

基于投影寻踪与内梅罗指数组合模型的地下水水质评价

李 琪, 赵志怀

(太原理工大学 水利科学与工程学院, 山西 太原 030024)

摘要: 针对内梅罗指数法注重最大污染指标及未充分考虑指标权重的问题, 通过投影寻踪模型构造投影指标函数, 采用实码加速遗传算法对投影指标函数进行优化得到指标权重, 与污染指数结合得到加权污染指数, 并将改进的最大污染指数加入权重与污染指数乘积最大值, 构成投影寻踪与内梅罗指数组合模型。结合阳泉市娘子关泉域地下水水质监测样本, 得到评价结果。结果表明, 组合模型的地下水水质评价结果与现行的水质评价方法比较在安全性与准确性方面均有一定优势。验证了组合模型的可行性与适用性。研究成果可为地下水水污染防治提供参考。

关键词: 内梅罗指数法; 投影寻踪; 实码加速遗传算法; 地下水水质; 评价

中图分类号: X824

文献标志码: A

1 引言

准确评价地下水水质对于水污染防治有积极意义。目前, 评价地下水水质的常用方法有主成分分析法^[1]、模糊综合评价法^[2]、人工神经网络法^[3]、内梅罗指数法等。这些评价方法均有各自的局限性, 如主成分分析法在水样较少时对结果的精度影响很大^[4]; 模糊综合评判法在确定指标权重方面存在明显不足^[5]; 人工神经网络法由于存在调参繁琐、稳定性差等缺陷, 最终导致精度不理想^[6]; 内梅罗指数法评价结果过分强调最大污染指数, 且没有为各指标赋予准确的权重^[7], 从而得不到准确的评价结果。因此, 本文将投影寻踪模型与内梅罗指数法结合构成组合模型, 以阳泉市娘子关泉域为例, 验证了该方法的可行性与适用性。

2 投影寻踪与内梅罗指数组合模型

2.1 内梅罗指数法

内梅罗指数法是一种强调最大污染指标的等权重多因子评价方法^[8], 其计算公式为:

$$F(i, j) = c(i, j) / s(j) \quad i = 1 \sim m; j = 1 \sim n \quad (1)$$

$$F_{ave}(i) = \frac{1}{n} \sum_{j=1}^n F(i, j) \quad (2)$$

$$P(i) = \sqrt{[F_{max}^2(i) + F_{ave}^2(i)] / 2} \quad (3)$$

式中, $c(i, j)$ 、 $F(i, j)$ 分别为第 i 个水质样本第 j 个评价因子的浓度监测值与污染指数; $s(j)$ 为第 j 个评价因子的标准值, 通常取地下水质量标准中Ⅲ类标准对应因子的上限值; $F_{ave}(i)$ 、 $F_{max}(i)$ 分别为第 i 个水质样本所有评价因子的平均污染指数与最大污染指数; $P(i)$ 为第 i 个水质样本的内梅罗指数。

2.2 投影寻踪法简介

投影寻踪模型^[7]是通过某种坐标组合将高维数据投影得到能反映其某种特征或结构的低维数据, 不同的投影方向代表的高维数据的结构或特征各不相同, 最佳投影方向最有可能反映高维数据某类结构特征, 求解最佳投影方向即是对投影指标函数的最大化寻优。

2.2.1 评价指标集的归一化处理

设各指标值(监测值与标准值)组成的样本集为 $\{c^*(i, j) \mid i = 1 \sim m, j = 1 \sim n\}$, 其中 $c^*(i, j)$ 为第 i 个样本的第 j 个指标, m 、 n 分别为样本数、指标数。由于各指标的量纲存在差异且某些指标数值相差过大, 故先对各指标采用归一化处理方法消除量纲影响。

归一化处理方式为:

收稿日期: 2019-01-10, 修回日期: 2019-02-23

作者简介: 李琪(1990-), 男, 硕士研究生, 研究方向为水资源与水环境, E-mail: 794484474@qq.com

通讯作者: 赵志怀(1964-), 男, 副教授, 研究方向为水资源与水环境、水文地质, E-mail: 2260977881@qq.com

$$c(i, j) = \frac{c^*(i, j) - c_{\min}(j)}{c_{\max}(j) - c_{\min}(j)} \quad (4)$$

式中, $c_{\max}(j)$ 、 $c_{\min}(j)$ 分别为样本中第 j 个指标的最大值、最小值; $c(i, j)$ 为指标经处理消除量纲后的值。

2.2.2 构造投影指标函数

投影寻踪模型是以 $a = \{a(1), a(2), \dots, a(n)\}$ 为投影方向将归一化处理后的 n 维数据 $\{c(i, j) \mid j=1 \sim n\}$ 投影到一维子空间上变换成一维投影值数据 $z(i) = \sum_{j=1}^n a(j)c(i, j) (i=1 \sim$

$m)$, 其中 $|a| = \sum_{j=1}^n a^2(j) = 1$ 。

投影指标函数 $Q(a)$ 可表示为:

$$Q(a) = S_Z D_Z \quad (5)$$

$$S_Z = \sqrt{\frac{1}{m-1} \sum_{i=1}^m [Z(i) - E(z)]^2} \quad (6)$$

$$D_Z = \sum_{i=1}^m \sum_{j=1}^n [R - r(i, j)] u[R - r(i, j)] \quad (7)$$

$$r(i, j) = |z(i) - z(j)| \quad (8)$$

式中, S_Z 、 E_Z 、 D_Z 分别为一维投影值 $z(i)$ 的标准差、期望及局部密度; R 为局部密度 D_Z 的窗口半径, 一般取经验值 $0.1S_Z$; $r(i, j)$ 为第 i 个与第 j 个样本之间的距离; $u(t)$ 为一单位阶跃函数, 当 $t \geq 0$ 时, 函数值为 1, 当 $t < 0$, 函数值为 0。

2.2.3 优化投影指标函数

在给定各评价指标样本集的情况下, 使投影指标函数值 $Q(a)$ 发生变化的唯一因素就是投影方向 a 。投影方向直接关系到高维数据的结构特征, 投影方向不同, 反映出的高维数据结构特征也不相同, 最佳投影方向比其他投影方向有更大可能揭示高维数据的某类结构特征, 是投影指标函数最大值求解问题得到的最优解。这是一个以 $\{a(j) \mid j=1 \sim n\}$ 为自变量, 以 $Q(a) = S_Z D_Z$ 为目标函数且在非线性约束条件 $|a| = \sum_{j=1}^n a^2(j) =$

1 下的最大值复杂优化问题, 无法用传统方法解决, 本文采用实码加速遗传算法寻求最优解。

实码加速遗传算法 (简称 RAGA) 是利用标准遗传算法 (SGA) 前两次进化迭代产生的优秀个体的区间 (变量的最优解区间) 作为变量的搜索区间^[9], 在缩小的搜索区间上使遗传算法加速运行, 重复前述操作继续缩小搜索区间, 继续加速直到个体目标函数值符合要求从而得到最优解, 提高了寻优效率, 解决了一般智能算法在寻优过程中过早停滞而导致精度低的问题, 得到更为准确的指标权重, 提高评价结果精度。

RAGA 的操作过程^[9]为: ①编码; ②初始化父代种群; ③对父代种群进行适应度评价; ④分别通过选择、杂交、变异操作产生第一、二、三代子群体; ⑤对三代子群体进行演化迭代; ⑥加速循环。

2.3 投影寻踪与内梅罗指数组合模型

在内梅罗指数法的基础上引入通过投影寻踪模型求解得到的指标权重 $\omega(j) (j=1 \sim m)$, 得到基于权重的加权污染指数 $F'_{ave}(i)$, 并进一步改进最大污染指数 $F'_{max}(i)$, 最终得到组合内梅罗指数 $P'(i)$, 单因子污染指数仍按式 (1) 计算。

投影寻踪与内梅罗指数组合模型的计算公式为:

$$F'_{ave}(i) = \sum_{j=1}^n \omega(j) F(i, j) \quad (9)$$

$$F'_{max}(i) = [F_{max}(i) + F_{\omega} + \max \omega(j) F(i, j)] / 3 \quad (10)$$

$$P'(i) = \sqrt{[F'^2_{max}(i) + F'^2_{ave}(i)] / 2} \quad (11)$$

式中, F_{ω} 为权重最大指标的污染指数; $\max \omega(j) \cdot F(i, j)$ 为权重与污染指数乘积最大的污染指数。

2.4 组合模型用于地下水水质评价的流程

组合模型水质评价流程见图 1。

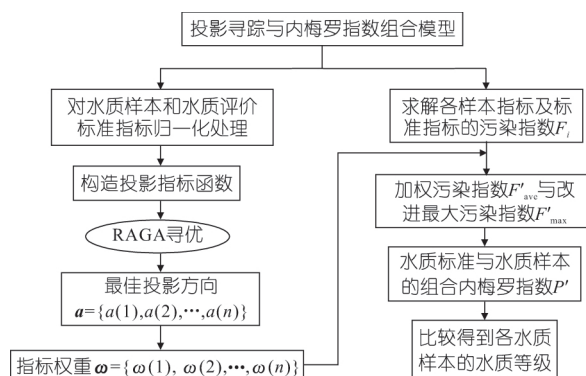


图 1 组合模型水质评价流程

Fig. 1 Combined model water quality evaluation flow chart

3 应用实例

娘子关泉域位于山西省东部, 太行山中段西侧, 地跨阳泉市城区、郊区、平定县、盂县, 晋中市榆次区、寿阳县、昔阳县、和顺县、左权县、太原市的迎泽区等地, 大多为所在区域内的水源地。以 2003 年娘子关泉域阳泉地区地下水重点供水水源地 20 例水质监测样本为例, 选取溶解性总固体 (TDS)、硫酸盐 (SO_4^{2-})、氯化物 (Cl^-)、硝酸盐 (NO_3^-)、亚硝酸盐 (NO_2^-)、氟化物 (F^-) 6 个指标作为水质评价指标, 采用基于投影寻踪与内梅罗指数的组合模型进行水质评价。表 1 为该地区水质评价指标的浓度监测值。

表 1 各水样评价指标监测浓度

Tab. 1 Monitoring concentration of each water

sample evaluation index		mg/L				
地点	TDS	SO ₄ ²⁻	Cl ⁻	NO ₃ ⁻	NO ₂ ⁻	F ⁻
乌玉深井	339.7	52.2	11.3	2.48	0	0.22
荫营煤矿水源地	1 374.3	769.5	40.0	4.87	0.000 6	0.56
韩庄	1 390.0	687.0	75.9	3.36	0.000 9	0.33
苇钢	712.0	344.0	16.0	0.71	0.000 6	0.39
柴家庄	1 190.0	523.0	48.2	1.82	0	0.62
大石门深井	630.0	338.0	16.3	0.37	0	0.67
俱乐部	1 122.4	511.5	66.0	9.44	0.001 5	0.62
西峪深井	526.0	185.0	12.4	0.58	0	0.35
河下	538.0	205.0	21.3	1.25	0	0.38
小河	646.3	230.7	45.2	7.75	0.000 6	0.65
王龙	430.0	125.0	28.4	1.33	0	0.32
龙庄深井	920.0	324.0	104.0	2.57	0.000 9	0.96
南坪	1 570.0	680.0	117.0	8.38	0.059 7	0.27
城西泉	644.0	224.0	52.1	1.21	0.000 9	0.96
坡底泉	658.0	214.0	49.3	1.19	0.000 9	0.94
五龙泉	682.0	214.0	56.0	0.78	0.000 6	0.62
苇泽关泉	698.0	244.0	54.9	0.65	0	0.52
集泉站	702.0	238.0	54.9	0.77	0	0.49
五矿贵石沟	543.8	231.7	14.8	2.97	0.000 9	0.40
李家庄	44.7	18.8	6.1	0.30	0	0.07

评价过程为:①在地下水质量标准的每级水质范围内随机产生五级各 200 个共计 1 000 个水质样本;②对随机水质样本的指标浓度及标准中的边界浓度用式(4)进行归一化处理;③对归一化后的水质样本通过加速遗传算法进行选择、杂交、变异、演化迭代和加速循环得到最佳投影方向 $\alpha = (0.412\ 4, 0.393\ 3, 0.395\ 5, 0.409\ 5, 0.448\ 5, 0.387\ 3)$;④权重向量为 $\omega = \alpha^2 = (0.170\ 1, 0.154\ 7, 0.156\ 4, 0.167\ 7, 0.201\ 2, 0.150\ 0)$,向量中各项分别为各指标的权重;⑤用式(1)对各级标准的边界浓度值进行处理得到各单因子边界污染指数;⑥将求得的指标权重通过式(9)、(10)与单因子边界污染指数结合得到 $F'_{ave}(i)$ 、 $F'_{max}(i)$;⑦将 $F'_{ave}(i)$ 与 $F'_{max}(i)$ 代入式(11)得到地下水水质分级标准,见表 2。

表 2 组合模型水质分级标准

Tab. 2 Combined model water quality grading standards

水质等级	I	II	III	IV	V
组合内梅罗指数	$\leq 0.338\ 4$	$\leq 0.451\ 4$	$\leq 0.877\ 0$	$\leq 2.971\ 6$	$> 2.971\ 6$

4 结果与分析

计算得出各水质样本的组合内梅罗指数,结合表 2 分级标准,得到各样本水质等级。同时采用单因子评价法、内梅罗指数法、改进的内梅罗指数法 I 型(仅改变平均污染指数)^[8]、II 型(仅改变

最大污染指数)^[7]、III 型(改变平均污染指数与最大污染指数)^[10]、投影寻踪与模糊物元组合模型(简称 PP-FME)^[4],对水质进行综合评价,并将评价结果与本文组合模型评价结果进行比较,各方法评价结果见表 3。将改进内梅罗指数法分别简称为改进 I 型、改进 II 型、改进 III 型。

表 3 各种方法水质评价结果

Tab. 3 Water quality evaluation results of various methods

地点	单因子评价法	内梅罗指数法	改进 I 型	改进 II 型	改进 III 型	PP-FME	投影寻踪与内梅罗指数组合模型 组合内梅罗指数	等级
乌玉深井	II	I	I	I	I	I	0.142 0	I
荫营煤矿水源地	V	IV	IV	IV	IV	IV	1.035 2	IV
韩庄	V	IV	IV	IV	IV	III	0.932 3	IV
苇钢	IV	IV	IV	III	III	II	0.473 1	III
柴家庄	V	IV	IV	III	III	III	0.738 2	III
大石门深井	IV	IV	IV	III	III	II	0.477 4	III
俱乐部	V	IV	IV	III	III	III	0.755 3	III
西峪深井	III	I	I	I	I	I	0.277 1	I
河下	III	I	I	I	I	I	0.307 5	I
小河	III	I	I	II	II	II	0.401 9	II
王龙	II	I	I	I	I	I	0.210 7	I
龙庄深井	IV	III	III	III	III	III	0.496 9	III
南坪	V	IV	IV	IV	IV	IV	0.979 6	IV
城西泉	III	II	II	II	II	II	0.403 1	II
坡底泉	III	I	I	II	II	II	0.394 9	II
五龙泉	III	I	I	I	I	II	0.357 4	II
苇泽关泉	III	II	II	II	II	II	0.381 6	II
集泉站	III	I	I	II	II	II	0.373 8	II
五矿贵石沟	III	I	I	I	I	I	0.343 9	II
李家庄	I	I	I	I	I	I	0.032 7	I

由表 3 可知,单因子评价法与组合模型得到的评价结果不一致。单因子评价法采用最差指标级别作为评价结果,仅考虑最差指标,不考虑其他指标,评价结果片面、不准确;而本文组合模型首先对所有评价指标以不同的权重对待,并以加权污染指数体现,同时突出最大污染指数、权重最大污染指数及权重与污染指数乘积最大(以下简称权积最大)污染指数,得到的结果准确、安全。

苇钢、柴家庄、大石门深井、俱乐部用内梅罗指数法得到的评价结果 IV 级差于组合模型的 III 级,这是因为内梅罗指数法会在结果中强调样本中的 IV、V 级指标,而组合模型引入了指标权重,将指标权重与污染指数结合得到加权污染指数,并在最大污染指数的基础上加入权重最大污染指数和权积最大污染指数,这样既不会太注重最大污染指数,又强调权重最大污染指数及权积最大污染指数,得到的评价结果更准确;而对于小河、坡底泉、五龙泉、集泉站以及五矿贵石沟,内梅罗指数法的评价等级 I 级优于组合模型的 II 级,但其内梅罗指数均接近 I 级水质上限,有向 II 级转化的趋势。产生上述差异的原因是,组合模型考虑了指标权重,并加入权重最大污染指数与权积

最大污染指数来修正最大污染指数,在评价结果中对指标权重及三项污染指数均有体现。

改进Ⅰ型与内梅罗指数法评价结果一致,与组合模型的评价结果相差较大,说明仅改进平均污染指数对评价结果影响不大;改进Ⅱ型中五龙泉及五矿贵石沟的评价结果与组合模型不一致,这是由于仅改变最大污染指数,评价结果仅强调改进的最大污染指数的信息,即最大污染指数与权重最大污染指数,但仍存在未赋权的问题;而对于改进Ⅲ型,五龙泉及五矿贵石沟的评价结果Ⅰ级均好于组合模型的Ⅱ级,从组合内梅罗指数来看,这两处Ⅱ级水质样本均接近Ⅰ级水质上限,而改进Ⅲ型的改进内梅罗指数均接近Ⅰ级水质上限,有向Ⅱ级转化的趋势,说明本文在改进最大污染指数时加入权重最大项不仅提高了准确性而且保证了水质安全。

本文组合模型的水质评价结果中韩庄、苇钢、大石门深井及五矿贵石沟的评价结果与 PP-FME 模型的评价结果不一致;这四组水质样本的共同的规律是,用组合模型评价得到的水质等级均比用 PP-FME 模型差一个等级,而 PP-FME 模型评价过程中水质样本得到的欧氏贴近度均接近各自所在水质级别上限,有向更差一级转化的趋势,说明组合模型评价结果的准确性,同时突显出在安全性方面的优点。

本文组合模型的某些水质样本评价结果较其他方法差一个等级;但用其他方法评价时这些水质样本均接近其所在水质等级的上限,有向更差一级转化的趋势。这说明本文评价模型对水污染具有一定的预测性,可通过评价结果及早发现水质问题并采取防治措施,避免水质进一步恶化。

5 结论

a. 组合模型通过 RAGA 优化投影寻踪法求

Groundwater Quality Evaluation Based on Combined Model of Projection Pursuit and Nemero Index

LI Qi, ZHAO Zhi-huai

(College of Hydro Science and Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: In view of the Nemero index method focusing on the maximum pollution index and not fully considering the index weight, this paper establishes the projection index function through the projection pursuit model, and uses the real code accelerated genetic algorithm to optimize the projection index function to obtain the index weight. Combined with the pollution index, the weighted pollution index is obtained. The maximum pollution index is added to the maximum value of the product of the weighted index and the pollution index to form a combined model of projection pursuit and Nemero index. Combined with the groundwater quality monitoring samples of the Niangzi Guanquan area in Yangquan City, the evaluation results were obtained. Compared with the current water quality assessment methods, the results show that the groundwater quality assessment by the combined model has some certain advantages in terms of safety and accuracy. The feasibility and applicability of the combined model are verified. The research results can provide reference for the prevention of groundwater pollution.

Key words: Nemero index method; projection pursuit; real code accelerated genetic algorithm; groundwater quality; evaluation

解指标权重,更加客观地为指标赋权;将平均污染指数改为加权污染指数,加入权重最大项 $\max \omega(j)F(i,j)$ 改进最大污染指数,对内梅罗指数法不赋权及其各改进型考虑因素不全面等问题进行了补充,提高了评价结果的准确性,同时可保证水质安全。

b. 将组合模型应用于工程实例,通过比较分析各种水质评价方法的评价结果,验证了组合模型的可行性与适用性。

参考文献:

- [1] 高卫东. 基于主成分分析的矿区地下水水质评价[J]. 安全与环境工程, 2009, 16(1): 28-30, 33.
- [2] 马玉杰, 郑西来, 李永霞, 等. 地下水质量模糊综合评价法的改进与应用[J]. 中国矿业大学学报, 2009, 38(5): 745-750.
- [3] 武强, 戴国锋, 吕华, 等. 基于 ANN 与 GIS 耦合技术的地下水污染敏感性评价[J]. 中国矿业大学学报, 2006, 35(4): 431-436.
- [4] 巩奕成, 张永祥, 任仲宇, 等. 基于投影寻踪和模糊物元组合模型的地下水水质评价[J]. 北京工业大学学报, 2014, 40(9): 1 402-1 407.
- [5] 巩奕成, 张永祥, 丁飞, 等. 基于萤火虫算法的投影寻踪地下水水质评价方法[J]. 中国矿业大学学报, 2015, 44(3): 566-572.
- [6] 白雁翎, 崔晨风, 董山琳. 基于改进萤火虫算法的投影寻踪太湖水质综合评价[J]. 节水灌溉, 2017(5): 72-76, 79.
- [7] 吴喜军, 董颖, 张亚宁. 改进的内梅罗污染指数法在黄河干流水质评价中的应用[J]. 节水灌溉, 2018(10): 51-53, 58.
- [8] 李小丽, 黎小东, 敖天其. 改进内梅罗指数法在西充河水质评价中的应用[J]. 人民黄河, 2016, 38(8): 65-68.
- [9] 谭永明, 孙秀玲. 基于加速遗传算法与投影寻踪的水质评价模型[J]. 水电能源科学, 2008, 26(6): 42-44.
- [10] 韩术鑫, 王利红, 赵长盛. 内梅罗指数法在环境质量评价中的适用性与修正原则[J]. 农业环境科学学报, 2017, 36(10): 2 153-2 160.