

湖北师范大学数学建模协会制

黄石市第三届“青磁湖杯”

数学建模竞赛（夏季赛）

承 诺 书

我们仔细阅读了“2022 年湖北师范大学第十七届文化科技节——第三届“青磁湖杯”黄石市大学生数学建模大赛（夏季赛）论文格式规范”。


我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们接受相应处理结果。

我们允许湖北师范大学数学建模协会批阅论文并进行发布，以供校友之间学习交流。

参赛队员（签名）：

队员 1: 

队员 2:

队员 3:

（没有可不填）参赛队教练员（签名）：

选题	2022 年湖北师范大学第十七届文化科技节——第三届“青磁湖杯”黄石市大学生数学建模大赛	参赛编号	组别
C		2022QCHB0020	L

黄石市 第三届 “青磁湖杯”

数学建模挑战赛论文

题 目 基于灰色预测与改进 TOPSIS 模型的水质预测与评价

关 键 词 灰色预测, TOPSIS, 熵权法

摘 要:

水资源的监测和保护对我国生态发展是一项重要工程。我国是一个水资源匮乏的国家, 保护和发展水资源生态很有必要, 但目前数据监测成本较大, 水质评价体系较为单一, 如何合理基于历史数据进行水质预测和综合评价是较为重要的问题。

对于问题一, 这是一个时间序列预测问题。由于只提供了 14 个月的数据并且存在离群点, 小体量下并不适合用过于复杂的流行时间序列建模方法。这里经过对比, 选择使用灰色预测模型进行合理建模, 所得到的 20 个序列 MSE 平均为 0.011, 相比于线性回归等拟合方法更加准确。最终预测的未来三月结果列在正文中。

对于问题二, 这是一个综合评价问题。将三个月的数据取平均值可以获得综合评价的数据, 参考附表中提供的水质评级对不同指标进行了预处理。进行综合评价的方法有很多, 这里选择使用 TOPSIS 模型 (优劣解距离法) 进行建模。考虑到不同指标对最终的评分影响不同, 基于熵权法对原有传统 TOPSIS 模型进行了赋权修正, 得到更为准确的评价体系。

对于问题三, 本质上是对问题二的延伸与补充。不同指标对最后评分的影响差异在问题二中已经通过赋权的方法解决, 所以只需要基于问题二的结论对不同指标进行阐释说明即可提出对应的建议。

综上, 模型的建模流程并不复杂, 但能够在一定程度上有效地基于历史数据进行水质的监测和评价建模。

参赛队号: 2022QCHB0020

所选题目: C 题

参赛密码: _____

(由组委会填写)

一. 问题重述

我国是一个水资源匮乏的国家,各地水资源均受到不同程度的污染,其阻碍了农业、工业等很多产业的发展。其次,在工业的不断发展过程中,大量硫化物排放至大气中后往往会导致酸雨等灾害。这种自然灾害不仅会影响农作物的生长,而且会破坏水生植物的生长环境,使原本的生态平衡遭到破坏,造成更大的环境效益损失。因此,保护水环境、提高水质质量迫在眉睫。

而目前对水质资源分析的主要目的一方面是为了预测水体各项指标随时间的变化以采取对应的应对措施,另一方面是对水体质量进行合理的综合评估以合理地将水体治理的力量更多地倾斜到更需要治理的地区。

为了有效治理水资源污染,专家对部分地区的历史水质指标进行了统计并将数据整理到附件中,希望通过分析数据得到更科学的建议以更好地治理水环境。现需要解决的问题如下:

(1) 附件一中给出了五个地区一段时间内的 pH^* 、DO、CODMn、 $\text{NH}_3\text{-N}$ 指标值,需要基于历史数据以预测五个地区未来三个月 pH^* 、DO、CODMn、 $\text{NH}_3\text{-N}$ 的含量。由于数据只有 14 个月,所以只需要预测三项即可。

(2) 附件二中给出了 2019 年 3 月至 2019 年 5 月中 17 个地区的一些水质指标和生态指标,需要对各地区水质质量状况进行评定,进而得到水质质量排名。本质上是需要构建一个综合评价模型

(3) 需要基于问题二寻找对这 17 个地区水质影响最大的因素,并向当地环保局提出建议。

二. 问题分析

本问题是一个小体量数据分析和综合评价的问题,对于这个问题我们的分析如下:

对于问题一,这是一个时间序列的预测问题。传统时间序列预测方法有很多,包括回归拟合、灰色预测、ARIMA 模型、以神经网络为代表的机器学习方法等。但首先我们需要注意两个问题:第一是数据体量其实并不大,只有 2003 年 6 月到 2004 年 7 月一共 14 个月的数据,如果使用 ARIMA 模型和神经网络等大规模序列模型容易学习不到良好的参数;第二是数据中存在离群点,也就是某地区的某个指标在某一日偏离平均水平较大,如果直接拟合容易造成较大偏差,故我们应该使用更鲁棒的策略。这里可以对比 GM(1,1)模型和岭回归、LASSO 回归等方法进行合理预测。

对于问题二,这是一个综合评价问题。给定的数据有三个月所以在进行综合评价时可以取平均值作为综合评价的标准。而多指标综合评价的核心主要是两点:权重和得分排名。所能够采取的策略也是多种多样的,包括层次分析法、TOPSIS 分析法、CRITIC 方法、主成分分析法、因子分析法等。由于给定的附表中还有各项指标的范围,所以还可以基于模糊隶属度进行模糊评价。本质上这些评价方法的效果其实很类似都是为了获得一个排名,这里我们可以使用优劣解距离法,也就是 TOPSIS 方法。注意到这里 PH 虽然是区间指标,但由于化学常识 PH 越接近 7 越好。另外 CODMn、 $\text{NH}_3\text{-N}$ 和垃圾密度是越小越好,在进行 TOPSIS 之前需要进行预处理。并且不同指标在评价过程中重要性不同,我们用熵权法进行权重确定。

对于问题三,影响因素的衡量本质上就是评价体系中的权重,基于问题二的结果就可以确定。

三. 模型假设

为更合理地建模以得到更确切的结论，我们对模型进行如下约定：

1. 假设所有的数据均足够真实，不考虑测量过程中由于意外造成的数据测量错误。这一假设保证数据真实性，否则模型的结果不具备说服力。
2. 假设不考虑不同地区之间水资源流通性造成的相互干扰，这一假设保证研究的对象是相互保持独立的，减少了观测对象之间的干扰因素。
3. 假设不考虑除问题所给定观测变量以外的指标对水质的影响，这一假设保证所使用的数据、体系等都是完备的，不需要另外考虑其他影响因素。

四. 符号约定

我们将文中出现的数学符号列在下表中：

符号定义	符号说明
Z	归一化的决策矩阵
z_{ij}	决策矩阵中第 i 个决策项第 j 个指标
w_j	熵权法求解的第 j 个指标的权重系数
Z^+, Z^-	正理想解和负理想解
D^+, D^-	决策项到正理想解和负理想解的距离
p_{ij}	对于指标 j ，第 i 类所占比例
e_j	指标 j 的熵值
n	指标 j 有多少类
$p(x)$	x 的概率
$x^{(0)}(t)$	原序列 x 在第 t 项的值
$x^{(1)}(t)$	1 阶累加序列 x 在第 t 项的值
$z(t)$	均值序列
a, b	方程的拟合参数

五. 模型建立与求解

5.1 问题一的建模与求解

为探索数据中蕴含的规律并选择最合适的建模策略，我们将不同地区的不同指标随时间变化绘制在图 1 中，如图 1 所示：

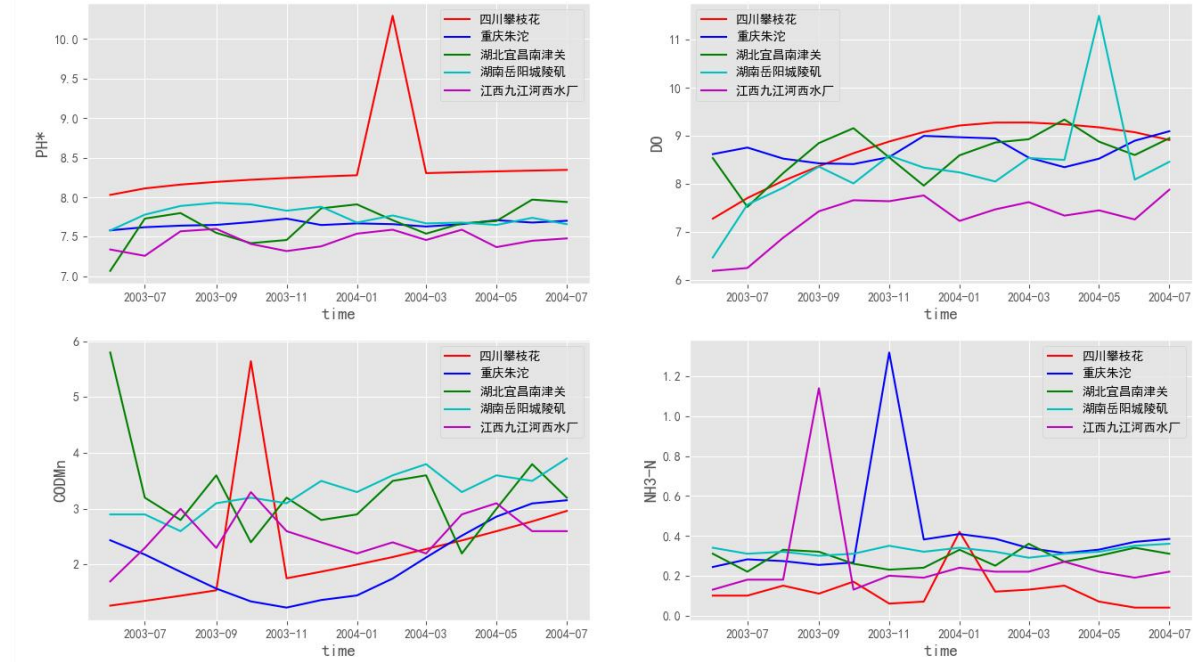


图 1. 不同地区的不同指标随时间变化

从图 1 中我们可以看到，大部分数据波动较大，并且存在个别水平远高于平均水平的异常数据，不适合直接进行拟合与回归。由于数据体量比较小只有 14 条，常用的 ARIMA 模型、神经网络等方法学习不到较为准确的参数，我们考虑使用灰色预测的方法进行时间序列预测。

灰色预测方法基于灰色系统原理，是一种适用于中短期时间序列体量下的系统建模方法[4]。对于给定的数据序列，我们的建模过程如下：

Step I: 生成 1 阶累加序列：

$$x^{(1)}(t) = \sum_{i=0}^t x^{(0)}(t) \quad (1)$$

Step II: 规定均值序列：

$$z(t) = \frac{x^{(1)}(t) + x^{(1)}(t+1)}{2} \quad (2)$$

基于均值序列和原序列我们可以建立灰色方程：

$$\frac{dx^{(1)}(t)}{dt} + az(t) = b \quad (3)$$

Step III: 利用最小二乘法求解对应的白化微分方程可以解得最优系数 a 和 b，并且得到方程通解形式为：

$$x^{(1)}(t) = \left(x^{(1)}(0) - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \quad (4)$$

Step IV: 利用逐差法还原为待预测序列即可得到结果。

为验证灰色模型的有效性，我们以湖南岳阳城陵矶和江西九江河西水厂的水中溶解氧(DO)为例构建对应的模型进行回归。由于数据中存在极个别的高偏数据，所以这类问题并不适合直接用 OLS 等方法建模而应该使用更鲁棒的回归方法例如岭回归。

将回归结果展示在图 2 中，如图 2 所示：

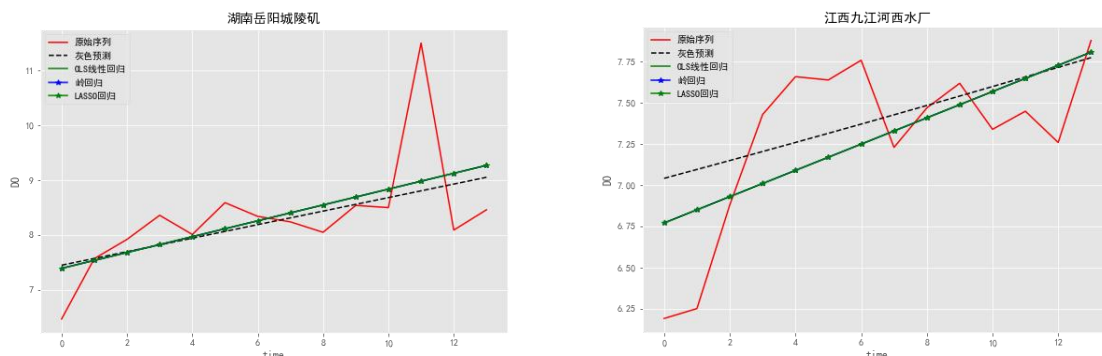


图 2. 不同模型的效果对比案例

总体而言灰色预测和几种线性方法的趋势变化并不十分明显，并且岭回归、LASSO 与普通的 OLS 之间差异不大。OLS 的 R^2 分数分别为 0.324 和 0.268，置信度均满足 0.05 置信度范围，说明确实效果不会很好。而灰色序列预测的分数略高一些，分别为 0.331 和 0.289，证明使用灰色预测的确是会更奏效一些的。

MSE 是衡量预测回归问题的常用指标，灰色预测结果的 MSE 如表 1 所示：

表 1. 灰色预测结果的 MSE

地区	pH*	DO	CODMn	NH3-N
四川攀枝花	0.026	0.0182	0.0279	0.0037
重庆朱沱	0.00013	0.0055	0.0144	0.0081
湖北宜昌南津关	0.0058	0.0191	0.0167	0.0061
湖南岳阳城陵矶	0.0014	0.0097	0.0131	0.0011
江西九江河西水厂	0.0015	0.0252	0.0073	0.0129

从表 1 的结果中我们可以看出，MSE 表示误差，而灰色预测的 MSE 普遍不高，说明预测的最终结果是较好的。

而灰色预测得到的接下来三个月的结果如表 2 所示：

表 2. 灰色预测结果

指标	月份	四川攀枝花	重庆朱沱	湖北宜昌南津关	湖南岳阳城陵矶	江西九江河西水厂
PH*	8	8.6440	7.6975	7.8671	7.6130	7.4995
	9	8.6731	7.7012	7.8867	7.5931	7.5041
	10	8.7024	7.7048	7.9064	7.5732	7.5087
DO	8	9.6498	8.8725	9.2159	8.8725	7.8354
	9	9.7570	8.8951	9.2898	8.8951	7.8954
	10	9.8655	8.9178	9.3644	8.9178	7.9558
CODMn	8	2.9376	3.4514	3.2528	4.0023	2.6527
	9	3.0196	3.7110	3.2736	4.0959	2.6583

	10	3.1039	3.9900	3.2945	4.1917	2.6640
NH3-N	8	0.0974	0.3932	0.3271	0.3417	0.1729
	9	0.0944	0.3914	0.3322	0.3442	0.1635
	10	0.0915	0.3895	0.3374	0.3466	0.1545

5.2 问题二的建模与求解

评价类问题常用的方法有层次分析法（AHP）、模糊综合判别法、灰色关联分析（GRA）、主成分分析（PCA）和优劣解距离法（TOPSIS）等[1]，但这些方法侧重点各有不同，前两者属于主观赋权评价法，在该问题中由于缺乏运输订购类问题中相关变量的权重分配研究，进行应用时权重会带有一定主观性和随机性，不利于量化。灰色关联分析更多用于分析序列之间的相关性而非重要性，主成分分析更多则用于降维减少变量维度，并且要求构建的主成分有充分的可解释性，在这一问题中不能很好地达到我们的目的，而 TOPSIS 方法是多目标决策分析中一种常用的有效方法，是一种趋近于理想解的排序法。它根据有限个评价对象与理想化目标的接近程度进行排序，在现有的对象中进行相对优劣的评价[2]，故我们选用 TOPSIS 方法进行求解。

TOPSIS 评价法是有限方案多目标决策分析中常用的一种科学方法，其基本思想为，对原始决策方案进行归一化，然后找出最优方案和最劣方案，对每一个决策计算其到最优方案和最劣方案的欧几里得距离，然后再计算相似度。若方案与最优方案相似度越高则越优先。基本流程如下：

Step1: 根据归一化得到的决策矩阵（这里我们选取 min-max 归一化）和权重向量构造规范化权重矩阵 Z ：

$$Z = [w_j z_{ij}] \quad (1)$$

Step2: 确定正理想解 Z^+ 和负理想解 Z^- ，其中 J^+, J^- 分别表示效益型指标和成本型指标：

$$\begin{aligned} Z^+ &= \left\{ \max_{j \in J^+} z_{ij}, \min_{j \in J^-} z_{ij} \right\} \\ Z^- &= \left\{ \max_{j \in J^+} z_{ij}, \min_{j \in J^-} z_{ij} \right\} \end{aligned} \quad (2)$$

Step3: 计算各评价对象 i ($i=1,2,3,\dots, 402$) 到正理想解和负理想解的欧几里得距离 D_i ：

$$\begin{aligned} D_i^+ &= \sqrt{\sum_{j=1}^n [z_{ij} - Z_j^+]^2} \\ D_i^- &= \sqrt{\sum_{j=1}^n [z_{ij} - Z_j^-]^2} \end{aligned} \quad (3)$$

Step4: 计算各评价对象的相似度 W_i ：

$$W_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (4)$$

可以看到，相似度是与负理想解与两理想解距离之和的比值，若占比越大则说明离负理想解越远，越优先选择。

Step5: 根据 W_i 大小排序可得到结果。

权重向量的构建是 TOPSIS 应用的核心，需要尽可能削弱其主观性。这里我们使用熵权法构建权重向量。熵权法基于信息论，基于信息论的熵值法是根据各指标所含信息

有序程度的差异性来确定指标权重的客观赋权方法，仅依赖于数据本身的离散程度[5]。熵用于度量不确定性，指标的离散程度越大，说明不确定性越大，则最终熵值越大，该指标值提供的信息量越多，则权重也相应越大。

TOPSIS 的思想流程大致如图 3 所示：

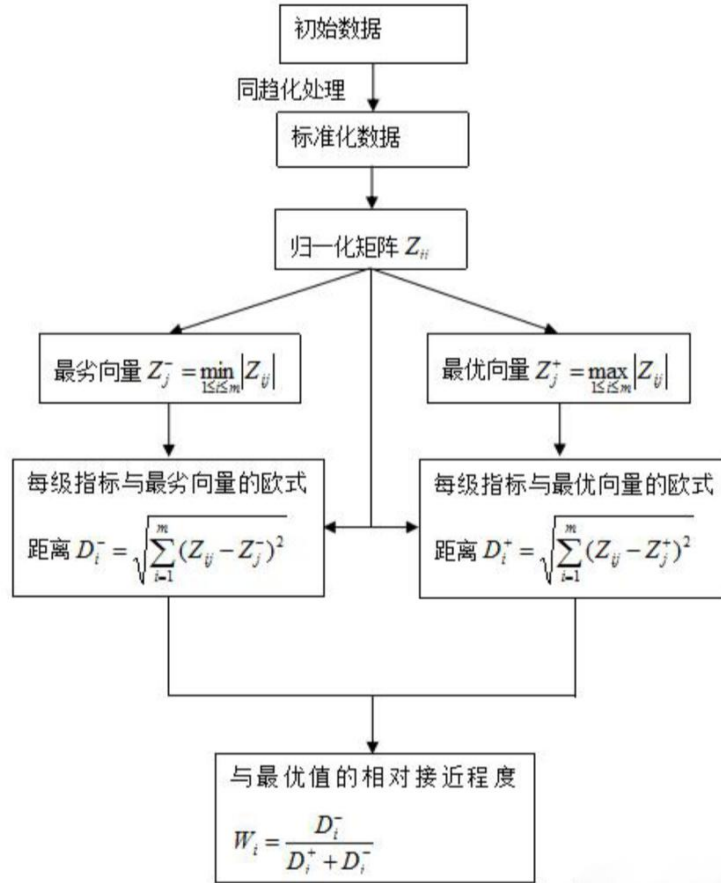


图 3. TOPSIS 方法的基本流程

由于不同指标在评价体系中往往重要性也不同，我们使用熵权法对不同指标计算赋权后再来计算距离。根据信息论中对熵的定义[3]，熵值 e 的计算如下所示：

$$e_j = - \frac{\sum_{i=1}^n p_{ij} \ln p_{ij}}{\ln n} \quad (5)$$

式子(4)中代表对于某一个属性 j ，第 i 类占样本的比例。 n 为属性 j 的取值数量。所以权重系数 w 定义为：

$$w_j = \frac{1 - e_j}{\sum_{i=1}^m (1 - e_i)} \quad (6)$$

当我们得到权重以后，计算距离的方法就更新为：

$$D_i^+ = \sqrt{\sum_{j=1}^n w_j [z_{ij} - Z_j^+]^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^n w_j [z_{ij} - Z_j^-]^2} \quad (7)$$

这样，我们构建了 TOPSIS 综合评价与熵权分析法的综合模型。

下面对问题二进行合理建模。

首先是对数据的预处理与规约化，由于给定的数据是三个月的面板数据，我们将其取平均值得到三个月的综合水平。经分析，CODMn、NH3-N、垃圾密度这三项指标是成本性指标，即越小越好；PH*为区间指标，但越接近 7 越好。所以我们将成本性指标取其倒数，而最优值指标变换为：

$$PH = |PH - 7| \quad (8)$$

得到结果后为防止指标差异过大或过小造成影响，我们将其进行 min-max 规约：

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

至此，预处理阶段完成。基于熵权法得到每项指标的权重，再基于欧几里得距离获得整体的最优解和最劣解，并评价每个地区与最优解和最劣解之间的接近分数可以得到不同地区的水质量情况排名如图 4 所示：

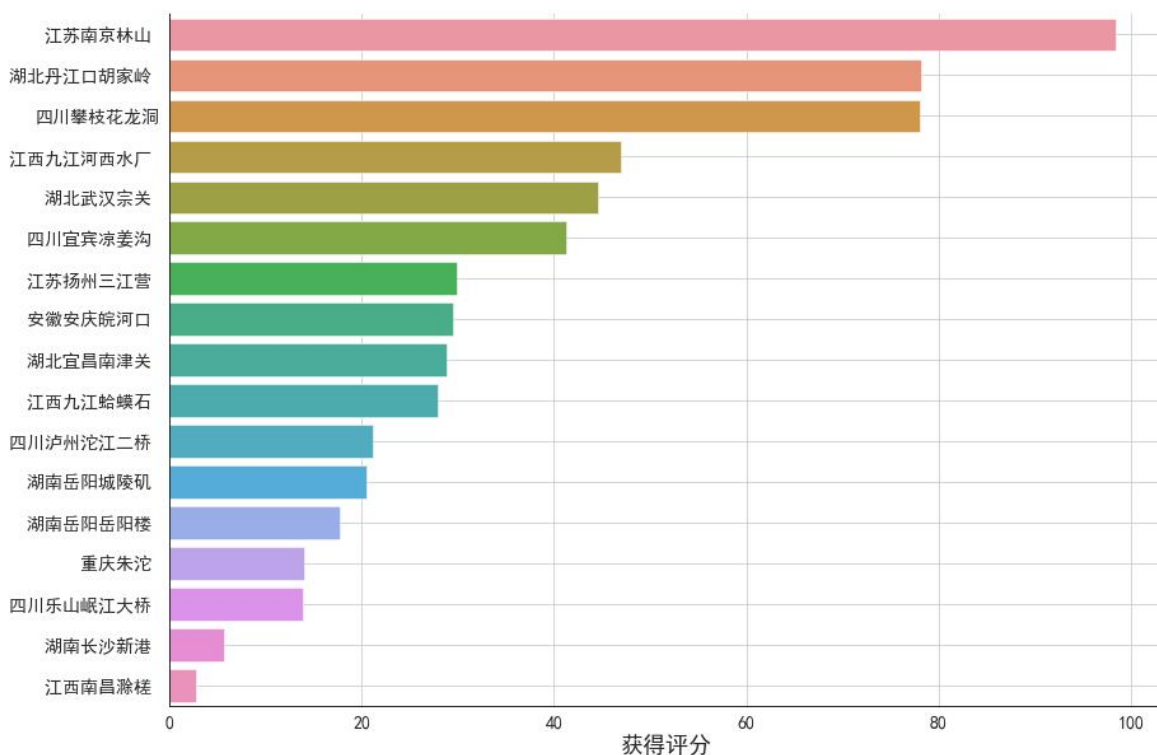


图 4. 不同地区水质量评分排名

从图 4 中可以看到江苏南京林山是最接近最优解的，按百分制计分获得了 98 分；其次湖北丹江口胡家岭、四川攀枝花龙洞等风景区的评分也不低，获得了 78 分；而到重庆朱沱、四川乐山岷江大桥、湖南长沙新港、江西南昌滁槎等工业区的评分则连 20 分都不到。这一结果充分反映，对水资源保护较好的风景名胜，自然风光美好，生态发展良好的同时也带动了当地经济发展，尤其是以湿地生态为主的江苏南京林山、丹江口胡家岭、攀枝花龙洞等地值得参考和借鉴；而过于发展工业的地区不仅生态环境不够好，产业结构也较为单一，不利于当地绿色经济的发展，需要引起有关部门注意。值得学习的是江西九江河西水厂在水资源治理方面的措施，水厂对水资源的消耗和污染应该是相对比较严重的，但江西九江河西水厂的评分排名还相对比较靠前。但总体而言，这 17 个地区的评分若按百分制计算大都没有及格，是需要引起有关部门注意的。

5.3 问题三的建模与求解

对水质影响最大的因素本质上是需要分析不同因素对最终得分贡献的权重，而这一权重在问题二的建模中已经利用熵权法给出。不同因素的权重分布如图 5 所示：

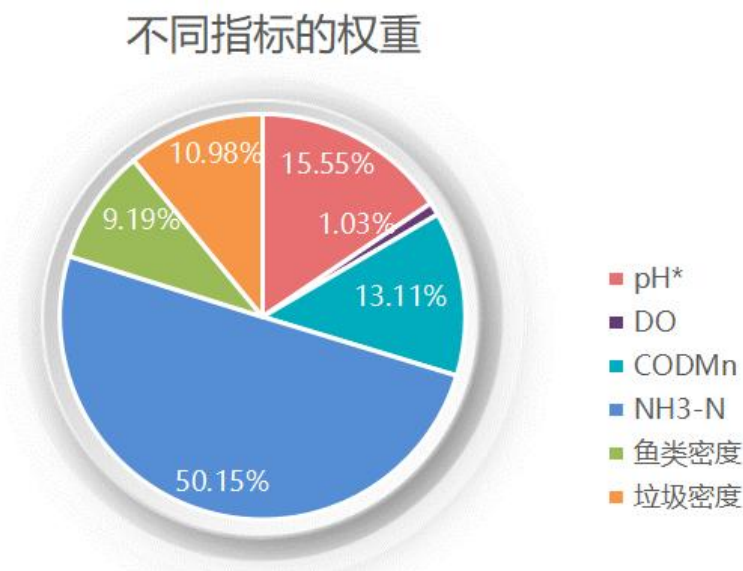


图 5. 不同指标的权重分布

其中，NH₃-N 由于提供信息熵最多，在评价体系中权重占比已经超过 50%。而最少的是水中溶解氧仅为 1.03%。

氨氮含量常做为废水指标，是由于氨氮这两个营养素水体富营养化和环境污染问题的一种关键污染成分，是现阶段导致我国河流湖泊富营养化的直接原因。大批量的氨氮废水排进江河湖泊给化工废水的处理造成了困难，氨氮废水处理中氨转变成为硝酸、硝酸盐更进一步转变成为亚硝酸铵具备比较严重的三致作用，直接危害人类身心健康[5]。

水中二氧化碳的含量是决定水体 PH 的最大因素之一，而水中二氧化碳的浓度又直接与水中浮游生物特别是水植物的含量和活跃程度有直接关系，例如：水中的浮游植物丰富，则白天光合作用强，消耗二氧化碳促进水体 PH 升高，而夜间水中植物由于呼吸作用增强，释放了二氧化碳，造成水中 PH 相对降低[6]。那么 PH 这一因素本质上也可以反映生物的活性，并且适合人体的饮用水 PH 在 6.5-8.5 之间。

CODMn 是锰元素的浓度，以锰元素、铅元素等为代表的重金属元素会对人体造成极大伤害，所以一定要严格控制其浓度不损害人体神经系统。

垃圾密度比较出乎意料，并不是最高的，这是因为现阶段垃圾处理的手段较为丰富并且效果良好，另一方面这些地区的垃圾密度差异并不明显，所以权重并不高。

水中含氧量和鱼类密度都可以反映水的生物活性，含氧量越高越适合生物存活，也就越适合人体饮用，这一点也非常符合常理。

综上，我们向有关部门提出建议：

1. 合理控制水体氨氮量，使得水体中有机物能够被及时分解不造成富营养化。
2. 化工厂在废水处理时进行酸碱度调节，使水体能够适合人体饮用。
3. 严格控制废水中重金属浓度，严防对人体造成伤害，若不合格者可进行惩罚。
4. 生活垃圾处理采取绿色环保的方式并垃圾分类，不直接将其倒入水中。
5. 发展生态经济，在水资源保护的同时建设整个地区生态使地区生态发展的同时以生态作为城市名片发展绿色新经济。

六. 模型优缺点分析

6.1 模型的优点

综合分析，我们认为我们的模型有以下优点：

（1）采用了合适的预处理，使结果更合理，避免了由于未考虑效益指标和成本指标的方向性不同带来的较大误差。

（2）所选取的指标既考虑了其在数据意义上的权重，又具有一定的实际意义，从而使问题描述更加清晰，具有更加明确的应用意义，较好地解释了水体质量与各项观测指标之间的关系。

（3）所用模型考虑全面，模型简洁高效，便于理解和运用：熵权法和 TOPSIS 的效率评价模型避免了数据方差不齐的问题；采用灰色预测的方法，问题解答层层递进，得到求解结果与实际情况较为接近，结果较为合理。

6.2 模型的缺点

尽管我们的模型已经做的比较合理，我们认为它仍然有一些不足之处：

（1）灰色预测的结果虽然比传统的回归方法要更好，但效果仍然欠佳。我们认为第一这是由于数据本身太少导致，第二是灰色方程中没有考虑离群点对模型的影响。

（2）熵权法和 TOPSIS 分析虽然能够合理对不同地区进行评分，却没有将可观测的指标抽象为更高级的原因而是直接获得评分系统。

6.3 模型的改进

针对上述问题，我们提出以下改进思路：

（1）当数据体量足够大时我们可以使用 ARIMA 系列模型进行合理预测。

（2）灰色预测模型的灰色方程与对应的白化微分方程不变，但在最小二乘拟合时可以仿照岭回归加入关于 a 的 L1 或 L2 正则化项再进行拟合，能够对离群点进行一些约束与控制。

（3）除了使用熵权法和 TOPSIS 分析外，主成分分析和因子分析法也能够进行评价模型，而对应的特征值就可以作为抽象级因子的权重。

参考文献

- [1]. 刘雪梅. 上市公司治理结构与盈余管理关系研究[D].华北电力大学（北京）,2008.
- [2]. 朱强军.基于改进的 TOPSIS 算法的企业信贷风险评价——以中小微企业为例[J].长春大学学报,2022,32(07):24-33.
- [3]. 谢宏,程浩忠,牛东晓.基于信息熵的粗糙集连续属性离散化算法[J].计算机学报,2005(09):1570-1574.
- [4]. 张二丽,胡玉林,汪太行.灰色 GM(1,1)模型在乙醇偶合制备 C₄ 烯烃预测中的应用[J].科技资讯,2022,20(15):220-223.DOI:10.16661/j.cnki.1672-3791.2201-5042-4430.
- [5]. 尹艳娥,胡中华,刘亚菲,唐文伟.生物活性炭纤维处理微污染原水研究[J].给水排水,2006(S1):56-58.DOI:10.13789/j.cnki.wwe1964.2006.s1.016.
- [6]. 刘音,马吉燕,程兴娜,陈森,杜玉娇,王婷婷.污泥基充填体浸出水样水质及健康风险评估[J].能源环境保护,2022,36(03):97-102.

python 环境: anaconda(windows) python 3.7.6

部分核心代码如下:

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
import numpy as np
import sympy as sp
x0=np.array([25723,30379,34473,38485,40514,42400,48337])
def GM11(x0):
    x0=np.array(x0)
    n=len(x0)
    x1=np.cumsum(x0) #求累加序列
    z=(x1[:-1]+x1[1:])/2.0
    B=np.vstack([-z,np.ones(n-1)]).T
    u=np.linalg.pinv(B)@x0[1:] #最小二乘法拟合参数

    t=sp.var('t')
    x=sp.var('x',cls=sp.Function) #定义符号变量和函数
    eq=x(t).diff(t)+u[0]*x(t)-u[1] #定义符号微分方程
    xt0=sp.dsolve(eq,ics={x(0):x0[0]}) #求解符号微分方程
    xt=xt0.args[1] #提取方程中的符号解
    xt=sp.lambdify(t,xt,'numpy') #转换为匿名函数
    t=np.arange(n,n+4)
    xt1=xt(t) #求模型的预测值
    x0_pred=np.diff(xt1) #还原数据
    t=np.arange(n)
    delta=np.diff(xt(t))-x0[:-1]
    print('预测值: ',x0_pred,'方程: ',xt0,'MSE: ',sum(delta**2/x0[:-1])/n)
    return x0_pred
GM11(x0)
newdata['pH*']=abs(newdata['pH*']-7)
newdata['CODMn']=1/newdata['CODMn']
newdata['NH3-N']=1/newdata['NH3-N']
newdata['垃圾密度']=1/newdata['垃圾密度']
def entropyWeight(data):
    data = np.array(data)
    # 归一化
    P = data / data.sum(axis=0)
    # 计算熵值
    E = np.nansum(-P * np.log(P) / np.log(len(data)), axis=0)
    # 计算权系数
    return (1 - E) / (1 - E).sum()
#TOPSIS 方法函数
def Topsis(A1):
    W0=[0.15547967,
        0.01030048,
        0.13105116,
        0.50152061,
        0.09189262,
        0.10975546]
    W=np.ones([A1.shape[1],A1.shape[1]],float)
    for i in range(len(W)):
        for j in range(len(W)):
            if i==j:
                W[i,j]=W0[j]
```

```

        else:
            W[i,j]=0
    Z=np.ones([A1.shape[0],A1.shape[1]],float)
    Z=np.dot(A1,W) #加权矩阵

    #计算正、负理想解
    Zmax=np.ones([1,A1.shape[1]],float)
    Zmin=np.ones([1,A1.shape[1]],float)
    for j in range(A1.shape[1]):
        Zmax[0,j]=max(Z[:,j])
        Zmin[0,j]=min(Z[:,j])

    #计算各个方案的相对贴适度 C
    C=[]
    for i in range(A1.shape[0]):
        Smax=np.sqrt(np.sum(np.square(Z[i,:]-Zmax[0,:])))
        Smin=np.sqrt(np.sum(np.square(Z[i,:]-Zmin[0,:])))
        C.append(Smin/(Smax+Smin))
    C=pd.DataFrame(C,index=['四川攀枝花龙洞', '重庆朱沱 ', '湖北宜昌南津关 ', '湖南岳阳城陵矶 ', '江西九江河西
水厂 ', '安徽安庆皖河口 ',
        '江苏南京林山 ', '四川乐山岷江大桥 ', '四川宜宾凉姜沟 ', '四川泸州沱江二桥 ', '湖北丹江口胡家岭 ', '湖南
长沙新港 ',
        '湖南岳阳岳阳楼 ', '湖北武汉宗关 ', '江西南昌滁槎 ', '江西九江蛤蟆石 ', '江苏扬州三江营 '])
    return C
result=Topsis(np.array(newdata))

```