关系数矩阵

Table 2 Correlation matrix

	Na ⁺ + K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻	CO ₃ ²⁻
Na ⁺ + K ⁺	1.000 0	0.067 7	0.305 1	0.821 1	0.453 7	0.245 0	0.386 9
Ca ²⁺	0.067 7	1.000 0	0.813 9	0.499 6	0.507 5	- 0.477 6	- 0.348 1
Mg ²⁺	0.305 1	0.813 9	1.000 0	0.673 1	0.674 8	- 0.453 7	- 0.273 1
Cl ⁻	0.821 1	0.499 6	0.673 1	1.000 0	0.601 3	- 0.280 5	- 0.071 6
SO ₄ ²⁻	0.453 7	0.507 5	0.674 8	0.601 3	1.000 0	- 0.370 8	- 0.082 6
HCO ₃ ⁻	0.245 0	- 0.477 6	- 0.453 7	- 0.280 5	- 0.370 8	1.000 0	0.646 8
CO ₃ ²⁻	0.386 9	- 0.348 1	- 0.273 1	- 0.071 6	- 0.082 6	0.646 8	1.000 0

利用因子分析中主成分分析法从离子相关系数矩阵中提取初始因子,得到了相关系数矩阵的 7 个特征值(主成分所解释的方差)(表 3)。提取初始因子时,一般依据特征值准则^[7],即取特征值大于等于 1 的主成分作为初始因子。本次选取 2 个因子,累计解释了水样信息中总方差的 78.4%。

表 3 方差分析表

Table 3 Total variance explained

主成分	特征值			特征值(提取后)		
	方差	方差 (%)	累积方差 (%)	方差	方差 (%)	累积方差 (%)
1	3.469 6	49.565 4	49.565 4	3.469 6	49.565 4	49.565 4
2	2.017 0	28.814 9	78.380 4	2.017 0	28.814 9	78.380 4
3	0.613 4	8.763 3	87.143 7			
4	0.470 3	6.719 0	93.862 7			
5	0.286 5	4.093 5	97.956 2			
6	0.142 2	2.031 7	99.987 9			
7	0.000 8	0.012 1	100.000 0			

选取的因子初始载荷矩阵一般需进行方差极大正交旋转,使各离子变量在同一因子上的载荷区别明显。表 4 为采用方差极大正交旋转方法迭代 3 次后的因子载荷矩阵 A,每个离子变量的信息被保留的程度体现在各自的公因子方差上。因子 f₁ 除与 CO₃²⁻ 载荷很小,与 HCO₃⁻ 载荷较小外,与其它各离子载荷均较大,而因子 f₂ 在 CO₃²⁻ 和 HCO₃⁻ 上载荷相对较大,故可把 f₁ 命名为 TDS 因子, f₂ 命名为碱度因子。

聚类分析需要对因子进行计量。表 5 给出了经正交旋转后的因子计量系数的回归估计值,因子计量系数乘以每个水样离子变量的标准化值就得到该水样的因子值。

表 4 旋转后因子载荷矩阵

Table 4 Rotated component matrix

	旋转因子载荷(A)		公因子方差 H _j ²
	f ₁	f ₂	
Na ⁺ + K ⁺	0.775 0	0.563 5	0.918 2
Ca ²⁺	0.616 1	- 0.574 1	0.709 2
Mg ²⁺	0.801 7	- 0.434 6	0.831 6
Cl ⁻	0.930 9	0.009 1	0.866 7
SO ₄ ²⁻	0.797 8	- 0.187 4	0.671 6
HCO ₃ ⁻	- 0.238 0	0.837 8	0.758 6
CO ₃ ²⁻	0.033 4	0.854 1	0.730 6

表 5 因子计量系数矩阵

Table 5 Component score coefficient matrix

因子计量系数	f ₁	f ₂
Na ⁺ + K ⁺	0.301 5	0.320 3
Ca ²⁺	0.155 4	- 0.210 4
Mg ²⁺	0.227 9	- 0.131 6
Cl ⁻	0.307 1	0.080 9
SO ₄ ²⁻	0.247 1	- 0.019 4
HCO ₃ ⁻	- 0.009 2	0.361 5
CO ₃ ²⁻	0.081 5	0.391 3

2.2 水质样品的系统聚类分析

为获取矿区地下水各子系统典型水样,利用系统

聚类法中的最短距离法、最长距离法、类平均法、重心法和离差平方和法等5种方法对水样进行聚类,距离测度使用平方欧氏距离,聚类变量即为正交因子(TDS因子和碱度因子)。

根据SPSS软件完成的102个水样的5种系统聚类谱系图,除最短距离系统聚类的谱系图成链状连接,类别模糊,分类效果差外,其它4种系统聚类均较好地在水样进行了合理划分。根据聚类谱系图并结合矿区水文地质条件分析,矿区7个含水组中新生界松散层、寒武系灰岩、下元古界片麻岩及煤系砂岩的水样大多能独聚成类,可视为独立的含水系统。“夹片”溶(裂)隙含水组仅有的5个水样均聚到下元古界片麻岩裂隙含水组中,说明它们有直接水力联系或具相同的补给源。新生界三含的4个水样大多聚到寒武系灰岩含水组中,说明三含与寒武系灰岩溶隙含水组有直接水力联系,通道可能就是局部存在的“天窗”。太原组及奥陶系灰岩含水组中所采集的9个水样,各含水组中均有分布,其特征不甚明显,该含水组中的典型水样尚无法提取。综合分析,该矿区的7个含水组可分为5个地下水子系统,即新生界孔隙含水系统(简称“新生界”)、推覆体寒武系灰岩溶隙含水系统(简称“寒灰”)、推覆体下元古界片麻岩—“夹片”裂(溶)隙含水系统(简称“片麻岩”)、煤系砂岩裂隙含水系统(简称“煤系”)及下伏太原组及奥陶系灰岩溶隙含水系统(简称“太灰”、“奥灰”)。

在4种系统聚类法中,最长距离法与重心法、类平均法与离差平方和法的结果较接近。从系统聚类法的性质来看,离差平方和法与类平均法提取样本的信息较多,聚类空间的浓缩性或扩张性也较适中,聚类结果也与实际情况比较相符^[5]。因此,以离差平方和法和类平均法系统聚类谱系图为基础,参考最长距离法和重心法的聚类结果,从4类聚类谱系图上提取各子系统典型水样:新生界15个、寒灰11个、片麻岩9个、煤

$$\begin{cases} y_1(x) = -0.0064[Na^+ + K^+] + 0.2120[Ca^{2+}] - 0.1917[Mg^{2+}] + 0.0228[SO_4^{2-}] - 5.8790 \\ y_2(x) = -0.0105[Na^+ + K^+] + 0.0813[Ca^{2+}] - 0.1076[Mg^{2+}] + 0.0485[SO_4^{2-}] - 11.5374 \\ y_3(x) = -0.0261[Na^+ + K^+] + 0.1502[Ca^{2+}] - 0.2589[Mg^{2+}] + 0.0988[SO_4^{2-}] - 48.4423 \\ y_4(x) = -0.0264[Na^+ + K^+] - 0.0834[Ca^{2+}] + 0.1249[Mg^{2+}] - 0.0052[SO_4^{2-}] - 7.9255 \end{cases} \quad (1)$$

式(1)中, y_1 、 y_2 、 y_3 、 y_4 分别代表新生界、寒灰、片麻岩及煤系含水系统的判别函数。利用这4个判别函数进行判别归类的原则是,把待判水样的 $Na^+ + K^+$ 、 Ca^{2+} 、 SO_4^{2-} 、 Mg^{2+} 的实际浓度值直接代入方程中,计算出函数值并比较,把水样划归为函数值最大的方程所代表的地下水子系统。

$$P(A_g/x) = e^{y_g(x)} / \sum_{i=1}^4 e^{y_i(x)} \quad (2)$$

系16个。

2.3 地下水系统水质的贝叶斯线性判别分析

判别分析要求得到矿区4个地下水子系统的水化学判别函数,分类信息即系统聚类得到的地下水各子系统的典型水样。贝叶斯线性判别一般要求母体变量观测值必须服从多元正态分布,而且要求各母体的协方差阵无显著差别。在应用中,当各母体的样本规模比较接近时,如最大一组的样本容量不超过最小一组样本容量的1.5倍时,违反协方差阵相等的假设条件,影响也不太大^[7]。

利用SPSS软件对4个地下水子系统的51组典型水样变量数据进行单变量正态性检验,各变量基本满足正态假设,但统计检验拒绝了母体协方差阵相等的假设,需从样本规模上进行调整。从新生界和煤系水样组中分别随机抽取2个和3个水样不参与建立判别函数,仅作为检验样本。前已表明水质离子变量之间的相关系数较高。基于克服多重共线性和选择合适的预测变量进入判别函数,使用逐步判别分析(表6)。

表6 逐步判别引入的变量

Table 6 Variables entered with stepwise method

步数	入选变量	wilks	f_1	f_2	f_3	F			
						统计值	f_1	f_2	P
1	$Na^+ + K^+$	0.1935	1	3	42	58.36	3	42.00	0.0000
2	Ca^{2+}	0.0571	2	3	42	43.51	6	82.00	0.0000
3	SO_4^{2-}	0.0222	3	3	42	40.94	9	97.50	0.0000
4	Mg^{2+}	0.0159	4	3	42	32.62	12	103.47	0.0000

从表6可以看出,逐步判别分析最终选择了 $Na^+ + K^+$ 、 Ca^{2+} 、 SO_4^{2-} 、 Mg^{2+} 4种离子作为影响显著的预测变量进入了判别函数。若假设各母体的先验概率相等,由于各母体的协方差阵不相等,分类协方差阵可以利用合并的类内协方差阵^[7]。经过SPSS逐步判别分析,得到矿区4个地下水子系统的贝叶斯线性判别函数如下:

$$\text{式中: } y_g(x) = y_g(x) - \max_{1 \leq g \leq 4} y_g(x)$$

根据函数值按式(2)计算水样属于各地下水子系统的后验概率。经验表明^[8],后验概率较高(大于85%),可以将该水样划入该类,后验概率较低(小于80%),说明有第二水源混入。

函数判别效果的统计检验,可通过引入判别函数的变量均值在各类间的差异来检验,结果表明变量类间均值差异不显著的概率均小于0.05,表明选入的4个变量判别能力显著。对类间的总体判别效果,用马

氏距离作 F 检验,结果均拒绝了类间马氏距离为零的零假设,说明判别函数是有效的。表 7 是随机抽取的 5 个检验水样的判别结果,各水样均未发生误判。

表 7 检验样品的判别结果

Table 7 Classification results of test water samples

水样 编号	实际 归属	判别 归属	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_4)$
5	y_1	y_1	0.998 4	0.001 6	0.000 0	0.000 0
10	y_1	y_1	0.999 9	0.000 1	0.000 0	0.000 0
46	y_4	y_4	0.002 4	0.006 8	0.000 0	0.990 1
60	y_4	y_4	0.000 0	0.003 9	0.000 0	0.996 1
76	y_4	y_4	0.000 0	0.000 0	0.000 0	1.000 0

3 结论及建议

对某矿区地下水水质资料多元耦合统计分析表明,该矿区地下 7 个含水组可分为 5 个地下水子系统,并获得了该矿区 4 个地下水系统水质线性判别函数。

地下水系统水化学分类判别模型的可靠程度,依赖于水样的数量、变量的选择及采样空间分布的合理性。由于本次收集的水样数据采样时间跨度较大、水样点空间分布较不均匀等因素影响,获取各子系统的典型水样数量十分有限,特征变量也只选取了常量离子,判别模型还有待完善。建议今后采集更多的地下

水样品,从中选择时间跨度较小且空间分布较均匀的水样作为统计样本,统计模型需不断调整和完善。模型仍属静态模型,今后应尝试将时间因子引进模型,建立水化学动态与水位动态预测模型,以实现各子系统内部及系统之间的水化学动态监测。

参考文献:

- [1] 尹国勋,杨娜,贺玉晓,等.焦作市市区地下水水质现状评价[J].环境工程,2004,22(4):66-69.
- [2] 贲旭东,郭黄海,解奕伟,等.模糊综合评判在矿井突水水源判别中的应用及探讨[J].矿业安全与环保,2006,33(3):57-59.
- [3] 苗成林,周宏.基于灰色模型的水质预测[J].中国农村水利水电,2007(2):126-128.
- [4] 地质矿产部水文地质工程地质研究所.水文地球化学理论与方法的研究[M].北京:地质出版社,1985.
- [5] 张尧庭,方开泰.多元统计分析引论[M].北京:科学出版社,1997.
- [6] 沈照理,朱宛华,钟佐荣.水文地球化学基础[M].北京:科学出版社,1993.
- [7] 郭志刚.社会统计分析方法—SPSS 软件应用[M].北京:中国人民大学出版社,1999.
- [8] 王广才,王秀辉,李竞生,等.平顶山矿区矿井突(涌)水水源判别模式[J].煤田地质与勘探,1998,26(3):47-50.

Hydrochemical classification and identification of groundwater in mining region using multivariate statistical analysis

JIANG Xiao-yi¹, CHENG Chun-qi²

(1. Zhejiang Institute of Hydraulics & Estuary, Hangzhou 310020, China; 2. Department of Natural Resources and Environmental Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: Taking the groundwater system in mining region as example, the applied research to classify and identify the groundwater occurred in different subsystems and stopes was actualized on the basis of hydrochemical data analysis with the techniques of multivariate statistics. Taking the concentration of the major ions dissolved in groundwater as variables, the factor analysis was performed to reduce the dimensions of variables and identify the underlying variables or the orthogonal factors which reflect the hydrochemical characteristics of groundwater and can be then used in cluster analysis. As the results of hierarchical cluster analyses, a set of statistically derived typical samples were obtained, which hydrochemically represent the groundwater in individual subsystems. Multigroup Bayes discriminant analysis stepwise method was conducted with the subsystems as groups and the major initial ions as variables, and then a set of linear discriminant functions was obtained. The functions were applied to predict the sources of some random water samples on the basis of the variables with unknown group membership and proved to have the ability of differentiation with statistical tests. The results show that it is an effective and practical way in classifying and identifying of groundwater or gushing water sources with high stability.

Key words: groundwater system; hydrochemical identification; factor analysis; cluster analysis; discriminant analysis

责任编辑:汪美华