

改进 RAGA – PPC 模型在河湖水质评价及应急监测中的应用

龚 慧¹, 姚 敏², 王 丽³, 吴荣荣¹

(1. 江苏省水文水资源勘测局常州分局, 江苏 常州 213000;

2. 江苏省水文水资源勘测局, 江苏 南京 210000;

3. 鲁南煤化工研究院, 山东 济宁 272000)

摘要:为得出应用于河湖水质评价和应急监测中的标准模型,本文基于 RAGA – PPC 模型对太湖水质进行了评价,并对 5 种不同水质指标数值进行了模拟,结果表明:该模型按照投影追踪的原理,将水质等级分为了 5 个等级,太湖 3 个区域核心区、缓冲区和开发控制利用区的水质分别属于 II 级、III 级和 II 级,结论与实测值基本一致,而基于该模型对水质指标进行模拟结果可知,模型模拟结果的误差较小,与实测值的一致性较高,表明模型模拟精度较高,验证了模型在水质评价与应急监测中的先进性和科学性。

关键词:水质评价; 应急监测; RAGA – PPC 模型; 水质等级; 模拟精度

中图分类号: X824

文献标识码: A

文章编号: 1007-7839(2019)09-0005-04

Application of improved RAGA – PPC model in water quality assessment and emergency monitoring of rivers and lakes

GONG Hui¹, YAO Min², WANG Li³, WU Rongrong¹

(1. Changzhou Hydrology and Water Resources Survey Bureau of Jiangsu Province, Changzhou 213022, Jiangsu;

2. Jiangsu Hydrology and Water Resources Survey Bureau, Nanjing 210029, Jiangsu;

3. Lunan Coal Chemical Industry Research Institute, Jining 272000, Shandong)

Abstract: In order to obtain the standard model applied in river water quality assessment and emergency monitoring, the water quality of Gehu based on RAGA – PPC model was evaluated, and the values of five different water quality indicators were simulated. The results showed that the model was divided into five grades according to the principle of projection tracking. The water quality of the three regional core areas, buffer zones and development control utilization areas of Gehu belonged to Grade II, Grade III and Grade II respectively. The measured values were basically the same. The error of the model simulation results was small, and the consistency with the measured values was high. It indicated that the model had high simulation accuracy, which verified the advanced and scientific nature of the model in water quality assessment and emergency monitoring.

Key words: water quality assessment; emergency monitoring; the RAGA – PPC model; water quality level; simulation accuracy

收稿日期: 2019-01-14

作者简介: 龚慧(1982—), 女, 硕士, 主要从事水文水资源相关工作。

0 引言

河流存在于人民生活区周围,是人们赖以生存和社会文化经济发展的基础,直接关系到人们的身体健康和居住质量^[1]。但是随着目前工业和经济的快速发展,工业废水、生活污水及污染企业的排水未经处理排水河道内的形象十分严重,已严重威胁到河流水质及周围居民的身体健康^[2-3]。河流水质评价主要通过一定的方法研究河流的健康程度,当河流水质健康程度较低时,就需要及时采取相应措施进行治理^[4]。同时由于河流水质在年内的变化趋势具有一定的规律性,找出合理的方法对河流水质进行监测,是河流水质防治预防措施制定的关键数据依据^[5]。因此,对河流水质进行评价与监测是保证河流水质良好的必要工作。

目前,针对河流水质评价的模型包括层次分析法、SPA 法、模糊评价法^[6-8]等,这些评价算法多从主观角度或客观角度的单一角度出发,无法全面反映事物的整体性。而针对河流水质预测模型主要集中于贝叶斯原理、机器学习模型、模糊评价算法等^[9-10],这些模型存在收敛速度慢、计算精度较低等缺点,限制了模型的应用。遗传算法—投影寻踪(RAGA-PPC)模型是近年提出一种新的技术,该模型将高维数据根据转化为低维数据,去除与数据结构无关的变量干扰。将该模型应用于河流水质监测及风险评价中,可利用模型在数据降维及主成分分析中的优势,合理分配各指标的权重,最终可较客观和真实地反映出各指标对河流健康风险评价的影响程度。

本研究拟将模型用于河流水质评价及应急监测预测中,可提高模型评价及预测的准确性及精度,基于 Matlab 软件建立河流水质评价及应急监测模型。同时基于实测数据,分析了实测数据的变化趋势,基于 RAGA-PPC 模型对河流各水质指标数值进行预测,本文的研究结论可为河流水质防护措施的制定提供理论依据。

1 RAGA-PPC 模型构建

1.1 投影寻踪模型构建

建立指标样本集 $\{x^*(i, j) \mid i = 1, 2, \dots, n; j = 1, 2, \dots, p\}$, 其中 i, j 分别代表样本 i 和指标值 j ; n, p 分别对应样本容量和指标数目。首先对数据进行预处理,对于越大越优指标和越小越优指标分别用公式(1)和(2)进行预处理,具体公式如下:

$$x(i, j) = \frac{x_{\max}(j) - x^*(i, j)}{x_{\max}(j) - x_{\min}(j)} \quad (1)$$

$$x(i, j) = \frac{x^*(i, j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad (2)$$

式中, $x_{\max}(j)$ 和 $x_{\min}(j)$ 分别为指标 j 的最大值和最小值。

其次建立投影寻踪模型 P, 假设 $a = \{a(1), a(2), a(3), \dots, a(p)\}$ 为投影方向的一维投影值, 其公式可表示如下:

$$z(i) = \sum_{j=1}^p a(j)x(i, j) \quad (3)$$

综合考虑局部投影点尽可能密集, 最好凝聚成若干个团, 而在整体上投影点团之间要尽可能散开, 构造一个投影指标函数, 可以表述如下:

$$Q(a) = S_z D_z \quad (4)$$

式中, S_z 为类间散开度, 可用 $z(i)$ 的标准差代替; D_z 为类内密集度, 可表示为 $z(i)$ 的局部密度, 具体计算公式如下:

$$S_z = \sqrt{\frac{\sum_{i=1}^n [z(i) - E(z)]^2}{n-1}} \quad (5)$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^n (R - r(i, j)) \cdot U(R - r(i, j)) \quad (6)$$

式中, $E(z)$ 为序列 $\{z(i) \mid i = 1, 2, \dots, n\}$ 的均值; R 为局部宽度参数, 由数据特征确定, 其值一般可取 0, 1; $r(i, j)$ 表示样本之间的距离, $r(i, j) = |z(i) - z(j)|$; 符号函数 $U(t)$ 为单位阶跃函数, 其中 $t = R - r(i, j)$, 当 $t \geq 0$ 时函数值取 1, 反之则取为 0。

1.2 投影指标函数优化

最佳投影方向是数据信息利用最充分、信息损失量最小的方向, 优化投影方向归根到底是找出某种意义上好的投影指标。因此, 通过求解投影指标函数最大化来估计最佳投影方向。

$$\text{目标函数最大化: } \max Q(a) = S_z D_z \quad (7)$$

$$\text{约束条件: } \sum_{j=1}^p a^2(j) = 1 \quad (8)$$

1.3 遗传算法优化模型

规定投影方向的长度范围为单位长度, 则投影矩阵 A 的模为 1, 同时确定投影指标 Q 为目标函数, 实属编码的加速遗传算法的优化问题是:

$$\max Q(A), \|A\| = 1 \quad (10)$$

$$\min Q(A), \|A\| = 1$$

根据投影指标的选择, 进行加速遗传算法程序, 确定最大及最小指标函数值, 其对应的投影方向即为最优投影方向^[11]。

2 结果与分析

2.1 水质指标实测数值变化趋势分析

本文选择江苏省太湖为研究区域,将全湖区分为核心区、缓冲区和开发控制利用区 3 个区域,分别测定区域内总氮、总磷、化学需氧量、五日生化需氧量和高锰酸盐指数 5 种指标的含量。图 1 为江苏省太湖 5 种水质指标年内变化趋势。图 1 显示,不同指标在年内的变化趋势基本均呈先降低后增加的趋势,各项指标在 1 月、2 月、11 月和 12 月的值普遍较高,在 3~10 月的含量较低,这可能是由于水体中植物起到了关键的净化水质作用,植物在 11 月~2 月温度较低时,存活率普遍偏低,在 3~10 月普遍存活率偏高,增加了对水体的净化作用,使各指标值较低。不同区域的总氮含量不同,总氮含量在缓冲区普遍较高,在开发控制利用区普遍较低,最高可高出 30.0%,缓冲区总氮含量最高达到了 4.65 mg/L,已严重超标;不同区域总磷含量在缓冲区普遍较高,在开发控制利用区普遍较低,最高可高出 31.0%,在 3~10 月,不同区域总磷含量基本维持在较低水平;不同区域化学需氧量在年内保持基本不变的趋势,不同区域 1~12 月的降低趋势仅维持在 0.43 mg·L⁻¹·a⁻¹左右;不同区域五日生化需氧量的变化趋势与化学需氧量基本一致,不同区域 1~12 月的降低趋势仅维持在 0.67 mg·L⁻¹·a⁻¹左右;高锰酸盐指数呈先降低后增加的趋势,在 1 月的值最高,在 3 月的值最低,不同区域 1~12 月的降低趋势维持在 0.59 mg·L⁻¹·a⁻¹左右。综上所述,应加强太湖在 11 月~2 月的水质防治措施,提高在该时段的水质。

2.2 基于 RAGA-PPC 模型的水质评价

根据《地表水环境质量标准》(GB3838-2002)中给定的水质参数取值及分级标准,选定初始种群 n=500,交叉概率 P_c=0.8,变异概率 P_m=0.8,优秀个体数量为 30,显著性检验水平 α=0.01,遗传算法中的遗传代数数为 200,得到的最大投影指标为 200.3766。利用 RAGA-PPC 模型进行数据转化,得出的最佳投影方向为 x^{*}=[0.431,0.537,0.079,0.149,0.274],地表水 5 个等级投影值计算结果可用表 1 来表示。

通过类似的步骤,求出不同区域实测指标的投影值,最终可得出不同区域的水质情况,结果见表 2 所示。由表 2 可以看出,核心区和开发控制利用区的水质等级为 II 类,而开发控制利用区水质等级为

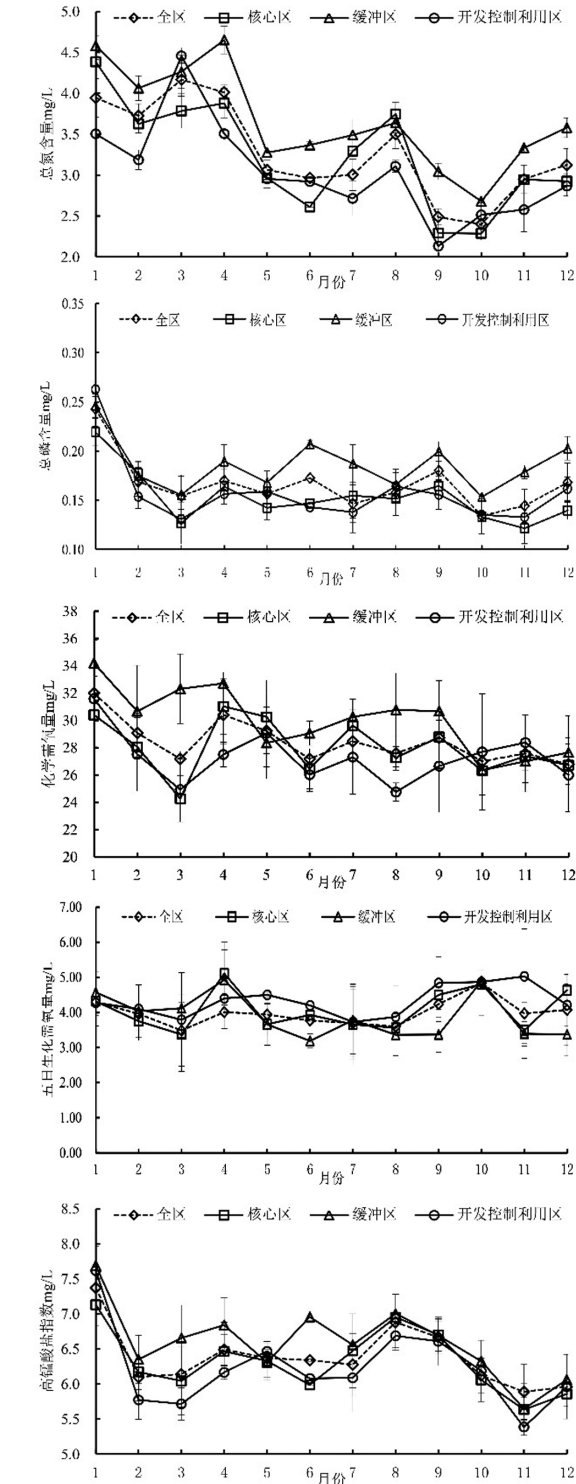


图 1 江苏省太湖不同水质指标年内变化趋势

表 1 地表水评价水平投影值分级结果

等级	I	II	III	IV	V
投影值区间	0	0.0364	0.710	1.200	1.539

III类,水质较差,3 个区域水质排名由高到低依次为开发控制利用区、核心区和缓冲区,与前文水质的

实测值结果基本一致,验证了 RAGA - PPC 模型的准确性。

表 2 不同区域水质评价结果

分区	投影值	最终评价等级
核心区	0.0301	II
缓冲区	0.536	III
开发控制利用区	0.0274	II

2.3 基于 RAGA - PPC 模型的水质应急监测结果分析

基于 RAGA - PPC 模型建立河湖水质的应急监测预测模型,通过实测数据分析,预测出河湖水质的变化情况,从而做到提前预警。预测结果的精度可由以下指标来进行判断,通过计算不同区域不同水质指标的模拟精度,从而得出模型的精度,验证模型的准确性。采用相对误差(RE)、模型效率系数(E_{ns})和决定系数(R^2)综合分析精度,具体计算公式如下:

$$RE = \frac{(Y_i - X_i)}{X_i} \times 100 \quad (11)$$

$$E_{ns} = 1 - \frac{\sum_{i=1}^m (Y_i - X_i)^2}{\sum_{i=1}^m (X_i - \bar{X})^2} \quad (12)$$

$$R^2 = \frac{[\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^m (X_i - \bar{X})^2 \sum_{i=1}^m (Y_i - \bar{Y})^2} \quad (13)$$

式中, Y_i 为模型的模拟值, \bar{Y} 为其平均值; X_i 为 P - M 模型计算的标准值, \bar{X} 为 X_i 的平均值; m 为数据样本数。 RE 越小,表明模型模拟误差越小; E_{ns} 和 R^2 越接近 1,表明模型模拟精度越高。

表 3 为基于 RAGA - PPC 模型对水质指标的模拟结果精度对比。由表 3 可以看出,该模型的计算精度较高,基于该模型模拟 5 种不同水质指标时,

相对误差 RE 均在 10.0% 以下,而 E_{ns} 和 R^2 基本均在 0.80 以上,且与实测值的相关性均通过了 $P < 0.01$ 的极显著水平检验,表明模型模拟值的误差较低而一致性较高。同时,缓冲区的模拟精度要低于开发控制利用区和核心区,这可能是由于缓冲区水质较差,水质变化规律不明显,造成了模型模拟时训练精度不够,降低了模型的模拟精度。

3 结论

本文基于 RAGA - PPC 模型模拟了河流水质,并对水质进行了评价,指出研究区域不同分区的水质分类级别,并对 5 种水质指标进行了模拟,结果表明,模型评价结果与实测结果基本一致,同时模拟结果的精度与实测值的误差较小,一致性较高,反映了 RAGA - PPC 模型在河湖水质评价及应急监测中的科学性。

参考文献:

- [1] Lau S S S, Lane S N. Biological and chemical factors - influencing shallow lake eutrophication: a long - term study[J]. The Science of the Total Environment, 2002, 288:167 - 181.
- [2] 刘国东,黄川友,丁晶. 水质综合评价的人工神经网络模型[J]. 中国环境科学, 1998, 18(6):514 - 517.
- [3] Zhou M, Shen Z, Yu R. Responses of a coastal phytoplankton community to increased nutrient input from the Changjiang (Yangtze) River[J]. Continental Shelf Research, 2008, 28:1483 - 1489.
- [4] 王宏伟,张伟,杨丽坤,等. 中国河流健康评价体系[J]. 河北大学学报(自然科学版), 2011, 31(6):669 - 672.
- [5] 朱剑锋. 淀浦河水体中氨氮、总氮和总磷污染变化趋势及相关性分析[J]. 北方环境, 2013(6):155 - 159.

(下转第 13 页)

表 3 模型模拟精度结果

指标	核心区			缓冲区			开发控制利用区		
	$RE(\%)$	E_{ns}	R^2	$RE(\%)$	E_{ns}	R^2	$RE(\%)$	E_{ns}	R^2
总氮	3.7	0.891**	0.880**	8.7	0.817**	0.851**	2.8	0.868**	0.900**
总磷	4.1	0.847**	0.838**	8.6	0.831**	0.824**	3.0	0.884**	0.879**
化学需氧量	4.0	0.882**	0.887**	9.7	0.826**	0.832**	2.7	0.878**	0.886**
五日生化需氧量	3.8	0.836**	0.837**	10.1	0.830**	0.808**	2.7	0.875**	0.892**
高锰酸盐指数	3.5	0.887**	0.853**	9.5	0.814**	0.823**	1.6	0.881**	0.903**

(上接第 8 页)

- [6] 宋述军,周万村. 沱江流域地表水水质的模糊综合评价[J]. 水土保持研究, 2007, 14(6):128-130+134.
- [7] 徐晨光,艾尼瓦尔·艾买提,杜青辉. 基于 SPA 和信息熵的可变模糊集水质评价方法[J]. 人民黄河, 2013, 35(6):56-58.
- [8] 卢文喜,李迪,张蕾,等. 基于层次分析法的模糊综合评价在水质评价中的应用[J]. 节水灌溉, 2011, (3):43-46.
- [9] 曹宇峰,林春梅,孙霞. 模糊数学法在海洋水质评价中的应用[J]. 海洋技术, 2011, 30(2):118-122.
- [10] 杨咪,屈文岗,钱会. 基于熵权的贝叶斯模型及其在水质评价中的应用[J]. 灌溉排水学报, 2018, 37(1):1-7.
- [11] 金菊良,杨晓华,丁晶. 基于实数编码的加速遗传算法[J]. 四川大学学报(工程科学版), 2000, 32(4):20-24.