

# 澳门半岛近岸海域水质时空变异分析

何万谦<sup>1,3</sup>, 黄金良<sup>2\*</sup>

(11 清华大学环境科学与工程系, 北京 100084; 2 厦门大学环境科学研究中心, 厦门 361005; 31 澳门民政总署渠务处, 澳门)

摘要: 采用聚类分析和判别分析技术对澳门半岛近岸海域 22 个监测点 2000~2005 年期间的监测数据进行水质时空分析, 旨在识别澳门半岛近岸海域水质时空变化特征, 从而为澳门半岛水质监测网络的优化和近岸海域水质的控制提供支持。监测点位取样时间聚类结果可分为 6~9 月和 1~5 月、10~12 月 2 组, 与传统的 4~9 月和 10~3 月的丰、枯水期划分存在差异; 按采样点位的空间聚类结果显示, 第 1 组的监测站点都位于西侧, 第 2 组的监测站点都位于东侧和南侧。通过后退式判别分析, 表征空间差异性的显著性指标为 pH、氯化物、TSS、颜色; 表征时间差异性的显著指标为总磷、氯化物、颜色、氨氮、溶解氧、化学需氧量。空间与时间差异性判别的正确率分别达到 84.82% 和 76.57%, 显示后退式判别分析具有较好的指标降维能力和判别能力。

关键词: 聚类分析; 判别分析; 水质; 澳门半岛

中图分类号: X834 文献标识码: A 文章编号: 025023301(2010)032060206

## Identification of Spatio-Temporal Variation in the Seaside Water Quality Along Macau Peninsula

HOMAN Wanqian<sup>1,3</sup>, HUANG Jinliang<sup>2</sup>

(11 Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China; 21 Environmental Research Center, Xiamen University, Xiamen 361005, China; 31 The Official Provisional Municipal Council of Macau, China)

Abstract Cluster analysis (CA) and discriminant analysis (DA) were coupled to identify the spatial and temporal variation of the seaside water quality for 14 parameters and 22 different sites along Macau Peninsula so as to identify the temporal and spatial variation for further optimizing the monitoring network and controlling the seaside water quality along Macau Peninsula. The results show that 12 months could be grouped into two periods: June-September and the remaining months, and the entire area was divided into two clusters: one located at western sides of Macau Peninsula and another one is in southeast and southern sides of the Macau peninsula. Through backward stepwise DA, pH, Cl<sup>-</sup>, TSS, color and TP, Cl<sup>-</sup>, color, NH<sub>4</sub><sup>+</sup>, DO, COD were identified as the discriminant variables of spatial and temporal variation, with 84.82% and 76.57% correct assignments, respectively. This study illustrates the application of multivariate statistical techniques is beneficial for understanding the temporal and spatial variation of seaside water quality and further effective water quality management.

Key words: cluster analysis; discriminant analysis; water quality; Macau Peninsula

多元统计分析方法包括聚类分析 (CA)、判别分析 (DA)、主成分分析 (PCA) 等, 是水质水析的有效辅助工具, 被成功地应用于评估河流水质特征<sup>[1~4]</sup>、雨水水质的时空变化<sup>[5,6]</sup>、识别不同城市下垫面降雨径流污染特征的影响因素<sup>[7,8]</sup>等研究。近年来, 针对水质监测点位多、时间序列长, 但缺乏有效信息挖掘的状况, Wunderlin 等<sup>[1]</sup>、Simeonov 等<sup>[2]</sup>、Singh 等<sup>[3,9]</sup>、Shrestha 等<sup>[4]</sup>一些国外学者在阿根廷、希腊、日本和印度等国家开展了基于多元统计技术的河流水质时空分布模式和污染源识别研究。国内近期周丰等<sup>[10~14]</sup>采用聚类分析、判别分析和主成分分析等多元统计分析方法, 在香港河流域和近岸海域进行了成功应用。这些研究显示多元统计分析方法在区域河流与海域水质时空数据分析与信息挖掘方面具

有强大的优势, 至今仍是研究的热点。

澳门近岸部分海域如筷子基南、北湾及游艇会, 偶发大量鱼类死亡。显然, 从整体上掌握澳门近岸海域水质时空分布状况, 对于进一步优化水质监测网络, 控制与管理澳门半岛近岸海域水质具有明显的意义。本研究采用 CA 和 DA 相结合对澳门半岛近岸海域 22 个监测点 2000~2005 年水质数据进行分析, 以期定量识别澳门半岛近岸海域采样时间与采样点位的时空相似性和差异性, 为水质监测网络的优化及水质控制与管理提供科学依据。

收稿日期: 2009204213; 修订日期: 2009210215

基金项目: 福建省青年人才项目 (2007F3093)

作者简介: 何万谦 (1968~), 男, 博士研究生, 主要研究方向为城市水环境污染及下水道生化反应。

\* 通讯联系人, E-mail: jhuang@xmu.edu.cn

## 1 材料与方法

### 1.1 监测点位、取样及分析项目

2000年开始,澳门民政总署渠务处对澳门近岸海域水质进行设置监测点(如图1),每月定期取样分析.监测点位分布大部分在合流渠道有溢流口的排放口附近.取沿岸水深1 m左右处的表层水进行化验.分析的水质参数包括粪大肠杆菌、埃希氏大肠菌、颜色、浊度、pH、电导率、DO、氯化物、NO<sub>2</sub>、NO<sub>3</sub>、NH<sub>4</sub><sup>+</sup>、TP、COD和TSS等14类.分析方法采用标准方法<sup>[15]</sup>.

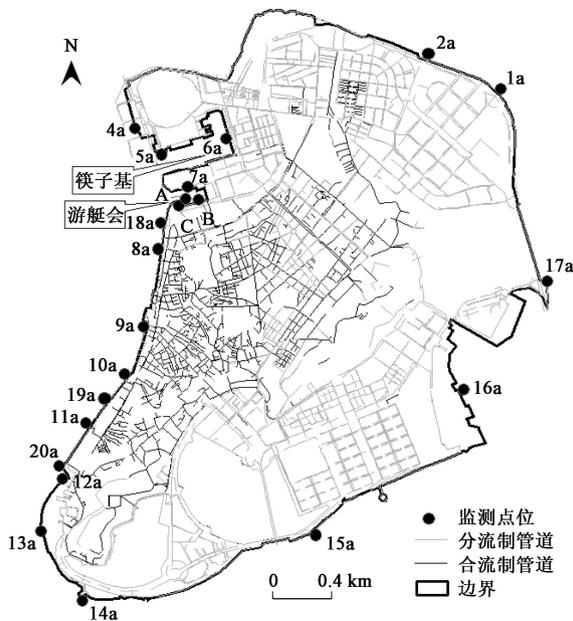


图1 澳门半岛沿岸水质监测点位分布

Fig 1 Location of sampling sites along Macau Peninsula

## 1.2 研究方法

### 1.2.1 数据分析特征

多元统计方法要求污染指标呈正态或接近正态分布<sup>[12]</sup>.所以,当进行CA和DA之前,需要检验污染指标的分布特征.本研究采用Kurtosis和Skewness方法检验.当数据结构过于偏离正态分布时,可以通过对数据转化 $x_c = \lg x$ 使污染指标接近正态分布,提高后续CA、DA多元统计方法的可信度<sup>[3, 10, 16]</sup>.

### 1.2.1.2 聚类分析

聚类分析根据目标间的自然属性和潜在的内在结构进行聚类,是根据事物本身的特性来研究个体分类的统计方法.层次聚类分析(hierarchical CA, HCA)是CA中应用最为广泛的探索性方法,其实质

是根据观察值或变量之间的亲疏程序,以逐次聚合的方法,将最相似的对象结合在一起,直到聚成一类<sup>[10, 16]</sup>.亲疏程度的计算包括2类:样本间距离和类间距离.本研究采用HCA进行时空相似性分析,采用的计算方法是欧氏距离平方和离差平方方法.

### 1.2.1.3 后退式判别分析

DA方法可通过判别函数判别CA分析结果和识别显著性的污染指标<sup>[1, 3, 16]</sup>.此方法分为3类:标准式、前进式和后退式.相比之下,后退式DA方法具有更好的指标降维能力和判别能力<sup>[1, 3, 10, 16]</sup>.所以本研究用后退式DA方法进行时空差异性分析,并采用交叉验证法来检验此方法的判别能力.相应的判别函数表达式为:

$$f(G_i) = k_i + \sum_{j=1}^n w_j P_j \quad (1)$$

式中,  $i$  为类型  $G$  的个数,  $n$  为污染指标个数,  $w_j$  为权重系数,  $P_j$  为显著性污染指标的浓度,  $f$  为判别函数值,  $k_j$  为常数.

## 2 结果与讨论

### 2.1 基础数据预处理

对原始数据进行Kurtosis和Skewness检验污染指数分布特征,结果分别为41247~6061629, -0.1714~241626, 太过于偏离正态分布,所以对数据进行对数处理,处理之后Kurtosis和Skewness分别为-0.1695~1.7106和-1.1831~0.1733.虽然还不是正态分布,但除个别污染指数外已经大大接近正态分布.最后将对数后的数据进行标准化处理,使其均值为0,方差为1.

### 2.1.2 时空相似性聚类分析

按照监测点位每个月的平均值,对12个月进行聚类,图2为其聚类结果,可以看出,在 $(D_{link}/D_{max}) @100 < 22$  ( $D_{link}/D_{max}$ 表示个案连锁距离与最大连锁距离之比)处可分为2组,第1组为6~9月,第2组为1~5月和10~12月.澳门全年降水量83%集中在4~9月,因此,这个聚类的结果基本上是对应于澳门的枯水期/丰水期,同时也说明如果直接按照枯水期/丰水期或是季节来进行水质监测会有一定的偏差.

考虑到时间尺度,对按时间聚类的第1组和第2组以及整体都进行空间的聚类,其结果基本相似,所以本研究采用整体样本进行空间聚类,聚类结果如图3所示.在 $(D_{link}/D_{max}) @100 < 12$ 处可分为2组,其中第1组包括a1、a2、a14、a15、a16、a17,

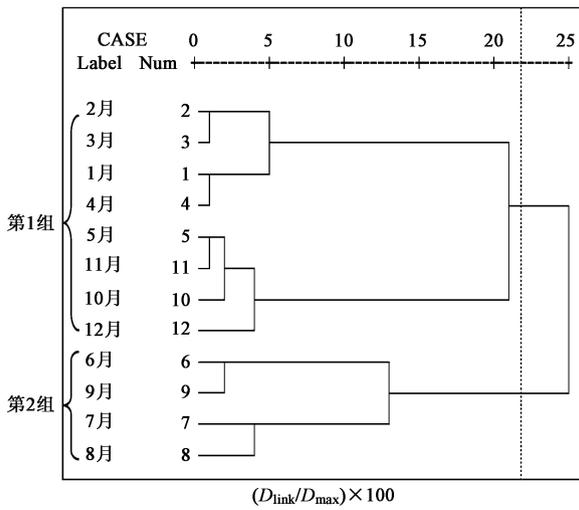


图 2 时间尺度聚类分析树形图

Fig 2 Temporal cluster analysis of monitoring periods based on Ward's method

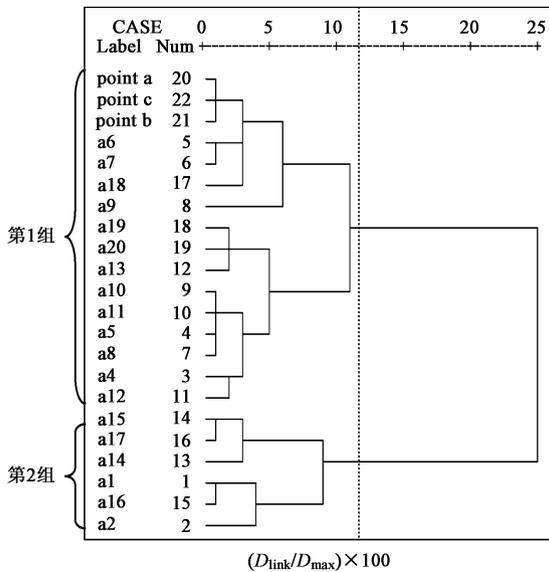


图 3 整体空间聚类分析树形图

Fig 3 Spatial cluster analysis of monitoring sites based on Ward's method

第 2 组包括 point a point b point c a4~ a13 a18 ~ a20 第 1 组的监测站点都位于澳门半岛东侧和南侧, 第 2 组的监测站点都位于西侧. 从各监测站点的水质基本情况可以看出, 监测点位分布在澳门半岛西侧的水质较东侧和南岸差, 与西侧位于老城区, 排水管网为合流式 (见图 1), 污染物容易溢流出海, 且西侧与珠海湾仔形成的狭窄河道不利污染物扩散有关. 而澳门半岛东侧和南侧为新城, 面对广阔海域, 水较澳门半岛西侧海域深, 有利于污染物稀释与扩散, 且有关区域的排水管网为分流式 (如图 1 所示), 污染物溢流出海的情况较小. 由此可见, 基于 CA 的空间聚类结果与实际情况相符. 对于水质监测的进一步启示即为: 今后近岸海域水质监测可以进一步优化水质监测网络, 在空间聚类划分的 2 组 (即西侧与东、南侧) 的监测点中选择若干有代表性的采样点, 这样既能达到监测的目的, 又可以降低成本.

2.1.3 时间差异性判别分析

2.1.3.1 空间差异性判别

针对整体空间聚类结果获得的 2 组类型, 将原始数据进行后退式判别分析. 在后退式判别分析的过程中以  $F = 2$  为下限剔出变量, 所得结果只剩下 pH、氯化物、TSS 3 个参数, 但对于监测样本分类判别正确率只有 43.18%. 因此, 如表 1 所示, 考察 4 个和 5 个参数的结果, 考虑到氨氮是一个重要的参数, 选用 5 个参数即 pH、氯化物、TSS、颜色和氨氮, 此时正确率达到 84.18%, 具有较好的指标降维能力和判别能力. 同时也说明虽然颜色和氨氮对于整体的显著性不高, 但对判别分析的结果的正确率具有较大影响.

图 4 列出 5 个显著性污染指标的空间差异, 第 1 组的 pH、颜色、氯化物和 TSS 都高于第 2 组, 而氨氮则小于第 2 组. 第 1 组监测站点位于澳门半岛的东侧和南侧, 是开阔的海域, 第 2 组监测站位于西

表 1 空间尺度后退式判别分析结果比较

Table Classification matrix for backward DA of spatial variations

参数个数	参数类别	Wilks	F	正确率 %
3 个参数	pH、氯化物、TSS	0.1628 5	1171.526 4	43.189
4 个参数	pH、氯化物、TSS、颜色	0.1620 4	911.167 63	86.147
5 个参数	pH、氯化物、TSS、颜色、氨氮	0.1614 8	741.621 54	84.182

侧, 与珠海湾仔形成狭窄的河道. 因此第 1 组监测站点受到海水的影响更大, 氯化物的含量也就更高. 另

外, 第 1 组采样点海域开阔, 珠江口的沿岸流、波浪等带来大量的泥沙, 导致该区域水体中 TSS 较高. 而

第2组采样点处于澳门半岛西侧位于老城区, 合流制排污口密集(图1), 生活污水的溢流排放较多, 氨氮作为主要源自于生活污水中含氮有机物受微生物作用的分解产物<sup>[17]</sup>, 其含量自然较高. 西侧海域由于受生活污水影响较大, 且西侧海域呈现小的河道, 不利于污染物的扩散, 令污染物积累, 总磷、氨氮增加, 水体中氧气较东、南侧开阔海域少, 较易形成有

机酸, 从而使水体中的 pH 值较小<sup>[4]</sup>. 从这5个具有空间显著差异的污染指标来看, 可以进一步估计污染相对严重的区域(西侧)污染物主要是与生活污染相关的营养盐, 而污染相对较轻的区域(东侧和南侧)的污染物主要是悬浮颗粒物, 所以有必要针对性地进行污染控制, 并建议增强显著污染指标的监测, 包括频率、点位和精度.

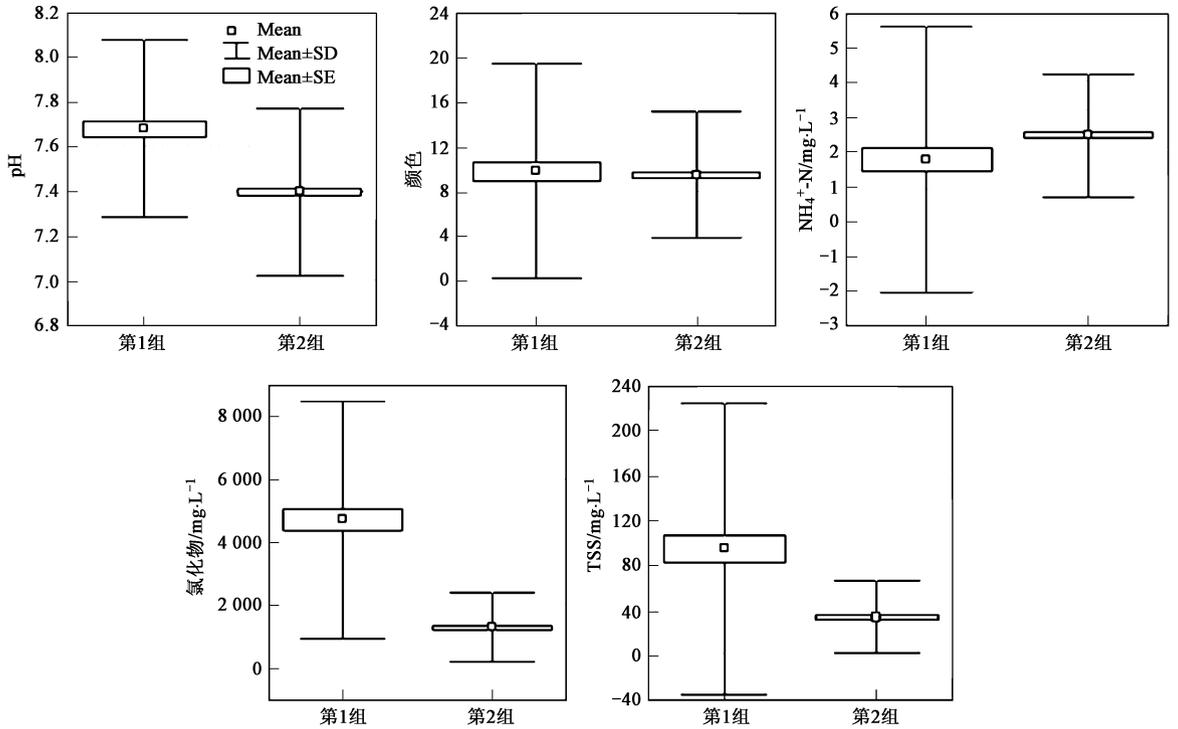


图4 显著性污染指数的空间差异性

Fig 4 Spatial variations of discriminant parameters derived from backward DA

此外, 后退式判别分析的结果可以指导该区域今后的海水水质的监测网络设计, 有针对性地选择这5个显著性的污染指标, 来反映空间尺度上水质的差异性, 从而达到优化和节约监测成本的目的.

2.1.3.2 时间差异性判别分析

用时间与参数的相关性矩阵来评价水质参数的时间差异, 除了粪大肠杆菌、埃希氏大肠菌、pH、亚硝酸盐、硝酸盐, 其他参数都与时间显著相关 ( $p < 0.05$ ). 氯化物的显著相关系数 (Spearman's  $R = 0.152$ ) 最大, 其后依次是颜色 ( $R = 0.128$ ), 溶解氧 ( $R = 0.126$ ), 电导率 ( $R = -0.119$ ), 总磷 ( $R = 0.113$ ), 氨氮 ( $R = 0.112$ ), TSS ( $R = -0.112$ ), 浊度 ( $R = -0.111$ ), 化学需氧量 ( $R = -0.108$ ). 这些参数是造成水质时间差异的主要因素.

针对时间聚类所划分的2组数据, 把原始数据进行后退式判别分析. 在后退式判别分析的过程中以  $F = 2$  为下限剔出变量, 所得结果为只剩下总磷、氯化物、颜色、氨氮4个参数, 但对于监测样本分类判别正确率只有63.104%. 因此, 如表2所示, 分别考察5个和6个参数的结果, 选用6个参数, 即: 总

磷、氯化物、颜色、氨氮、溶解氧、化学需氧量

表2 时间尺度后退式判别分析结果比较

Table 2 Classification matrix for backward DA of temporal variations

参数个数	参数类别	Wilks	F	正确率/%
4个参数	总磷、氯化物、颜色、氨氮	0.1809	351.17800	63.104
5个参数	总磷、氯化物、颜色、氨氮、溶解氧	0.17977	301.20425	74.126
6个参数	总磷、氯化物、颜色、氨氮、溶解氧、化学需氧量	0.17905	261.28137	76.157

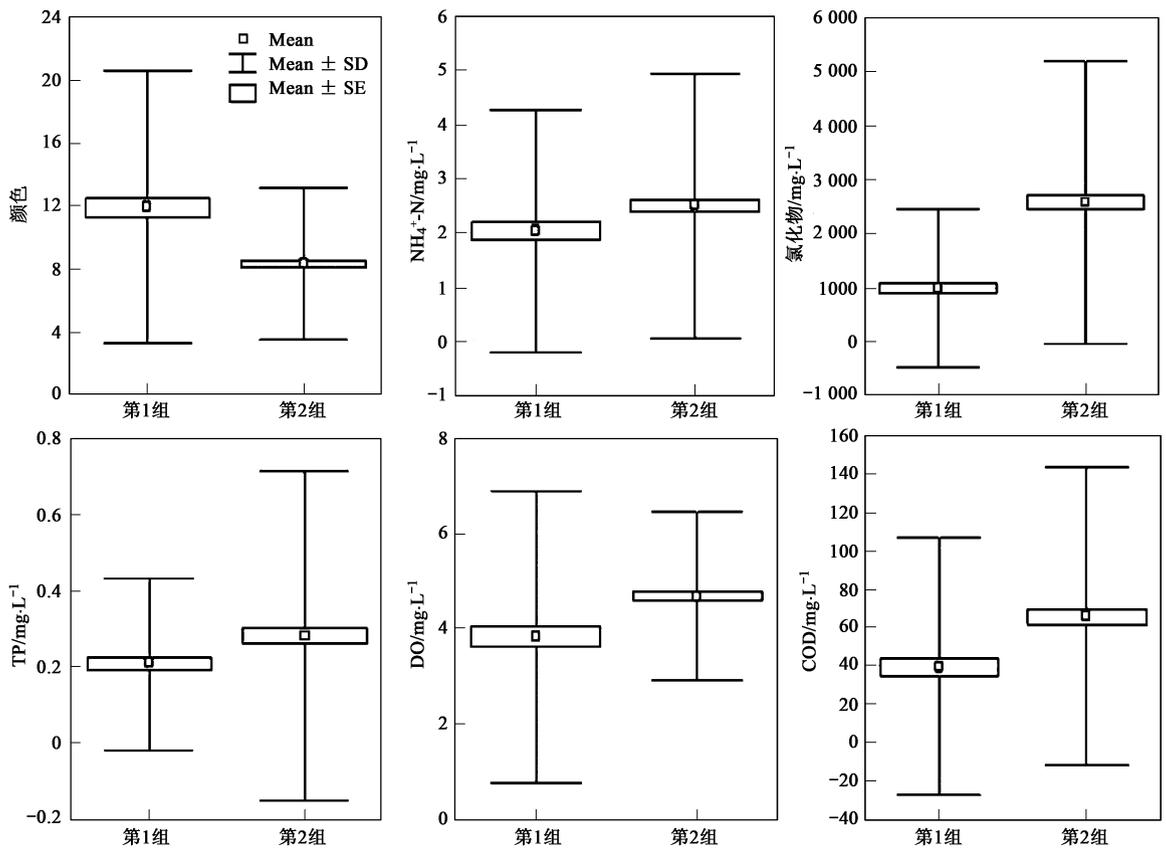


图5 显著性污染指标的时间差异性

Fig 5 Temporal variations of discriminant parameters derived from backward DA

磷、氯化物、颜色、氨氮、溶解氧、化学需氧量,此时正确率达到 7615%, 具有较好的指标降维能力和判别能力. 同时也说明虽然溶解氧、化学需氧量对于整体的显著性不高, 但对判别分析的结果的正确率还是有很大影响.

从图 5 可知, 在时间聚类结果的显著性污染指标中, 第 1 组的颜色较高, 而氨氮、氯化物、总磷、溶解氧、化学需氧量是第 2 组比第 1 组高. 第 1 组是处在丰水期, 雨水冲刷强度较大和持续时间较长, 陆源带来的泥沙较大, 水质混浊, 因此颜色较深. 相反第 2 组是处在枯、平水期, 雨水量较小, 主要受到海水的影响, 所以氯化物含量高. 由于溶解氧、化学需氧量、氨氮和总磷的主要污染来源是生活污水, 生活污水的量在各个时期没有大的变化, 水量减少, 浓度则变大, 水质则变差.

### 3 结论

(1)按监测点位取样时间聚类结果显示, 可分为 6~9 月和 1~5 月、10~12 月 2 组, 与传统的 4~9 月和 10~3 月的丰、枯水期划分存在差异; 按采样

点位的空间聚类结果显示, 第 1 组的监测站点都位于西侧, 第 2 组的监测站点都位于东侧和南侧.

(2)通过后退式 DA 分析, 表征空间差异性的显著性指标为 pH、氯化物、TSS、颜色; 表征时间差异性的显著指标为总磷、氯化物、颜色、氨氮、溶解氧、化学需氧量. 空间与时间差异性判别的正确率分别达到 8418% 和 7615%, 显示后退式 DA 具有较好的指标降维能力和判别能力.

致谢: 感谢澳门民政总署李梅香女士、北京大学环境学院周丰博士在监测数据分析方面提供的帮助.

参考文献:

- [1] Wunderlin D A, Díaz M D P, Ame M V, et al. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality: A case study: Suquia River Basin (Cordoba, Argentina) [J]. *Water Research* 2001, 35 (12): 2881-2894.
- [2] Simeonov V, Stratis V, Samara J A, et al. Assessment of the surface water quality in Northern Greece [J]. *Water Research* 2003, 37: 4119-4124.
- [3] Singh K P, Mali A, Mohan D, et al. Multivariate statistical

- techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) a case study [J]. *Water Research*, 2004, 38: 3980-3992
- [ 4 ] Shrestha S, Kazama F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan [J]. *Environmental Modelling & Software*, 2007, 22: 464-475.
- [ 5 ] Vazquez A, Costoya M, Pena R M, et al. A rainwater quality monitoring network: a preliminary study of the composition of rainwater in Galicia (NW Spain) [J]. *Chemosphere*, 2003, 51: 375-386.
- [ 6 ] Hamers T, van den Brink P J, Mos L, et al. Estrogenic and esterase-inhibiting potency in rainwater in relation to pesticide concentrations, sampling season and location [J]. *Environmental Pollution*, 2003, 123: 47-63.
- [ 7 ] Goonetilleke A, Thomas E, Ginne S, et al. Understanding the role of land use in urban stormwater quality management [J]. *Journal of Environmental Management*, 2005, 74: 31-42.
- [ 8 ] Huang J, Du P, Ao C, et al. Multivariate analysis for stormwater quality characteristics identification from different urban surface types in Macau [J]. *Bulletin of Environmental Contamination and Toxicology*, 2007, 79: 650-654.
- [ 9 ] Singh K P, Malik A, Sinha S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study [J]. *Analytical Chimica Acta*, 2005, 538: 355-374.
- [ 10 ] Zhou F, Liu Y, Guo H. Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern new territories, Hong Kong [J]. *Environmental Monitoring Assessment*, 2007, 132: 121-3.
- [ 11 ] Zhou F, Guo H, Liu Y, et al. Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong [J]. *Marine Pollution Bulletin*, 2007, 54: 745-756.
- [ 12 ] Zhou F, Huang G H, Guo H, et al. Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong [J]. *Water Research*, 2007, 41: 3429-3439.
- [ 13 ] 周丰, 郭怀成, 刘永, 等. 基于多元统计分析和 RBF NNs 的水质评价方法 [J]. *环境科学学报*, 2007, 27(5): 846-853.
- [ 14 ] 周丰, 郭怀成, 黄凯, 等. 基于多元统计方法的河流水质空间分析 [J]. *水利学报*, 2007, 18(4): 544-551.
- [ 15 ] American Public Health Association, American Water Works Association, and Water Environment Federation. Standard methods for the examination of water and wastewater [M]. Washington D C: Am Public Health Assoc, 1992.
- [ 16 ] Lattin J, Carroll D, Green P. *Analyzing multivariate Data* [M]. New York: Duxbury Press, 2003.
- [ 17 ] 奚旦立, 孙裕生, 刘秀英. *环境监测* [M]. 北京: 高等教育出版社, 1999.