

武汉理工大学

数学建模暑期培训论文

第 5 题

基于投影寻踪法的黄河健康状态评价

第 1 组

姓名	方向
阮滨（组长）	编程
王宇	建模
王家柯	写作

2020 年 8 月 24 日

摘要

黄河作为宁夏地区的主要供水水源，其水生态保护具有重要意义。本文通过**经验正交函数分解**提取了水污染的时空规律，运用**模糊综合评价法**对断面水质进行综合评级，最后使用**实数编码加速的遗传算法**求解**投影寻踪评价模型**，得到了各断面处黄河水生态系统评分。

针对问题一，从水质的**综合污染情况**和**单一指标超标情况**分别对黄河宁夏河段水质进行分析。在综合污染情况的分析中，本文在国标规定的以最劣指标确定水质等级的基础之上，通过**模糊综合评价**对当前水质的整体情况以及与目标水质之间的差距进行补充说明；在各指标超标情况的分析中，本文以目标水质对应各指标的下限作为阈值，统计了水污染指标超标的种类以及次数。发现虽然较多断面水质未达国家标准，但实际上综合水质情况较为乐观，仅有个别指标不达要求，后续可以采取针对性措施的对水质进行治理。

针对问题二，本文通过**经验正交函数分解法**对黄河宁夏河段水质的时空规律进行分析。考虑到**枯水期**和**丰水期**黄河流量的较大差异，本文首先将数据按此标准进行划分，基于**经验正交函数分解法**分别对两个时期的**时间模态**和**空间模态**进行提取，通过**Mann-Kandell 检验**得到各水质指标在时间维度的趋势，并运用**North 检验**验证了经验正交函数分解法的可信度。从时间维度分析发现从1月到8月，**氨氮浓度**、**生化需氧量**等四个指标呈增加趋势，而**铜**和**硫化物**具有减少趋势；而在空间维度发现整体水质由上游到下游逐渐变好，认为与国家对于出境水质的监管和处罚规定有关。

针对问题三，综合考虑黄河水生态中的各要素，构建指标体系，并建立了**投影寻踪模型**进行综合评级。首先通过查阅资料，从**水质**、**水量**以及**人类开发**等六个维度出发，选择了包括**河岸基质类型**、**径流变化率**等十六个指标，并建立**投影寻踪模型**对各断面的水生态进行综合评分。通过**加速遗传算法**对模型进行求解，得到了相应的得分。之后将各指标的等级阈值代入到**投影寻踪评价模型**中，得到了相应的分数阈值，最后得到了黄河水生态综合状况所属的等级，发现存在较多的水生态整体情况不达标，仍有较大的提升空间。

本文的优点包括：1. 通过**经验正交函数分解法**得到了水质指标的时空模态，既能分别获取水污染在时间和空间上的规律，又能分析任意两断面指标变化的关联性。2. 采用**投影寻踪模型**对黄河宁夏河段水生态环境的整体情况进行评价，其结果为连续评分与离散等级相结合，更为科学合理。

关键词： 水质评价 模糊综合评价 正交经验分解 投影寻踪 加速遗传算法

目录

1	问题重述	1
1.1	问题背景.....	1
1.2	问题重述.....	1
2	模型的假设	1
3	符号说明	2
4	问题一的模型建立与求解	2
4.1	问题分析.....	2
4.2	数据预处理.....	3
4.3	模糊综合评价模型.....	3
4.4	模型求解.....	4
4.5	结果分析.....	4
4.5.1	综合污染情况分析	4
4.5.2	单一指标超标情况分析	6
5	问题二模型的建立与求解	6
5.1	问题分析.....	6
5.2	建立经验正交函数分解模型（EOF）	7
5.2.1	Manner-Kendall 非参数检验法.....	8
5.3	模型求解.....	8
5.4	模型检验及结果分析.....	10
5.4.1	North 检验	10
5.4.2	结果分析	11
6	问题三模型建立与求解	12
6.1	问题分析.....	12
6.2	数据补充.....	13
6.3	建立基于加速遗传算法的投影寻踪评价模型.....	14
6.4	加速遗传（AGA）算法求解	16
6.5	结果分析.....	19
7	模型总结与评价	19
7.1	模型优点.....	19
7.2	模型缺点.....	19
7.3	模型改进.....	20

参考文献	21
附录 A 代码	22
A.1 模糊综合评价 python 源程序	22

1 问题重述

1.1 问题背景

黄河作为中华民族的“母亲河”，在宁夏段由于长期接纳工业废水、生活污水和农田退水，使得宁夏全区地面水环境已受到不同程度的污染。作为宁夏最重要的地表水资源，如何评价分析黄河的水质情况对黄河的治理十分必要。



图 1 水污染治理

1.2 问题重述

问题一：结合附件中的内容，通过处理文中的数据得出结论分析黄河宁夏段某河段上的水环境污染现况。

问题二：结合数据，建立数学模型，得到黄河某段水污染的现状规律。

问题三：通过提出一个合理的评价体系，建立一个评价模型评价黄河水环境、水生态。

2 模型的假设

- 采样断面点位置设置合理，数据具有有效性；
- 一周内采样数据数值稳定不变；
- 选取指标能够有效反应水质。

3 符号说明

符号	意义
Y	隶属度
w_i	权重
λ	特征
$x_{(i,j)}$	样板集
R_{zy}	标准差

注：其他未列出的符号以第一次出现时的解释为准。

4 问题一的模型建立与求解

4.1 问题分析

根据《地表水环境质量标准 GB3838-2002》的相关要求，水质应该按照所有检测指标中最差的等级进行评级。此方案具有合理性，因为任何一个成分超标都会导致水源无法正常使用。

但是此评级方案无法区分同一类水质中的差别，且很可能仅因一个成分超标就导致评级结果远超目标值，无法客观反映当前水质与目标水质之间的距离^[1]。故本文拟通过模糊综合评价法得到水质综合等级的对原本依照国标确定的水质等级进行补充说明^[2]。

此外，本文拟通过对各断面水质的单一成分超标情况进行统计，分析黄河主要容纳的污染成分。

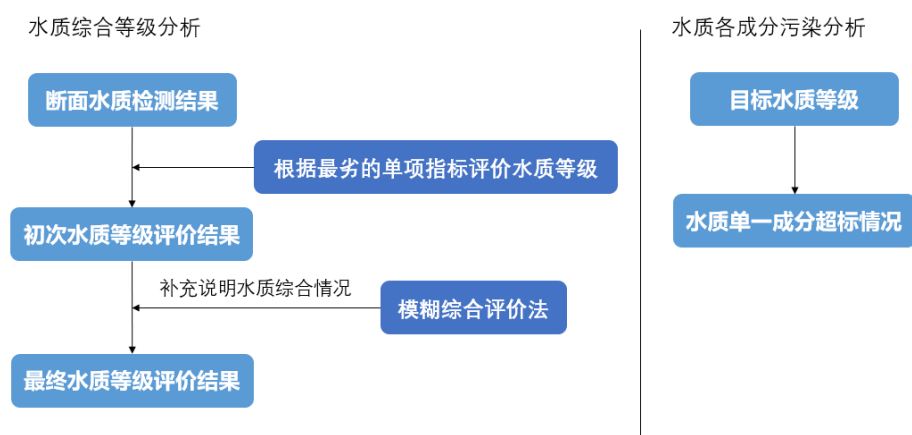


图2 问题一思路图

4.2 数据预处理

通过观察附件中的所给数据，可以发现 A 站点在七月份和八月份的所有数据都为 -1，属于异常数据。根据备注中的说明，这些异常数据的产生是因为监测点被堵塞，故本文选择剔除这些数据。

此外，通过查阅《地表水环境质量标准 GB3838-2002》，发现在水质检测的过程中，已经进行了重复实验保证结果的可靠性，故本文对附件中标红的超标数据不做额外处理。

4.3 模糊综合评价模型

模糊数学综合评价法的基础理论是隶属度理论，即用属于分别五类水质标准的程度大小去代替属于或不属于该种标准，用隶属度对受多种条件限制的事物作出总的评价。以下是具体步骤：

1) 建立综合评价因素集

因素集是有影响评价水质的因素组成的集合，通常用 $U = (U_1, U_2, U_3, \dots, U_m)$ ，其中 m 是评价因素的个数。本文选择 21 个评价因素，即 $U = (U_1, U_2, U_3, \dots, U_{21})$ ，其中分别表示为每种重金属的含量。

2) 建立综合评价的评价集

评价集是由各评价对象可能获得的结果所组成的集合，通常用 V 表示， $V = (V_1, V_2, V_3, \dots, V_n)$ ，其中 n 表示评价集的个数。水质评价按功能高低共有 I V 五类级别，所以，本文设定评价集为 $V = (V_1, V_2, V_3, V_4, V_5)$ 。

3) 计算隶属度

将某种事物属于某种标准的程度称为隶属度。水体水质好坏的程度界定较为模糊，可以用隶属函数来刻画模糊的水质等级边界。隶属度可以用隶属函数表示：

$$Y = \frac{X - X_0}{X_1 - X_0} \text{ 或 } \frac{X_1 - X}{X_1 - X_0}, \quad (1)$$

式中， Y 为隶属度； X 为实际监测值； X_0 、 X_1 为两相邻水质标准所对应的值。

4) 定各因子在综合评价中的权重

权重公式为 $W_i = \frac{C_i}{S_i}$ 式中， W_i

$$W_i = \frac{C_i}{S_i}. \quad (2)$$

式中， W_i 为评价参数权重； C_i 为 i 项检测指标浓度的实测值； S_i 为 i 项检测指标的标准浓度平均值。

5) 模糊矩阵的复合运算

将权重行列式与隶属度矩阵进行综合运算，得到模糊综合评价向量，根据最大隶属原则可知，综合评价向量中的最大隶属度所对应的评价等级即为模糊评价的水质等级。

4.4 模型求解

通过 python 编程进行求解，本文依据国标对每一个类型的水质进行模糊综合评价将附件中五类中的每一类水再细分为 A、B、C、D、E（品质由高到低）五个类别。如图3所示方式进行编码。

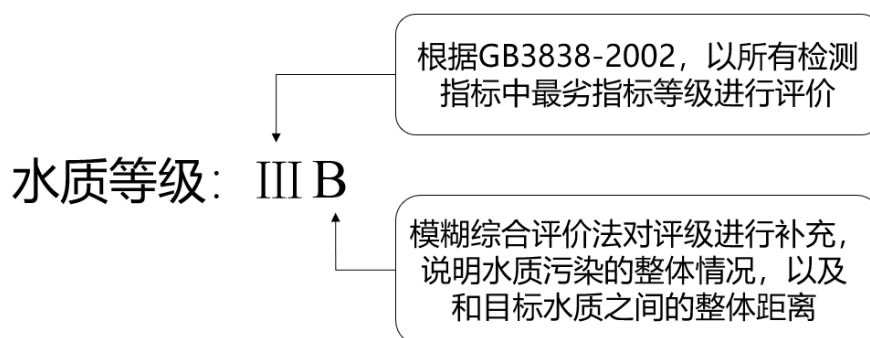


图3 水质等级说明

表1 各月份各界面水质细分表

	1月	2月	3月	4月	5月	6月	7月	8月
A	VE	VE	VE	IVD	IVC	IVC	\	\
B	VE	VD	VE	VC	VB	VC	IVA	IIIA
C	VE	IVD	IVD	IIIB	VE	IVC	IIIA	IIIA
D	IIIA	IIIA	VE	IIIB	IVC	IIIC	IIIA	IIIA
E	IIB	IIIC	IIIC	IIIA	VD	IVB	VE	IIIA

计算结果如表1所示，其为利用模糊综合评价对国家标准水质等级补充后形成的新的水质检验结果表，例如：E 站点中六月的水质情况为 IV 类水中的 B 类品质，其中 V 类表示此次检测中，水质各指标最差的等级为 IV 类；而 B 则表示其水质综合情况较好，大部分指标都能够满足要求。综合来看，虽然最差的指标属于 IV 类，但是该水质综合情况良好，具有较大的提升潜力。

4.5 结果分析

4.5.1 综合污染情况分析

图4为水质分布图，图5为贡献率最大的几种污染物的超标次数频数统计图。

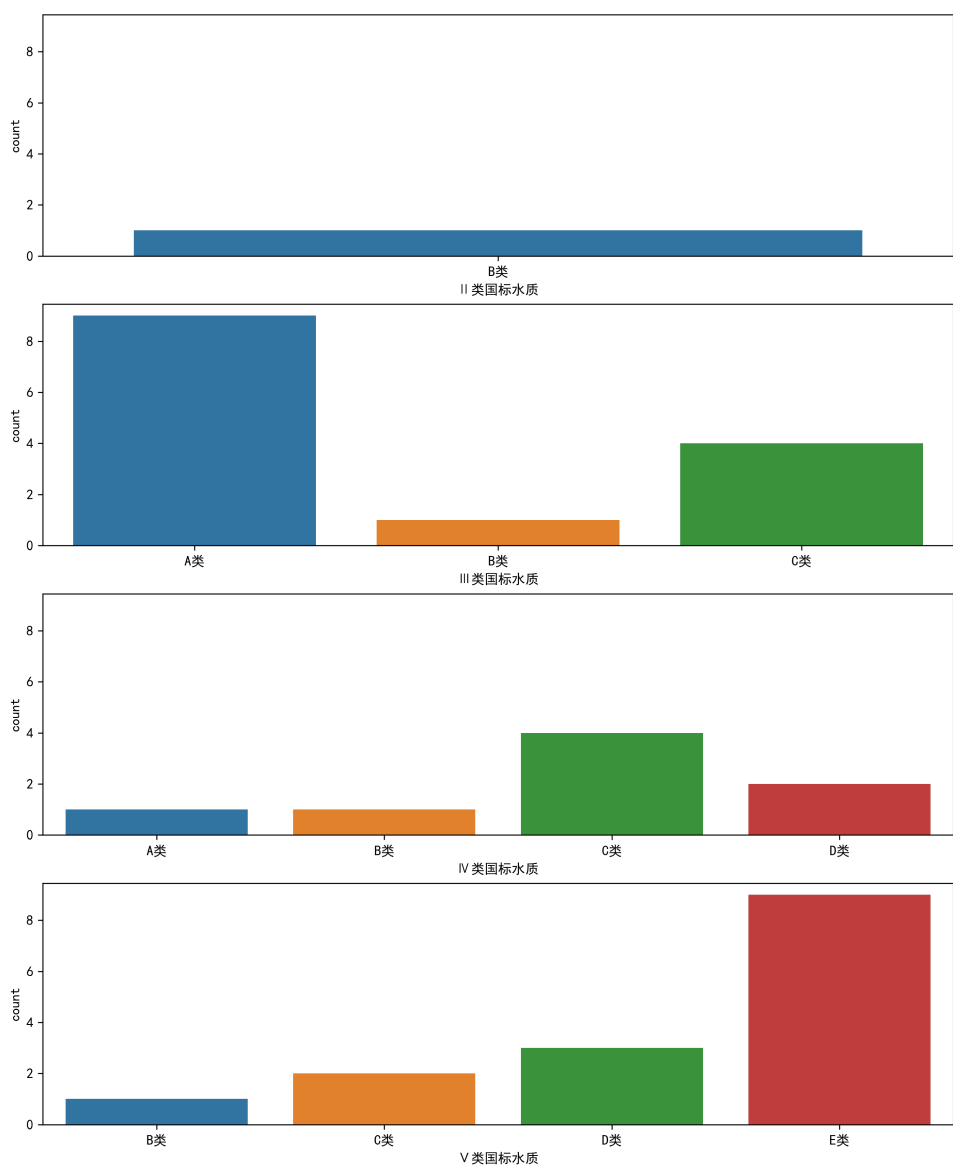


图4 水质分布示意图

从图中可以看出，宁夏水质情况处于 II-V 类水质，大部分为 IV、V 类水质。与宁夏河段 III 类目标水质进行对比仍有大部分水质处于不达标状态，但是观察每个等级中模糊综合评价的结果，发现尤其是 III 类和 IV 类水，水质的综合情况往往比最差指标所属的等级要好，故说明这些水往往是由于个别指标严重超标导致的水质不合格需要针对性的进行管控；而对于 V 类水，其模糊综合评价的结果也大多数属于 E 类，说明该水质的整体情况也不容乐观，需要进行全方面的管控。

4.5.2 单一指标超标情况分析

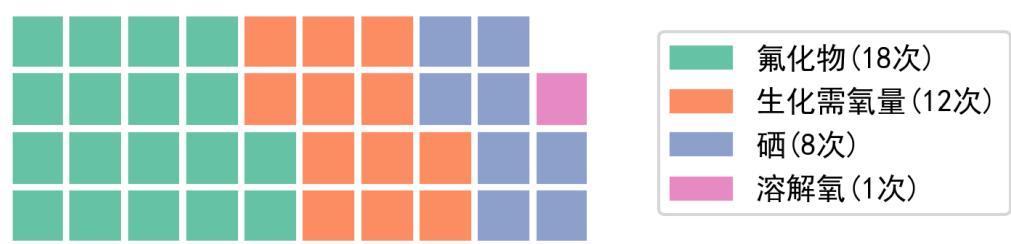


图 5 超标次数频数统计图

图中为单污染分析下的每一种主要污染物的污染次数，可以发现黄河宁夏河段的污染物种类比较有限，仅有生化需氧量、溶解氧、硒以及氟化物四类污染物超标。其中主要污染的指标为氟化物和生化需氧量，其对水质污染的贡献率高达 46.2%，而生化需氧量的贡献率也达到了 30.8%。说明了黄河宁夏段上游受到含有氟化物的工业废水污染较为严重，后续需要针对分析成因，重点排查相关企业。

5 问题二模型的建立与求解

5.1 问题分析

黄河具有着明显的枯水期和丰水期，两个时间段内的流量具有明显差异，对污染物的稀释能力也不相同。刘姜艳在《黄河流域水污染现状分析及控制》^[3]中指出黄河的枯水期为 10 月至次年 6 月，而丰水期则为 7 月至 9 月。本文拟将附件中八个月的断面水质检测数据按照上述方法分成两组分别进行研究。

为提取水环境污染的规律，本文从时间和空间两个角度进行考虑。故拟通过经验正交函数分解法对样本进行分析，分别得到时间和空间模态，并对时间模态进行趋势检验。

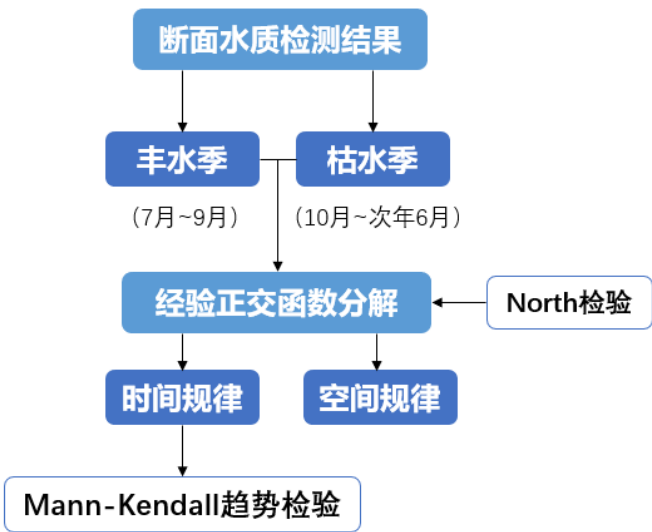


图 6 问题二思路图

5.2 建立经验正交函数分解模型 (EOF)

为获取黄河宁夏段各前面水质情况的时空分布, 应采取模型将水质情况分解至时间维度和空间维度。本文认为水质情况分别在时间维度和空间维度都具有相应的模态。因此本文选用正交经验分解模型, 拟分别提取出经典的时空模态。另外考虑到数据的分布不一定为正态分布, 故采用 Mann-Kendall 趋势检验。此方法不依赖数据的分布, 且能都避免异常数据干扰。以下是模型建立的具体步骤:

1) 数据预处理 选定要分析的水质数据, 进行数据的预处理, 通常处理成距平的形式得到一个数据矩阵 $X_{m \times n}$

2) 得到方阵 计算 X 与其转直矩阵 X 的交叉积, 得到方阵

$$C_{m \times m} = \frac{1}{n} X \times X^T. \quad (3)$$

如果 X 是已经处理成距平的话, 则称 C 为协方差阵; 如果 X 已经标准化 (即 C 中每一行数据的平均值为 0 标准差为 1) 则称为相关系数矩阵。

3) 计算特征根和特征向量 计算方阵 C 的特征根 $(\lambda_1, \lambda_2, \dots, \lambda_m)$ 和特征向量 $V_{m \times m}$, 二者满足

$$C_{m \times m} \times V_{m \times m} = V_{m \times m} \times E_{m \times m}, \quad (4)$$

其中 E 是 $m \times m$ 维对角阵, 即

$$E = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}$$

一般将特征根 λ 按从大到小排列。因为数据 X 是真实的观测值, 所以 λ 应该大于等于 0。每一个非 0 的特征根对应一列特征向量值, 也称 EOF。

4) 计算主成分 将 EOF 投影到原始资料矩阵上, 就得到所有空间特征向量对应的时间系数 (即主成分), 即

$$PC_{m \times n} = V_{m \times m}^T \times X_{m \times n}, \quad (5)$$

其中 PC 中的每一行数据就是对应每个特征向量的时间系数。

5) 计算贡献率 矩阵 X 的方差大小可以简单的用特征根的大小来表示。 λ 越高说明对应的模态越重要, 对总方差的贡献越大。第 k 个模态对总的方差解释率为

$$\frac{\lambda_k}{\sum_{i=1}^m \lambda_i} \times 100\%. \quad (6)$$

6) 显著性检验 实际资料分析中得到的空间模态是否是随机的，需要进行统计检验。North 等（1982）的研究提出，在 95% 置信度水平下的特征根的误差

$$\delta\lambda = \lambda\sqrt{\frac{2}{N^*}}, \quad (7)$$

其中 λ 是特征根， N^* 是数据的有效自由度。将 λ 按顺序依次检查，标上误差范围。如果前后两个 λ 之间误差范围有重叠，则没有通过显著性检验。

5.2.1 Manner-Kendall 非参数检验法

Manner-Kendall(M-K) 非参数检验法常用于分析降水、径流、气温等要素时间序列的变化趋势。本文采用此方法对 1 到 6 月枯水期的数据进行 M-k 趋势检验。

设有 n 个样本量的时间序列 x_1, x_2, \dots, x_n , 对于所有 $k, j \leq n$, 且 x_k 和 x_j 的分布是懂得，计算检验统计量 s , 公式如下：

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{Sgn}(x_j - x_k), \quad (8)$$

其中

$$\text{Sgn}(x_j - x_k) = \begin{cases} +1(x_j - x_k) > 0, \\ 0(x_j - x_k) = 0, \\ -1(x_j - x_k) < 0; \end{cases} \quad (9)$$

s 为正态分布，均值为 0，方差 $\text{var}(s) = n(n-1)(2n+5)/18$ 。当 $n > 10$ 时，标准正态统计变量通过下式计算：

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{var}(S)}} > 0, \\ \frac{S+1}{\sqrt{\text{var}(S)}} < 0. \end{cases} \quad (10)$$

5.3 模型求解

利用 matlab 进行求解，可以得到各个污染物的变化趋势，如下表3，图10为变化明显的污染物变化趋势大小关系图。图7是部分污染物的时间系数变化。

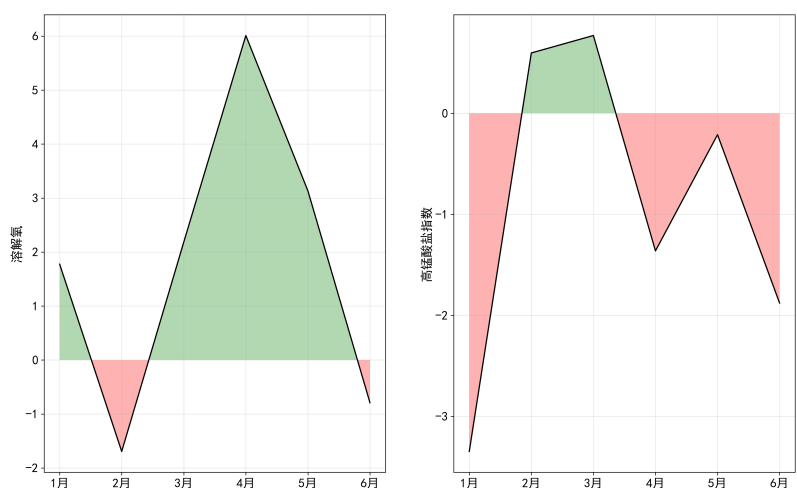


图7 时间系数趋势图

图7为部分污染物的时间模态图，图中可以分别看出溶解氧和高锰酸盐污染物污染物时间上的变化趋势。其中模态的正负表示各断面变化过程中的正负相关性，以图7（a）为例，当1月份增加一个单位时，根据统计学意义，模态同为正的3月、4月、5月都会按照一定的比例增加，而2月和6月会按照一定比例减少。

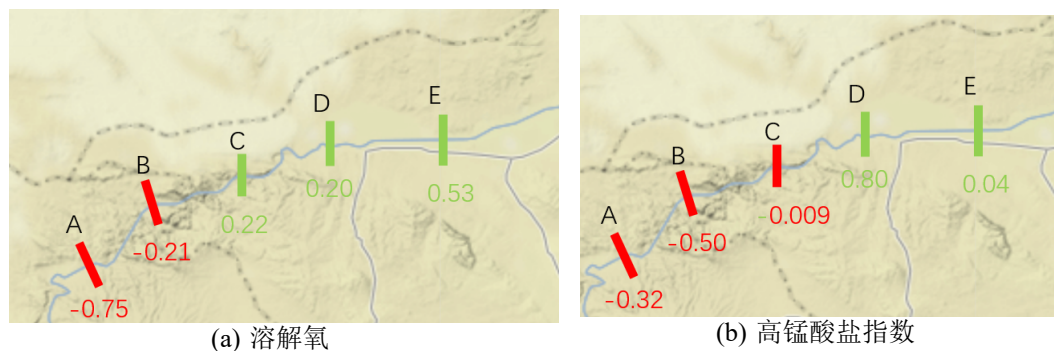


图8 空间趋势图

图8为部分污染物的空间系数图，图中可以分别看出溶解氧和高锰酸盐污染物在各断面处的情况。模态正负的含义与上文中时间模态正负的含义相同。从图中不难看出，黄河宁夏河段上游污染情况与下游污染情况具有明显的差异性，且不难看出下游的污染情况远比上游要轻。

表 2 污染物变化趋势

污染物	变化趋势
生化需氧量	增加
氨氮	增加
铜	减少
砷	增加
六价铬	增加
硫化物	减少
其污染物	无明显趋势

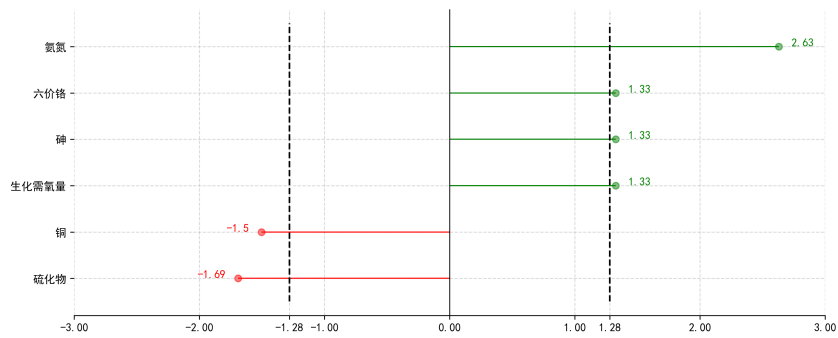


图 9 污染物趋势分布图

通过图片和表格可以看出，在未来一段时间里氨氮、六价铬、砷、生化需氧量这些污染物浓度有增加趋势，其中氨氮浓度增大趋势最明显，而铜含量和硫化物则有明显的减少趋势。因此具有增加趋势的污染物需要特别注意。

5.4 模型检验及结果分析

5.4.1 North 检验

为了确定各个模块之间是否相互独立，North 检验是最简单也是必须要做的检验。North 检验，是计算特征值误差范围来进行显著性检验。特征值 λ 的误差范围

$$e_j = \lambda_j \left(\frac{2}{n} \right)^{\frac{1}{2}} \quad (11)$$

其中 n 为样本量，当相邻特征值 $\lambda_j + 1$ 满足

$$\lambda_{j+1} - \lambda_j \geq e_j$$

时，认为这两个特征值对应的经验正交函数是有价值的信号。下表是 North 检验结果：

表 3 污染物变化趋势

污染物种类	North 检验结果
高锰酸盐指数	12.0249>11.1121
生化需氧量	40.5699>30.8463
氨氮	14.8622>10.1053
石油类	0.0004>0.000252982
挥发酚	1.6e-08>1.01193e-08
...	...
铜	0.0001944>0.000122949
硒	1.5552e-05>9.83595e-06
砷	1.23709e-05>9.75207e-06
阴离子表面活性剂	0.008>0.00505964
硫化物	0.000341383>0.000237382

通过 North 检验，可以发现每一个污染物指标都能通过检验，因此可以说明每一个模块的数据都有价值，此模型可以较为成功的反应实际问题。

5.4.2 结果分析

通过正交经验分解（EOF），分别得到了污染物变化的时间模态和空间模态。从空间维度上来说，污染物的分布情况与上下游之间具有明显的相关性，具体反映在污染情况上则为上游污染更为严重，而下游污染较轻。经过查阅《中华人民共和国水污染防治法》，发现对于各省出境的水质具有严格要求，不达标则会受到严重的处罚，因此政府可能着重于管理出境河段的污染。因此很有必要完善相关的法律法规，使得全河段的污染情况都得到严格的管控。

对时间模态经过 M-K 趋势检验后，本文得到了几种污染物未来的变换趋势，基于此类趋势可以检验当地污染排放的种类。通过查阅相关资料，发现生化、化工工厂的废水均含有氨氮、六价铬、砷等污染物，因此建议宁夏有关部门重点排查相关企业，做好排放前的污水治理工作，以此减少黄河污染程度。

6 问题三模型建立与求解

6.1 问题分析

为建立评价黄河水生态的评价模型，首先需要构建合理的指标体系。在众多学者对生态评价的研究中，都将问题转化为生态承载能力的评价^{[4]~[6]}，故指标体系中不仅需要考虑到黄河水质，还需要考虑到人类开发程度、黄河河岸及其周边环境等。

目前，国内大量学者对黄河流域的相关指标进行了探索。田野在《黄河内蒙古段河流健康评价》中提出了对河床、河岸等物理结构的相关指标，张丽洁在《黄河流域水资源评价研究》中提出了水质、水量等维度的指标。本文在上文中选取了具有代表性的 6 个一级指标以及 16 个二级指标。

此外，本文拟通过实数编码加速的投影寻踪法^[7]对各指标进行赋权，得到黄河水生态的最终评价结果。并且，本文将权重与上述指标的标准阈值评价，对比分析后得到各检测结果的生态等级。

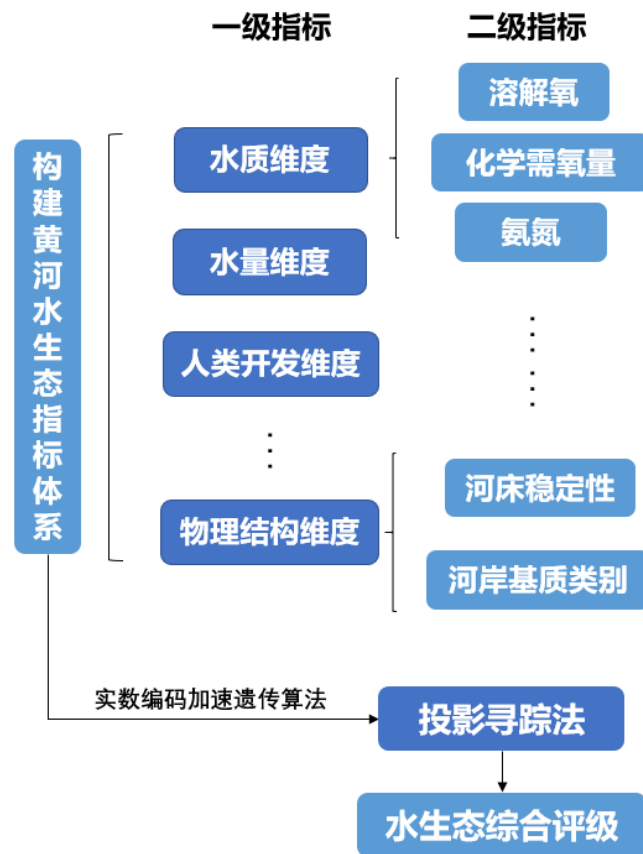


图 10 问题三思路图

6.2 数据补充

水生态的评价比较复杂，需要通过查阅相关文献以获取更多例如径流变化率等水文数据。收集到的数据如表4所示。分级指标如表5所示。

表 4 数据及来源

指标名称	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	数据来源
人均水资源占有量 (立方米)					960				《2018 年宁夏水资源公报》
径流变化率/%	0.56	0.47	0.51	0.68	1.09	1.15	1.8	1.76	《黄河上游宁蒙河段径流特征分析及预测》
含沙量变化率/%					0.2145				《中国河流泥沙公报》
生态流量保证程度/%		0.3214				0.4125			
万元生产总值用水量 (立方米)					179				《2018 年宁夏水资源公报》
万元工业增加值用水量 (立方米)					91				
农业灌溉水利用系数					0.42				
水资源开发利用率/%					48.4				《宁夏地下水资源分布特征及开发利用现状研究》
河床稳定性					砂卵河床				《黄河上游宁蒙河段径流特征分析及预测》
河岸基质类别					非黏土河岸				《中国河流泥沙公报》
河流平均比降/%					11.6				《黄河上游宁蒙河段径流特征分析及预测》

表 5 分级标准

指标类别	指标名称	I 类	II 类	III 类	IV 类	V 类	数据来源
水量维度	人均水资源占有量 (立方米)	3500	2200	1700	1000	500	《黄河流域水资源评价研究》
水文维度	径流变化率/%	100	95	90	80	65	《黄河内蒙古段河流健康评价》
	含沙量变化率/%	0.05	0.15	0.2	0.25	0.4	《地表水环境质量标准》
	生态流量保证程度/%	0.5	0.4	0.3	0.1	0	
人类开发维度	万元生产总值用水量 (立方米)	24	60	140	220	500	《黄河流域水资源评价研究》
	万元工业增加值用水量 (立方米)	0	15	50	100	300	
	农业灌溉水利用系数	0.6	0.5	0.4	0.3	0	
	水资源开发利用效率/%	10	20	40	60	100	
物理结构维度	河床稳定性	1	2	3	4	5	《黄河流域水资源评价研究》
	河岸基质类别	基岩	基岩	岩土河岸	黏土河岸	非黏土河岸	
水流维度	河流平均比降/%	10	20	30	40	50	《黄河流域水资源评价研究》

6.3 建立基于加速遗传算法的投影寻踪评价模型

投影寻踪 (PP) 属于直接由样本数据驱动的探索性数据分析方法, 可以将高维数据通过组合投影到低维空间, 通过寻找最优投影函数来使得投影能反映高维数据的结构或者特征。

1) 评价指标归一化

设样本集为 $x_{(i,j)}^*, i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$ 其中 $x_{(i,j)}$ 为第 i 个样本第 j 个指标值。为消除各指标的量纲和统一化各指标的变化范围, 可采用下式进行极值进行归一化处理:

$$x_{(x,j)} = \frac{x_{(i,j)}^* - Ex(j)}{Sx(j)}, \quad (12)$$

式中, $Ex(j)$ 、 $Sx(j)$ 分别为第 j 个指标值的均值和标准差。

2) 构造投影指标函数

将 p 维数据 $\{x_{(i,j)} | j = 1, 2, \dots, p\}$ 以 $a = (a_1, a_2, \dots, a_p)$ 方向投影为一维值 Z_i

$$Z_i = \sum_{j=1}^p a_j x_{(i,j)}, \quad (13)$$

式中 a 为单位长度向量。

在综合投影值时, 要求投影值 Z_i 散布特征为: 1、投影值 Z_i 尽可能多的提取到

$\{x_{(i,j)}\}$ 中的变异信息, 即 Z_i 的标准差 S_z 达到尽可能大, 同时要求 Z_i 与已知标准等级 y_i 的相关系数的绝对值 $|R_{zy}|$ 达到尽可能大。为此, 投影指标函数可构造为

$$Q(a) = S_z |R_{zy}|, \quad (14)$$

式中, S_z 为投影值 R_{zy} 的标准差。式中 S_z 、 R_{zy} 分别为

$$Q(a) = S_z |R_{zy}|, \quad (15)$$

式中, S_z 和 R_{zy} 分别为

$$\begin{aligned} S_z &= \left[\frac{\sum_{i=1}^n (Z_i - EZ)^2}{n-1} \right]^{0.5}, \\ R_z &= \frac{\sum_{i=1}^n (Z_i - EZ)(y_i - Ey)}{[\sum_{i=1}^n (Z_i - EZ)^2 \sum_{i=1}^n (y_i - Ey)^2]^{0.5}}. \end{aligned} \quad (16)$$

3) 优化投影指标函数 当各个指标值的样本集给定时间, 投影函数值 $Q(a)$ 只随着投影方向 a 的变化而变化。不同的投影方向反映不同的数据特征, 最佳投影方向就是最大可能暴露高纬度数据某类特征的投影方向。可通过求解投影指标函数最大化问题来估计最佳投影方向:

$$\begin{aligned} \max Q(a) &= S_z |R_{zy}|, \\ \text{s.t. } \sum_{j=1}^p a^2(j) &= 1. \end{aligned} \quad (17)$$

4) 建立投影寻踪等级评价模型

将步骤三求得的最佳投影方向的估计 a^* 代入投影值公式, 得第 i 个样本投影值的计算值 Z_i^* , 根据 $Z_i^* - y_i$ 的散点图建立相应的等级评价模型。 Z_i^* 与 $y(i)$ 之间一般呈现单调非降关系, 当 Z_i^* 值超过某门限值就会判定最高等级, 当指标值低于另门限值就会判定为最低等级, 当 Z_i^* 介于这两点中间时则为中等等级, 可以用逻辑斯蒂曲线来描述:

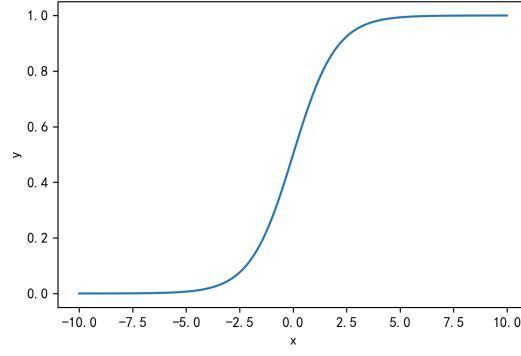


图 11 逻辑斯蒂曲线示意图

$$y_i^* = \frac{N}{1 + e^{c_1 - c_2 Z_i^*}}, \quad (18)$$

式中 y_i^* 为第 i 个样本等级的计算值；模型参数 c_1 、 c_2 分别为积分常数和增长率，可通过求解如下优化问题来确定：

$$\min F(c_1, c_2) = \sum_{i=1}^n (y_i^* - y_i)^2. \quad (19)$$

6.4 加速遗传（AGA）算法求解

加速遗传（AGA）算法对于普通遗传算法的改进：

- 1) 控制参数设置：每次加速循环中的 AGA 只循环两次 SGA 的进化迭代。
- 2) p 个变量、加数循环 q 次，优秀个体包围最优点的概率为 $(1 - 0.5^{2s})^{pq}$ 。

以下是 AGA 求解伪代码：

Algorithm 1: Function AGAPPGE(x)

```

1 StandardScaler(x) /* 归一化：为消除量纲与指标值的变换范围，对数据进行归一化 */
2 Initialize /* 初始话：初始化种群 */
3 iteration = 1; Repeat for  $i \in [1, N]$  do
4   ComputeFeasibility( $x_i$ ) /* 计算适应度 */
5   Sort( $x_i$ ) /* 按照适应度进行排序 */
6   Evaluate( $x_i$ ) /* 计算基于序的评价函数 */
7   Select( $x_i$ ) /* 选择操作 */
8   Mutate( $x_i$ ) /* 变异操作 */
9   Merge( $x_i$ ) /* 合并交叉部分和剩余部分 */
10 end
11 iteration  $\leftarrow$  iteration+1; 进入下一次迭代
   until 达到迭代次数。

```

通过 matlab 求解，得到一到八月每个节点的水生态评分如表8，表中分数越低则说明该月该节点的水生态越恶劣。

表 6 各月份各断面水质细分表

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月
A	2.231876	3.051511	2.474963	2.332324	2.801374	2.508442	\	\
B	2.137685	2.330664	2.766444	2.019016	2.478643	2.005526	2.797785	2.796999
C	2.329781	2.933338	2.799032	2.54852	2.493417	2.345483	2.818519	2.938506
D	3.574317	3.452619	2.802116	2.943436	2.763737	2.921222	2.839983	2.880777
E	3.410205	3.307502	3.171557	2.795494	2.499318	2.351724	2.183836	3.085028

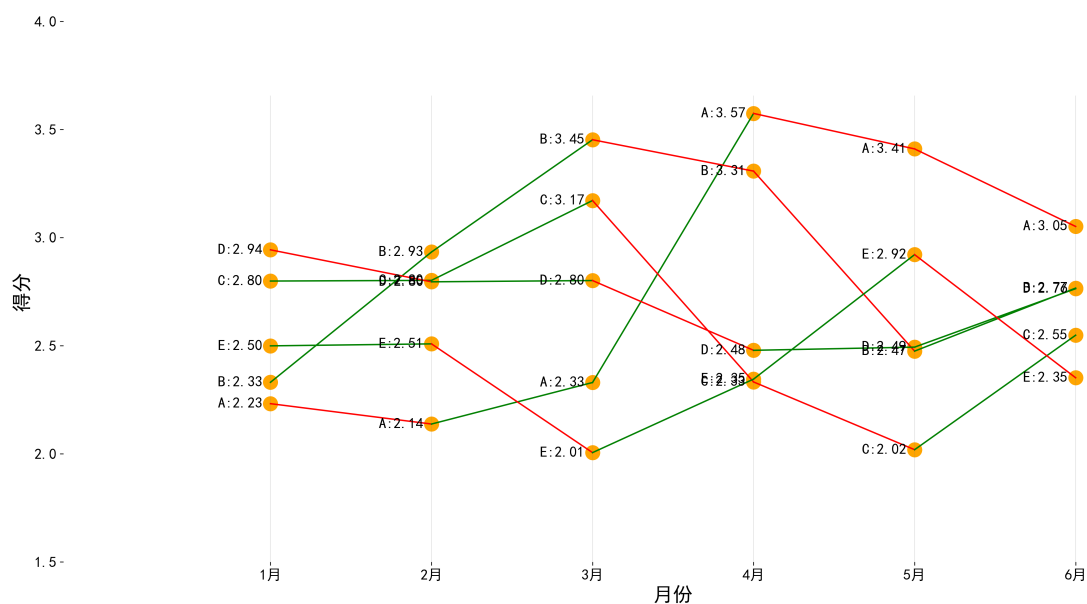


图 12 各个断面 1 到 6 月得分趋势图

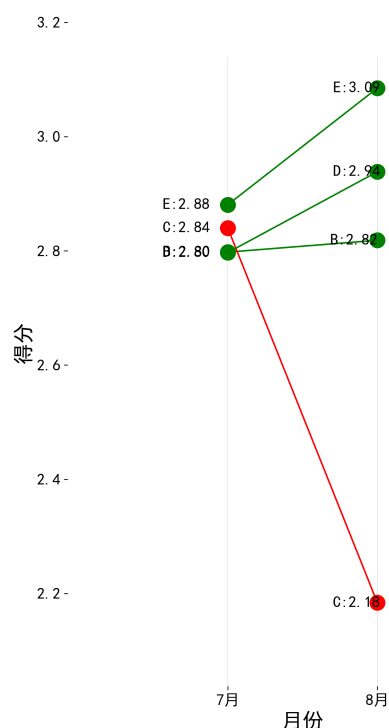


图 13 各个断面 7 到 8 月得分趋势图

图12、图13为枯水期和丰水期各个断面随着时间变化的趋势折线图。每个断面的评分为利用投影寻踪法求出的水质评分。通过评分可以观察每个截面水质在时间上的变化。

在计算出各断面的水生态得分后，为更客观的反映相应水质的优劣，本文拟引入各指标的分级标准（具体分级阈值见表5分级标准），并将每一级的指标阈值代入到上述的投影寻踪评价模型中，得到水生态综合等级划分阈值，见表7：

表 7 分类等级评分

分类等级	I 类	II 类	III 类	IV 类	V 类
评分	4.045987	3.327785	2.436266	1.232433	0.299109

将表 6 各月份各断面水质细分表的数据分别代入至表 7 分类等级评分中，最终得到各断面的水生态等级。需要说明，其中 I 类表示水生态健康，V 类表示水生态严重不健康。

表 8 各月份各断面水质细分表

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月
A	IV	III	III	IV	III	III	\	\
B	IV	IV	III	IV	III	IV	III	III
C	IV	III	III	III	III	IV	III	III
D	II	II	III	III	III	III	III	III
E	II	III	III	III	III	IV	IV	III

6.5 结果分析

从整体分析，存在较多的生态整体情况不达标。从时间上看枯水期的生态情况不容乐观，一月和二月的评分较低，究其原因可能为黄河冰冻期，河流流速缓慢并且冬季取暖导致污水排放增加，而三、四月份河流解冻流速和流量都增加所以导致污染物浓度下降，评分增加。四五月份为枯水期界面流量的下降导致了浓度的增加。从数据上来看，污染物浓度变化影响的水质指标的变化对水生态评分的影响最大。

综上，基于相关资料，本文判断此评价模型能较为客观的反映出黄河宁夏段的水生态情况。同时本文认为维护水生态最快速的方式就是改善水质情况，减少水污染。从产业的角度来说，如今整个黄河流域产业体系上偏重，煤炭、有色占了比较大的比例，资源性重度污染在黄河流域上游出现较多，并且目前仍有大量企业在黄河沿岸随意排放污水，因此如何监管企业的污水处理情况十分重要，有关部门需要加大监管力度。

7 模型总结与评价

7.1 模型优点

1. 模糊综合评价从整体水质的角度进行了描述，与国家标准中规定的以最劣指标划分等级的方法相结合，使得评价结果更为全面。
2. 运用了投影寻踪法对水生态情况进行分析，能够获得客观、科学的评价，且评分为连续的数值，比等级评价的结果更为细致。
3. 通过经验正交函数分解法得到了水质指标的时空模态，并从时间和空间两个维度进行分析。除了能够分析单一指标的时空趋势，还能够得到任意两断面之间变化的关联性。

7.2 模型缺点

1. 由于附件未标明数据来源的年份和具体经纬度，故在第三问补充数据时仅能保证数据皆为同一个月份，导致评价结果存在一定偏差。

7.3 模型改进

水污染的动力学模拟为当前的研究热点，国内外的诸多学者也提出了诸如 WASP、StreeterPhelps、QUAL 等水质模型。由于附件中缺少断面具体位置的描述，故未能实施。若将来能获取更为全面的数据，可基于污染物扩散机理，对污染源回溯、污染物降解过程进一步模拟分析。

参考文献

- [1] 景朝霞, 夏军, 张翔, 等. 汉江中下游干流水质状况时空分布特征及变化规律 [J]. 环境科学研究, 2019, 32(01): 104-115.
- [2] 彭亚辉, 周科平, 蒋俊伟. 湘江流域长株潭段水污染负荷时空分布规律及成因 [J]. 中国农业大学学报, 2018, 23(09): 108-116.
- [3] 贺磊. 基于熵权法 - 正态云模型的辽宁省水生态承载力评价 [J]. 水资源开发与管
理, 2020(07): 16-22.
- [4] 隋欣, 齐晔. 基于生态系统健康的生态承载力调控模式研究 [J]. 应用基础与工程科学
学报, 2006(04): 479-487.
- [5] 阿提卡木·阿布来提. 不同水质评价方法在喀什地区地下水水质评价中的适用性分
析 [J]. 地下水, 2020, 42(04): 49-50+80.
- [6] 陈明霞, 熊贵耀, 张佳鹏, 等. 湘江流域水质综合评价及其时空演变分析 [J]. 环境工
程, 2019, 37(10): 83-90+104.
- [7] 李琪, 赵志怀. 基于投影寻踪与内梅罗指数组合模型的地下水水质评价 [J]. 水电能源
科学, 2019, 37(11): 70-73.

附录 A 代码

A.1 模糊综合评价 python 源程序

```
def compute_gbstd(df, gbdata):
    roma=[' 类',' 类',' 类',' 类',' 类']
    ms = np.zeros((1, len(df.index)))

    for idx, col in enumerate(df.index):
        if col not in ['溶解氧']:
            if df[col] < np.float64(gbdata[gbdata['feature'] == col][roma[0]]):
                ms[0, idx] = 1
            elif df[col] > np.float64(gbdata[gbdata['feature'] ==
                col][roma[-1]]):
                ms[0, idx] = 5
            else:
                for i in range(5):
                    if df[col] < np.float64(gbdata[gbdata['feature'] ==
                        col][roma[i]]):
                        ms[0, idx] = i+1
                        break
        else:
            if df[col] > np.float64(gbdata[gbdata['feature'] == col][roma[0]]):
                ms[0, idx] = 1
            elif df[col] < np.float64(gbdata[gbdata['feature'] ==
                col][roma[-1]]):
                ms[0, idx] = 5
            else:
                for i in range(5):
                    if df[col] > np.float64(gbdata[gbdata['feature'] ==
                        col][roma[i]]):
                        ms[0, idx] = i+1
                        break
    return max(max(ms))
```