

DOI :10.3969/j.issn.1004-6933.2011.01.009

对应分析方法在白城市洮北区地下水水质评价中的应用

龚 磊,卢文喜,辛 欣,伊燕平

(吉林大学环境与资源学院,吉林 长春 130026)

摘要 :应用对应分析方法研究了白城市洮北区地下水水质状况和影响水质的主要因素,确定了主要的污染因子和污染原因。结果显示 :白城市洮北区地下水污染农村最轻,市区次之,城郊结合部最严重,主要污染因子是硝酸盐和亚硝酸盐,水质污染来源是垃圾渗滤液和生活污水等。将对应分析结果与综合指数法评价结果进行对比发现,两种方法对水样点水质污染程度的分区基本一致,但对应分析方法综合考虑了样本和变量之间的关系,并根据变量的载荷大小和变量与样本水质分区间的关系,确定出研究区内各水质分区的主要污染因子。

关键词 :对应分析 ;地下水 ;水质评价 ;白城市

中图分类号 :P641.12 ;X823

文献标识码 :A

文章编号 :1004-6933(2011)01-0038-04

Application of correspondence analysis method in the assessment of groundwater quality in Taobei area of Baicheng City

GONG Lei ,LU Wen-xi ,XIN Xin ,YI Yan-ping

(College of Environment and Resources ,Jilin University ,Changchun 130026 ,China)

Abstract :The correspondence analysis method was applied to study the groundwater quality status and the main factors which affect the groundwater quality in Taobei area of Baicheng City ,and the cause of pollution and the main pollution factors were determined. The status of groundwater pollution in Taobei area of Baicheng City was that the least serious area was in rural ,followed by the urban areas ,and the worst area was around the suburbs. The main pollution factors which affected this area were nitrate and nitrite ,and the sources of water pollution were landfill leachate and sewage. Comparing the results from correspondence analysis method with the results by comprehensive index method ,it could found that the division of water pollution degree by two methods was generally the same ,but the correspondence analysis method took the relationship between the sample and variables into account ,and identified the main pollution factors in each zone according to the quantity of the variable load and the relationship between variables and water quality zone.

Key words :correspondence analysis ; groundwater ; quality assessment ; Baicheng City

地下水水质评价是地下水环境质量评价的一项重要内容。目前用于水质评价的数学方法很多,如单项组分评价法、综合指数法、模糊数学法、灰色聚类法等。这些传统方法,在水质评价中一般侧重于就水样点的污染程度进行评价、分级,很少揭示样本和变量(评价指标)间的内在规律^[1-4]。对应分析方法是

由法国数学家 J P Bcozecri 在 1970 年首次提出的,它是在因子分析的基础上发展起来的一种多元统计分析方法。它将样本和变量有机地结合起来进行分析,同时研究样品和变量之间的相互关系,用因子进行推断和解释观测数据^[5]。笔者用对应分析方法对白城市洮北区地下水状况进行分析,不仅对洮北区环境保护具有现实意义,而且对于加强该地区的饮水安全、提高当地人民的生活质量也具有一定

基金项目 :吉林省科技发展计划项目(20080456)

作者简介 :龚磊(1987 —)男,湖北荆门人,硕士研究生,研究方向为数字流域及流域水管理。E-mail :gonglei09@mails.jlu.edu.cn

通讯作者 :卢文喜,教授。E-mail :luwenxi@jlu.edu.cn

的参考价值。

1 水质评价中的对应分析方法

1.1 基本原理及特点

对应分析的基本思想是将一个矩阵中的行和列各元素的比例结构以点的形式在较低维的空间中表示出来。在地下水水质评价中,水质变量空间中的 n 个样品点的协方差矩阵 A 与样本空间的 k 个变量点的协方差矩阵 B 具有相同的特征值,而特征值表示各个因子所提供的方差,即各个公因子对所有公因子的方差贡献是相同的。因此可用同一坐标轴来表示水质检测指标和样本。同时也能够利用 R 型分析的结果推出 Q 型分析的结果(即特征根和特征向量),从而克服了样本容量大而难于进行 Q 型分析计算的困难^[2 6-7]。

对应分析综合了 R 型(对指标变量的)因子分析和 Q 型(对样点的)因子分析的优点,可以把变量和样点的载荷反映在相同的公因子轴上,以达到把变量和样点联系起来加以分析、解释和推断的目的^[2]。

具体来讲,对应分析具有以下特点:

- a. 统一性:将样品与变量结合起来分析,其结果能统一反映在一张图上,便于进行解释和推断;
- b. 对称性:通过变量变换,将原始数据 X_{ij} 变换成中间变量 Y_{ij} ,变量与样品是对称的,于是可将 R 型因子分析和 Q 型因子分析结合起来;
- c. 互推性:在实际运算中,只进行 R 型因子分析,经过简单换算,便可得到两种因子分析的结果,从而节省了计算 Q 型因子分析的时间,也克服了由于样品容量大,在进行 Q 型因子分析时带来的计算上的困难^[7]。

1.2 数据预处理和计算步骤

根据对应分析的数学原理^[2 6 8-12],结合水质评价问题的特点,将水质评价的对应分析方法中的数据预处理和计算步骤阐述如下。

1.2.1 数据预处理

在水质评价中,为了消除量纲差异^[13]对对应分析结果的影响以及满足对应分析的数据要求($x_{ij} > 0$),需对原始数据进行变换,笔者采用单项指数指标(标准值取为 GB/T 14848—93《地下水质量标准》Ⅲ类)对原始数据进行处理,经处理后的数据为各指标相对于Ⅲ类水质指标的相对值,具体处理方法如下^[14]:

水质中各组分的单项污染指数(P_{ij})通过以下公式计算:

$$P_{ij} = C_{ij}/S_{ij} \quad (1)$$

对 pH 值,

$$P_{ij} = (C_{ij} - 7.0)/(8.5 - 7.0) \text{ pH} \geq 7.0)$$

或

$$P_{ij} = (7.0 - C_{ij})/(7.0 - 6.5) \text{ pH} \leq 7.0) \quad (2)$$

式中: P_{ij} 为第 i 组水样第 j 项指标的单项污染指数; $i = 1, 2, \dots, n$, n 为水样数目; $j = 1, 2, \dots, k$, k 为水样指标数目; C_{ij} 为第 i 组水样第 j 项指标的实测浓度; S_{ij} 为第 i 组水样第 j 项指标的标准浓度。

将变换后的资料矩阵称作对应分析矩阵,为方便分析,记作 $X = \{P_{ij}\}_{n \times k}$

1.2.2 主要计算步骤

a. 建立水质监测数据矩阵 $X = \{P_{ij}\}_{n \times k}$,将矩阵按行求和,并求出总和。

$$\text{行和: } x_{i \cdot} = \sum_{j=1}^n x_{ij} \quad (3)$$

$$\text{列和: } x_{\cdot j} = \sum_{i=1}^k x_{ij} \quad (4)$$

$$\text{总和: } T = \sum_{i=1}^k \sum_{j=1}^n x_{ij} \quad (5)$$

b. 对矩阵作对应变换,计算矩阵 $Z = (z_{ij})$

$$\text{其中 } z_{ij} = \frac{x_{ij} - \frac{x_{i \cdot} \cdot x_{\cdot j}}{T}}{\sqrt{x_{i \cdot} \cdot x_{\cdot j}}} \quad (6)$$

c. 作 R 型因子分析:计算矩阵 $A = ZZ'$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$,再按照其累加百分比 $\sum_{\alpha=1}^h \lambda_{\alpha} / \sum_{\alpha=1}^r \lambda_{\alpha} \geq 80\%$ 取其 h 个特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_h$,确定前 h 个公因子,并计算它们的特征向量 $u_1, u_2, \dots, u_h$,R 型因子载荷矩阵为

$$\begin{bmatrix} u_{11} \sqrt{\lambda_1} & u_{12} \sqrt{\lambda_2} & \dots & u_{1k} \sqrt{\lambda_k} \\ u_{21} \sqrt{\lambda_1} & u_{22} \sqrt{\lambda_2} & \dots & u_{2k} \sqrt{\lambda_k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{h1} \sqrt{\lambda_1} & u_{h2} \sqrt{\lambda_2} & \dots & u_{hk} \sqrt{\lambda_k} \end{bmatrix}$$

并在两两因子轴上作指标变量点图。

d. 作 Q 型因子分析:取 $B = Z'Z$,则 B 与 A 具有相同的非零特征根, B 的非零特征根对应的特征向量为 $v_i = Z'u_i$,再将其单位化得 v_i^* ,由此可以得到 Q 型因子载荷矩阵,见表 1,并在两两因子平面图上作水质样点图。

表 1 Q 型因子载荷

样品	G_1	G_2	...	G_h
1	$v_{11}\sqrt{\lambda_1}$	$v_{12}\sqrt{\lambda_2}$	...	$v_{1h}\sqrt{\lambda_h}$
2	$v_{21}\sqrt{\lambda_1}$	$v_{22}\sqrt{\lambda_2}$	...	$v_{2h}\sqrt{\lambda_h}$
...	...	...	...	...
n	$v_{n1}\sqrt{\lambda_1}$	$v_{n2}\sqrt{\lambda_2}$	...	$v_{nh}\sqrt{\lambda_h}$

e. 根据以上所作水质指标变量和样点的主因子平面点聚图,对水质污染状况进行综合评价、解释

和推断。

2 应用案例

2.1 研究区概况

研究区位于吉林省白城市中北部的洮北区。洮北区下辖 8 个乡 5 个镇,156 个居民委员会^[15]。研究区位于洮儿河冲积形成的扇形平原上,主要为剥蚀堆积地貌。地下水补给来源包括降水入渗、河流渗漏和侧向地下水径流补给,以降水入渗占主导地位。

笔者结合 GB/T 14848—93《地下水质量标准》,选取洮北区 13 个水样点($Q_1 \sim Q_{13}$)水质数据中的 9 项指标作为评价因子,即 pH 值、总矿化度(TDS)、总硬度、COD、 Cl^- 、 SO_4^{2-} 、 NO_3^- 、 NO_2^- 、 F^- 。水质样点分布如图 1 所示。

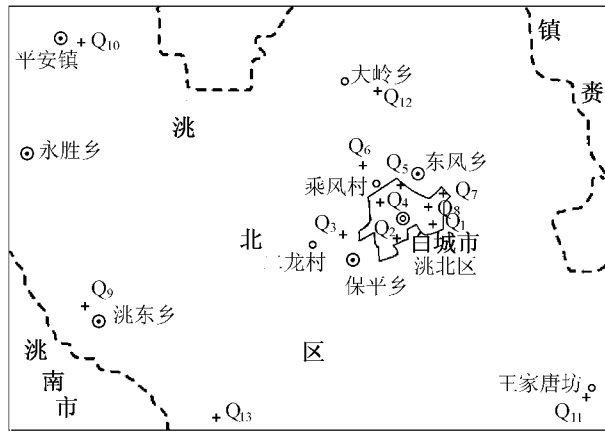


图 1 水质样点分布

2.2 主要计算结果

将原始数据进行处理,并按照 1.2.2 的步骤方法进行分析计算。表 2 是 R 型分析的相关系数矩阵 A 的特征值及方差贡献率。

表 2 R 型分析的相关系数矩阵 A 的特征值及
累计方差贡献率

序号	特征值	百分比	累计方差贡献率/%
1	0.204 73	0.541 89	54.19
2	0.107 55	0.284 66	82.66
3	0.048 04	0.127 16	95.37
4	0.007 94	0.021 03	97.47
5	0.007 09	0.018 75	99.35
6	0.001 82	0.004 81	99.83
7	0.000 61	0.001 62	99.99
8	0.000 03	0.000 07	1.00
9	0	0	1.00

由于前两个特征根的累计方差贡献率达到 82.66%,故确定主因子个数为 2 个,依据计算方法得到 R 型和 Q 型主因子载荷,如表 3 所示。

根据主因子载荷作出变量和样品点的主因子载荷平面投影图,如图 2 所示。

表 3 R 型和 Q 型主因子载荷

变量	R 型主因子载荷		样品	Q 型主因子载荷	
	F_1	F_2		G_1	G_2
Cl^-	-0.048 8	0.040 8	Q_1	-0.094 0	0.040 0
SO_4^{2-}	-0.040 3	0.033 7	Q_2	-0.045 3	0.123 5
NO_3^-	-0.013 4	0.231 0	Q_3	0.027 1	0.035 9
NO_2^-	0.397 9	-0.072 8	Q_4	-0.086 7	0.095 5
F^-	-0.114 6	-0.158 4	Q_5	0.001 5	0.089 8
TDS	-0.083 5	-0.019 5	Q_6	0.365 0	-0.084 1
总硬度	-0.097 0	-0.023 0	Q_7	0.081 3	0.083 8
COD	-0.105 0	-0.090 0	Q_8	-0.028 5	-0.073 3
pH	-0.040 9	-0.109 5	Q_9	-0.100 1	-0.098 7
			Q_{10}	-0.114 5	-0.107 4
			Q_{11}	-0.100 5	-0.172 1
			Q_{12}	-0.106 9	0.004 5
			Q_{13}	0.015 6	-0.042 8

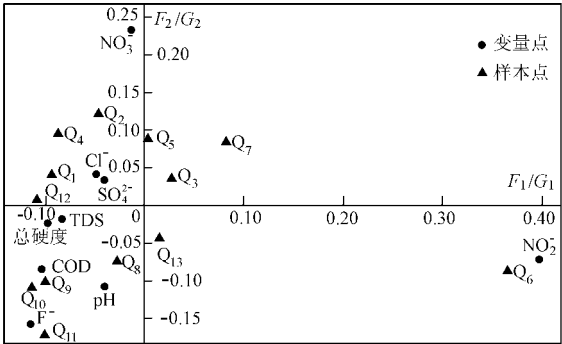


图 2 对应分析因子载荷平面投影图

2.3 分析与结论

根据因子载荷平面投影图(图 2)和因子载荷表(表 3)进行分析,结论如下:

a. 从整体上看,13 个样本点被明显地分为 3 个区。Ⅰ区: Q_6 ;Ⅱ区: $Q_2, Q_1, Q_3, Q_4, Q_5, Q_7, Q_{12}$;Ⅲ区: $Q_9, Q_{10}, Q_{11}, Q_8, Q_{13}$ 。Ⅰ区地下水水质主要受 NO_2^- 影响,Ⅱ区地下水水质主要受 NO_3^- 影响,Ⅲ区位于 F_1 和 F_2 轴负方向区域,地下水水质主要受 F^- 和 COD 的影响。

b. 从变量对主因子轴的影响分析。各离子坐标(载荷)绝对值反映了其与不同主因子轴的相关性程度大小。在 F_1 和 F_2 两个变量主因子轴上,分别有 NO_2^- 和 NO_3^- 的载荷绝对值远大于其他因子载荷,由此可以得出研究区的主要污染因子是 NO_2^- 和 NO_3^- ,说明洮北区地下水主要受到垃圾渗滤液、生活污水等有机污染。

c. 对各主因子轴进行单独分析。在第 1 主因子轴 F_1 上,正向上是 NO_2^- 载荷绝对值最大,负向上是 F^- 和 COD。正向上 F_1 值越大,水体越趋于厌氧环境,负向上 F_1 值绝对值越大,水体受到氟污染和有机物污染越严重。

在第 2 主因子轴 F_2 上,正向上是 NO_3^- 载荷绝对值

最大,负向是 pH。正向上 F_2 值越大,水体越趋于氧化环境,负向上 F_2 值绝对值越大,水体的碱性越强。

从样本水质的角度来说,第 2 主轴也反映了水质整体分异性。根据特征向量符号的取法和 G_2 值的大小,从整体趋势上来说,地下水样本的水质污染程度沿第 2 主轴正方向由轻向重过渡。 Q_6 由于与 NO_2^- 关系密切,水质状况最差。比照分区,可以得出Ⅰ区的地下水水质污染程度最重,Ⅱ区次之,Ⅲ区最轻。

d. 将因子载荷平面投影图 and 水质样本分布图结合起来分析。Ⅲ区的样本点大体分布在农村周边,而Ⅱ区和Ⅰ区则主要位于白城市区和周边城郊。污染最严重的城郊地区由于城市生活垃圾等的随意堆放,一些小型企业污水的随意排放以及配套处理设施的不完善,城乡接合部基础生活设施的落后,致使垃圾渗滤液、工业废水、生活污水等进入地下水中,对水质造成严重影响;白城市内城市卫生环境设施差,部分企业经明渠和渗水井排放的污水、废气等溶于雨水后渗入地下均造成地下水的污染,而城市排污管网的不完善,虽然也是造成地下水受到污染的一个因素,但在一定程度上也防止了地下水遭受污染,因此市区内水质污染程度相对城郊要轻;在农村周边地区,居民生活分散,污水排放量不大,对地下水质的影响较小,水质受污染程度也最轻。

e. 用综合指数法对水样水质进行分析。结果为(括号内数据为水样得分): $Q_9(0.38) \wedge Q_8(0.63) \wedge Q_{13}(0.66) \wedge Q_{10}(0.68) \wedge Q_{11}(0.74)$,属于未污染; $Q_1(1.14) \wedge Q_{12}(1.36) \wedge Q_4(1.54) \wedge Q_3(1.54) \wedge Q_2(2.16) \wedge Q_7(2.46)$,属于轻微污染; $Q_5(2.53) \wedge Q_6(3.06)$,属于中等污染。对应分析分区结果与此基本一致,说明对应分析结果可信,且对应分析根据变量的载荷大小和变量与样本水质分区间的关系,确定出各分区内样本水质的主要污染因子,并结合样本的实际位置分布对样本水质受污染原因进行了分析。

3 结 论

笔者将对应分析方法应用于白城市洮北区地下水水质评价中,参考地下水质量标准确定评价指标,采用单项指数法对原始数据进行处理,根据对应分析原理计算因子载荷,得出对应分析因子载荷图并结合水样点分布图进行分析,结果显示,影响白城市洮北区地下水的污染因子主要是 NO_2^- 和 NO_3^- ,其次是 F^- 和 COD 等。引起本区地下水水质污染的污染物来源主要是垃圾渗滤液和生活污水等。地下水污染程度在农村最轻,市区次之,城郊接合部最重。将对应分析结果与综合指数法评价结果进行对比发现,两种方法对水样点水质污染程度的分区基本一

致。对应分析方法考虑了样本和变量(评价指标)之间的内在联系,对洮北区地下水水质污染状况和原因的分析结论准确可靠,为日后提出有效的防治措施提供了强有力的科学依据。

对应分析方法结合了 Q 型因子分析和 R 型因子分析的优点,简化了数据分析的复杂度,能够较为方便地利用计算机进行分析计算,且运算速度快、结论性强。但是对应分析只对变量和样本间的关系进行了定性研究,未进行定量分析,同时对于既有宏量组分又有微量组分参与的分析分类效果不是太明显,有待于将来进一步研究。

参考文献:

[1] 虞登梅,江晓益.地下水水质评价的人工神经网络方法[J].西安科技学院学报,2003,23(1):27.

[2] 王晓鹏.水质评价的对应分析方法及在湟水流域的应用[J].青海师范大学学报:自然科学版,2001(1):66-69.

[3] 杨悦所.对应分析方法在研究地下水化学成分中的应用[J].水文地质工程地质,1992,19(2):51.

[4] 李保中.对应分析方法在地下水资源评价中的应用[J].勘察科学技术,1986(2):13-15.

[5] 煤炭科学研究院地质勘探研究所,西安矿业学院数学教研室.数学地质基础与方法[M].北京:煤炭工业出版社,1981:482-499.

[6] SUSAN G O, JAMES R C. Correspondence analysis of water quality data: implications for fauna deaths at Stillwater Lakes, Nevada[J]. Math Geol, 1990, 22(6): 665-698.

[7] 苏艺,许兆义,鄯贵权.对应分析方法在地下水环境系统分析中的应用[J].北方交通大学学报,2004,28(4):93-95.

[8] DAVID M, CAMPIGLIO C, DARLING R. Progresses in R-and Q-Mode analysis: correspondence analysis and its application to the study of geological processes[J]. Can J Earth Sci, 1974, 11: 131-146.

[9] 倪福全,宫辉力.对应分析在地下水水质中的应用[J].地下水,1993,15(3):97-100.

[10] EMMANUEL N, AVNER Bar-Henc. Influence function and correspondence analysis[J]. Journal of Statistical Planning and Inference, 2005, 134(1): 26-35.

[11] 金一鸣.西北太平洋海温与东海台风路径的对应分析研究[J].海洋通报,1983,2(1):7-9.

[12] AMBROGI F, BIGANZOLI E, BORACCHI P. Multiple correspondence analysis in S-plu[J]. Computer Methods and Programs in Biomedicine, 2005, 79: 161-167.

[13] 薛觉民,韩春璉.对应分析应用中的量纲问题[J].山西大学学报:自然科学版,1997,20(4):435-439.

[14] 杨炳超.地下水质量综合评价方法的研究[D].西安:长安大学,2004.

[15] 吉林百科全书编纂委员会,中国大百科全书出版社.吉林百科全书[M].北京:中国大百科全书出版社,2003:126-129.

(收稿日期:2009-09-13 编辑:高渭文)