两阶段模糊法在海河水系水质评价中的应用

邹志红*, 云逸, 王惠文

摘要: 两阶段模糊法用于海河水系监测断面的水质评价, 与传统模糊聚类法不同, 该方括首先进将国家地表水指标转化为等级打分, 再构造模
糊隶属度矩阵, 确定评价对象对标准水质水平的隶属度, 之后运用改进的模糊 e-均值法 (FCM 矩) 选取适宜模糊聚类和先验类中心, 对评价对象
进行聚类, 得到群类中水质指标的值域; 组织了 0~100 之间, 以消除现实评价, 计算了海河水系 2004~2006 年 4 个断面的 432 个数据
, 推算出季节波动规律和不同年度变化的情况, 对水质进行了综合评价, 两阶段模糊法对水质评价综合指标有较好的分类能力, 能够得到
分析出的水评价结果, 且计算过程相对 FCM 简单, 可用于河流水质综合评价。

关键词: 模糊聚类; FCM; 水质评价; 海河水系


Application of two-stage fuzzy set theory to water quality evaluation in the Haihe River system

ZOU Zhihong*, YUN Yi, WANG Huixwen

School of Economics and Management, Beijing University of Aeronautics and Astronautics, Beijing 100083

Received 17 June 2007; revised 27 December 2007; accepted 29 January 2008

Abstract: Two-stage fuzzy set theory was used to evaluate the water quality in Haihe River system monitoring sections. Unlike traditional fuzzy clustering methods, the national surface water indexes were first transferred into ranking marks. Fuzzy membership functions were formed to determine the membership degree of evaluation objects for each standard water level. The improved fuzzy c-means (FCM) methodology was used afterwards, and the proper fuzzy index and precedent class center were chosen for classification. Then an overall index, the River quality index (RQI), was constructed to represent the original monitoring data. The value of RQI was designed by the interval [0, 100], to make intuitive evaluation. Altogether, 432 data points from 4 monitoring sections in the Haihe River system during the years 2004~2006 were calculated. Seasonal fluctuations and different situations among years were considered, and the water quality was synthetically evaluated. Two-stage fuzzy set theory is capable of classifying water quality. Accurate scientific results are obtained, and the calculation process is relatively simple compared with usual TCM, so it can be applied in water quality evaluation in Chinese river systems.

Keywords: fuzzy clustering; FCM; water quality evaluation; Haihe River

1 引言 (Introduction)

水质评价的定量化对我国水资源的可持续利用有着十分重要的意义(彭志远等, 2004), 评价结果可确定目标水体水域的客观质量水平, 并为今后的水质保持或污染治理提供可靠依据。

国内外目前在水质评价方面的研究成果颇丰。概括起来, 国外的水质评价研究比较注重多介质、多参数的水质数据研究, 而我国水质评价则集中于对评价因子处理的应用, 不够突出研究水域的背景信息, 存在人为水质分类的特点 (邱志红等, 2002)。因此, 如何使评价方法更具科学性和合理性是本研究中进行水质评价的出发点。

水质评价数据的最大特点是随机性, 基于模糊数学原理所建立的模糊综合评价方法是一种解决数据随机性的可靠程序 (秦寿康, 2003), 它已被广
泛地应用于水质评价（孙靖等，2005；邹志红等，2005；Zou et al., 2006）。另外，模糊聚类理论也可应用于环境数据的分析评价（Tao, 1998）。模糊方法被证明适合用于环境指标，因为它可以解决不相容数据或是应用于数据缺陷的决策情况，并将定性的科学测量结果与抽象的社会目标很好地结合在一起（Silvert, 2000）。

模糊C-均值法（FCM算法）是模糊聚类理论中的一种常用方法，可以处理具有模糊性数据的聚类问题（刘增等，1998），但在环境方面的应用尚不多见。Tan等人将模糊C-均值法应用于土壤中重金属的评价（Tan et al., 2006）。Liou等人则将改进后的模糊C-均值法应用于台湾地区河流水质评价（Liou et al., 2006）。本研究中拟将此方法结合中国水质评价体系应用于海河流域的水质评价，期望为其在水质评价中的应用提供参考。

2 资料来源与方法（Materials and methods）

2.1 背景简介

海河流域地跨北京、天津、河北、河南、山东、山西、内蒙及辽宁8省（自治区、直辖市），人口密集，大中城市众多，是我国政治文化中心和经济发达地区。海河水系由北运河、永定河、大清河、子牙河、南运河5大支流水系组成，众多河流以幅射状汇聚京津，东流入海（陈家山，2005）。

在国家环保总局颁布的《全国主要流域重点断面水质自动监测通报》中，属于海河水系的断面有8个，我们选取其中的4个断面作为研究对象（见图1），分别为北京密云古北口潮河（密云水库入口）、北京门头沟沿河城永定河（官厅水库出口）、天津滨海新区潮河（于桥水库入口）和天津三岔口潮河（入海口）。这4个断面主要位于北三河与永定河流域，总体上反映了京津地区的水质状况。图中四圈标示了4个断面的大体位置。本研究中采用2004～2006年《全国主要流域重点断面水质自动监测通报》的所有数据，每个监测断面选取3个监测指标（溶解氧（DO）、高锰酸盐指数（COD_{Mn}）和氨氮（NH\textsubscript{3}-N）），进行月平均计算处理后共得到432个数据进行水质评价。

2.2 模糊聚类 FCM算法（Liou et al., 2006）

传统的FCM算法将含有n个监测面的数据或者向量划分为c类（c≤n），通过迭代过程使目标函数J_{m}最小化，标准方程为：

$$J_{m}(U,V,X) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^{m} ||x_i - \mu_k||^2$$

$$1 \leq m \leq \infty$$

（1）

式中，$x_i \in \mathbb{R}^p$ 代表第k个监测点的数据，$p$为监测指标数；$X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{n \times p}$ 表示整个监测数据的矩阵，即有n个监测面，每个监测点测量包含p个指标数据；隶属度 $\mu_{ik} \in [0, 1]$ 代表了观测值$x_i$属于第$i$类的相似度；$U \in \mathbb{R}^{n \times c}$ 是相似度矩阵；$\mu_{ik} \in \mathbb{R}^c$ 是第$i$个聚类标准值（i=1, ..., c）；$V \in \mathbb{R}^{p \times c}$ 是分类质心矩阵；$||x_i - \mu_k||$表示距离函数，当所有观测值的协方差等同于单元矩阵$I$时，距离矩阵就变成欧氏距离，$m \in [1, \infty]$，被称为模糊指标，决定了分类的模糊程度。在以下限制下最小化$J_{m}$就产生迭代最小拟算法：(1) $0 \leq \mu_{ik} \leq 1 \forall k, i$；(2) $\sum_{i=1}^{n} \mu_{ik}^{m} = 1 \forall k$；(3) $0 \leq \sum_{k=1}^{c} \mu_{ik}^m < n$，即FCM算法。

2.3 两阶段模糊法

两阶段模糊法是基于FCM方法的拓展（Liou et al., 2006）。因为相似度衡量了观测对象和所定义的标准水平的相似程度，则将观测值$x_i$对于所有质量水平的相似度 $\mu_{ik}(i=1, \ldots, c)$ 加起来，便可得到替代原观测值的一个全局质量指标，称为水质指标，符号定义为$I_{SO}$（the River quality index）。观测值和良好质量水平共性越大，则它的全局值得到的越大。

设溶解氧（DO）、高锰酸盐指数（COD_{Mn}）、氨氮（NH\textsubscript{3}-N）3个监测指标为变量，构造三维变量 $X = \{DO, COD_{Mn}, NH\textsubscript{3}-N\}$。
(x_d, x_M, x_N), 以及这些变量的分段线性隶属函数，下标 D 代表溶解氧, M 代表高锰酸盐指数, N 代表氨氮. 各个污染水平指标即为国家标准分级数据. 3 个分段线性隶属函数分别为 f_D(x_D), f_M(x_M), f_N(x_N):

\[
\begin{align*}
  f_D(x_D) &= \begin{cases} 
  1 & x_D > 7.5 \\
  0.75 + 0.25 \times \frac{x_D - 6}{7.5 - 6} & 6 < x_D \leq 7.5 \\
  0.5 + 0.25 \times \frac{x_D - 5}{6 - 5} & 5 < x_D \leq 6 \\
  0.25 + 0.25 \times \frac{x_D - 3}{5 - 3} & 3 < x_D \leq 5 \\
  0.25 \times \frac{x_D - 2}{3 - 2} & 2 < x_D \leq 3 \\
  1 & x_D \leq 2
  \end{cases}
\end{align*}
\]

\[
\begin{align*}
  f_M(x_M) &= \begin{cases} 
  1 & x_M > 4 \\
  0.75 + 0.25 \times \frac{4 - x_M}{4 - 2} & 2 < x_M \leq 4 \\
  0.5 + 0.25 \times \frac{6 - x_M}{6 - 4} & 4 < x_M \leq 6 \\
  0.25 + 0.25 \times \frac{10 - x_M}{10 - 6} & 6 < x_M \leq 10 \\
  0.25 \times \frac{15 - x_M}{15 - 10} & 10 < x_M \leq 15 \\
  1 & x_M \leq 15
  \end{cases}
\end{align*}
\]

\[
\begin{align*}
  f_N(x_N) &= \begin{cases} 
  1 & x_N > 0.15 \\
  0.75 + 0.25 \times \frac{0.5 - x_N}{0.5 - 0.15} & 0.15 < x_N \leq 0.5 \\
  0.5 + 0.25 \times \frac{1 - x_N}{1 - 0.5} & 0.5 < x_N \leq 1 \\
  0.25 + 0.25 \times \frac{1.5 - x_N}{1.5 - 1} & 1 < x_N \leq 1.5 \\
  0.25 \times \frac{2 - x_N}{2 - 1} & 1.5 < x_N \leq 2 \\
  0 & x_N \leq 2
  \end{cases}
\end{align*}
\]

式中 \( f_i \) 为数据从浓度为质量隶属度的值, 标准值 \( z_i \) 为预定水质标准水平, \( \mu_{ai} \in [0, 1] \) 为 \( f_i \) 对第 \( i \) 号水质水平的相似度指数. 由 0 和 1 构成的 5 维测量空间 \( f_i \) (k = 1, ..., n), 即终值为 \( f_i = (0, 0, 0, 0) \) 至 \( f_i = (1, 1, 1, 1) \), 被用来计算相似度和 \( I_{aq} \). 选取欧氏距离作为距离函数. \( z_i \) 为标准质量水平, 根据国家水质标准取 \( z_i = (0, 0, 0) \), \( z_1 = (0.25, 0.25, 0.25, 0.25) \), \( z_2 = (0.5, 0.5, 0.5, 0.5) \), \( z_3 = (0.75, 0.75, 0.75) \), \( z_4 = (1, 1, 1) \), 其中 \( z_4 = (0, 0, 0) \) 和 \( z_1 = (1, 1, 1) \) 表示了标准质量水平中的 I 类水和 V 类水; 这样就可以计算观测值 \( z_i \) 到各标准水平之间的相似度 \( \mu_{ai} \), 取值在 [0, 1] 区间内.

目标观测值 \( z_i \) 的全局指标计算为:

\[
I_{aq} = \left( \sum_{i=1}^{5} (\mu_{ai} \times q_i) \right) \times 100
\]

式中, \( q_i \) 为每个质量水平的权重, 水质等级较差的应被赋予更大的权重水平, 因此, \( q_1 = 0, q_2 = 0.25, q_3 = 0.5, q_4 = 0.75, q_5 = 1 \). \( I_{aq} \) 的取值为 0 ~ 100.

3 实例研究 (Case study)

以海河水系 4 个断面 2004 ~ 2006 年所有周报告数据作为研究对象，首先将数据进行处处理，计算 1 个断面每段数据的平均值作为评价对象。这样做是为了更好地反映水质波动状况，而且弥补个别周报告数据缺乏的情况而不影响评价结果。如果某月因河流断流整个月份没有数据，则该月无效，如果有一个月或者几个四周有数据，则只取这几个周的平均值。作图时遵循这一合理假设，即无数据的月份的 \( I_{aq} \) 值用上下 2 个月的平均值来代替，只为表示水质水平的大致趋势，否则在折线图中不利于观察与比较，而且这样的点只是少数。

3.1 不同年份同一监测断面水质评价结果

图 2 显示了密云、口头河、果河桥和三岔口 4 个断面各自在 2004 ~ 2006 年 12 个月 \( I_{aq} \) 值的情况。从图中可以看出，密云、果河桥和三岔口这 3 个断面水质有明显的季节性波动，这是因为前两者为水库的入水，三岔口为海河入海口；三个断面的水质变化趋势相似，水质较差的时期都发生在夏季汛期，这是由于夏季华北地区进入雨季，位于河流中
上游的密云和果河桥监测断面水质最差的时间点与三岔口监测断面发生要早一些，这也符合实际的地理位置引起的差异。门头沟监测断面虽比1974年有显著改善，也没有进一步恶化的趋势。

以上分析可以看出，$I_{BQ}$数值代表的水质情况很好地反映了之前定性分析的结果。表达直观、易于理解，更便于对数据进行归纳比较。

密云和门头沟监测断面分别代表由密云水库和官厅水库进入北京的水质。从图2可知，北京水源地的水质较好，天津果河桥监测断面的数值则说明来自于桥水库的天津水源的水质也较好。河流经过京东西两侧，入海口的三岔口监测断面处水质变得比较恶劣。这主要由于城市与郊区各种工业农业污染的注入，因此，对各种污染源和非污染源的控制仍需十分必要。

### 图2 4个断面2004-2006年月平均数据$I_{BQ}$值折线图

![图2 4个断面2004-2006年月平均数据$I_{BQ}$值折线图](image)

**Fig. 2 Monthly average $I_{BQ}$ value broken line graphs during 2004-2006 for 4 monitoring sections**

### 3.2 同一年份不同监测断面水质评价结果

将4个断面2006年全年的$I_{BQ}$值绘制折线图（图3），从图中可以看出，密云和果河桥监测断面的水质基本在Ⅰ～Ⅲ类，而门头沟监测断面处水质较前两者差，说明水质为Ⅳ～Ⅴ类的时间增长；而三岔口监测断面水质最差，水质降为劣Ⅴ类的时间较多。通过比较可知，天津入海口处附近的水质需要密切关注，并注意控制上游污染源，以达到切实提高河河水系整体水质的目的。门头沟监测断面反映的官厅水库水质出现轻微恶化的趋势，也应引起一定的重视。

### 图3 2006年各监测断面数据$I_{BQ}$值折线图

![图3 2006年各监测断面数据$I_{BQ}$值折线图](image)

**Fig. 3 Monthly average $I_{BQ}$ value broken line graphs of 4 monitoring sections in year 2006**
4 结论（Conclusions）

1) 选取了海河水系 2004~2006 年 4 个断面的 432 个数据，运用两阶段模糊法对影响京津地区主要水源水质进行了综合评价。结果表明，海河水系总体仍属于劣 V 类水质，京津地区总体水质随季节波动规律明显，控制污染势在必行。

2) 两阶段模糊法是一种拓展的 FCM 法，对处理具体明确取样的水质综合评价问题十分有效，可以根据评价标准区间值对评价对象实现较好的分类。同时它将所有水质水平量化成一个综合指标，可用于时间轴和评价对象之间的比较分析，直观清晰，易于更广泛人群的了解认知。

3) 实例证明，该方法在我国现行水质评价体系下十分适用，可以预先确定水质评价分类级别，有利于模糊聚类的准确性。

References:


Tao Y. 1998. Fuzzy comprehensive assessment; fuzzy clustering analysis and its application for urban traffic environment quality evaluation [J]. Transport Research, 3 (1); 51—57


中文参考文献:

陈茂山. 2005. 海河流域水环境变迁与水资源承载力的历史研究 [D]. 北京: 中国水利水电科学研究院, 16—18


刘增良. 1998. 模糊技术与应用选编 [M]. 北京: 北京航空航天大学出版社, 447—450

彭文启, 周怀东, 郑晓雯, 等. 2004. 三次全国地表水水环境评价综合水环境保护, 1: 37—39

秦寿根. 2003. 综合评价原理与应用 [M]. 北京: 电子工业出版社, 1—8

申献辰, 郑晓雯, 杜震. 2002. 中国地表水资源质量评价方法的研究 [J]. 水利学报, 12; 63—67


孙志红, 孙振南, 任广平. 2005. 模糊评价因子的熵权法赋权及其在水质评价中的应用 [J]. 环境科学学报, 25 (4); 552—556