

基于线性回归的环境质量评价普适指数公式

薛文博, 易爱华, 张增强

(西北农林科技大学资源环境学院, 陕西 杨凌 712100)

摘要: 阐述了利用遗传算法优化污染损害指数评价模型中参数时的优点和不足, 提出采用普适质数公式对污染损害指数评价模型中参数进行线性优化时, 可以大幅简化优化过程, 便于实际应用。将适用于大气质量评价的普适指数公式与通过 GA 优化的评价结果进行对比, 证实了该公式在线性优化中的可行性。

关键词: 环境质量评价; 普适指数公式; 线性回归; 参数优化

中图分类号: X820 **文献标识码:** C **文章编号:** 1006-2009(2006)06-0041-03

Universal Index Formula of Environmental Quality Assessment Based on Linear Regression

XUE Wen-bo, YIA i-hua, ZHANG Zeng-qiang

(College of Resources and Environment, Northwest A&F University, Yangling, Shanxi 712100, China)

Abstract: The advantage and shortcoming of using genetic algorithm (GA) to optimize parameter in pollution damage index assessment model was discussed. The paper present a linear regression optimized method for parameters of pollution damage index assessment model, which simplified the optimization process sharply and made practical application more convenient. After using this method to atmospheric quality assessment, the paper got a universal index formula of atmospheric environmental quality assessment, proved the feasibility of linear optimization eventually through compared assessment results of it with the GA optimization.

Key words: Environmental quality assessment; Universal index formula; Linear regression; Parameter optimization

有关环境质量评价的方法很多,不同的评价方法都各有利弊^[1-4],其中污染损害率评价法与环境质量和污染物对环境损害的程度相关联,并体现了污染物危害的程度与浓度之间的非线性关系,使评价结果具有明确的物理意义。但是,由于不同污染物的特性不同,就会造成不同污染物同级标准数值的差异很大,故不同污染物的损害率公式中的参数值各不相同,不便使用^[1]。

在设置各污染物“基准”浓度值的情况下,若用每种污染物实测浓度的相对值代替污染损害率公式中的实测浓度值,则公式中的参数与污染物的特性无关。通过数学方法对参数优化,就可以得到对多种污染物都能普遍适用的环境污染损害率计算公式^[1]。应用该方法评价的关键就是普适公式的参数优化,而目前大多采用的遗传算法(GA)优

化方法比较复杂,不便实际应用。普适参数的线性回归优化方法,可以大幅简化优化过程,便于实际应用。

1 计算环境污染损害率的普适公式

每种污染物对环境质量危害的程度都不是简单的线性关系,而是呈 S 型曲线变化。因此,污染物对环境质量损害的程度可以用污染损害率 R_j 和污染损害指数 I_j 表示:

$$R_j = 1 / (1 + a_j e^{-b_j x_j}) \quad (1)$$

$$I_j = 10 / (1 + a_j e^{-b_j x_j}) \quad (2)$$

式中: a_j 、 b_j ——与污染物特性有关的待定参数;

收稿日期: 2005-11-08; 修订日期: 2006-08-10

作者简介: 薛文博(1981—),男,陕西眉县人,硕士研究生,从事固体废弃物资源化与利用研究。

x_j —— j 污染物的相对浓度。

x_j 可用下式计算：

$$x_j = c_j / c_0 \tag{3}$$

式中 c_0 为设定的 j 污染物的“基准”浓度值 (或本底值), 而 c_0 的设定目的是使不同污染物相对浓度值 x_j 的差异不是很大, 最好处于同一数量级。从而可以认为, 采用相对浓度值后污染损害率公式中的参数 a_j 、 b_j 与污染物特性无关, 对所有污染物都可取相同的待定参数 a 、 b 。因而可定义适用于多种污染物的污染损害率计算公式为：

$$R_j = 1 / (1 + ae^{-bx_j}) \tag{4}$$

同时定义相同的污染损害指数 I_j 为：

$$I_j = 10 / (1 + ae^{-bx_j}) \tag{5}$$

该文是以大气质量评价为例, 显然, R_j 满足 $0 < R_j < 1$ 。当 $x_j = c_0 / c_0 = 1$ 时, $R_0 = 0.01$; 当 $x_j = c_p / c_0$ 时, $R_p = 0.99$ 。将 R_j [0.01, 0.99] 按“等差分级, 等比赋值”分级原则, 划分为 $k = 0, 1, 2, \dots, 9$ 共 10 个级别, 其中任意相邻两级之间 R_j 的比值为: $= (R_p / R_0)^{1/9} = 1.6662$, 第 k 级的 $R_k = R_0 (R_p / R_0)^{k/9}$, 而 $k = 0, 3, 5, 7, 9$ 时, 分别与大气“基准级、级、级、级、级 (明显危害级) 相对应。与 5 个级别相对应的污染损害率的目标值 R_{ke} 分别为: 0.01、0.0463、0.1284、0.3562、0.99。

2 普适参数的线性优化

通过线性回归优化 (4) 式中的参数 a 、 b , 对

(4) 式取倒、移项、取对数可得 (6) 式：

$$\ln\left(\frac{1}{R_j} - 1\right) = \ln a - bx_j \tag{6}$$

通过最小二乘法估计参数 a 、 b 。根据最小二乘法的定义^[5], 构造目标函数：

$$f_{\min}(a, b) = \sum_{k=0}^m \sum_{j=1}^n \left[\ln\left(\frac{1}{R_j} - 1\right) - \ln\left(\frac{1}{R_{ke}} - 1\right) \right]^2 \tag{7}$$

式中: m ——污染物种类数目；

$K = 5$ ——大气分级数目 (包括“基准级和明显危害浓度级)；

R_j ——通过式 (4) 计算出的 j 污染物的 k 级标准的污染损害率；

R_{ke} —— k 级标准的污染损害率目标值。

当 $f(a, b)$ 取得最小值时, 所对应的 a 、 b 值为需要的最佳优化参数。也可以对 (6) 式作图, 当所得直线的截距为 $\ln a$, 斜率为 $-b$ 时, 可以获得 a 、 b 的最佳优化结果。

3 通过线性优化确定大气质量评价普适指数公式

3.1 确定普适参数

几种主要大气污染物的国家标准 GB 3095-1996《环境空气质量标准》浓度限值 c_{jl} ($l = \dots$,)、本底浓度值 (“基准值”) c_0 、明显危害浓度限值 c_{kl} 及由公式 (3) 计算出的各级浓度相对值 x_{jl} ($l = 0, 1, 2, 3, 4$) 见表 1。

表 1 空气质量标准及各级浓度相对值

污染物	0		1		2		3		4	
	c_0	x_0	c_1	x_1	c_2	x_2	c_3	x_3	c_4	x_4
SO ₂	0.02	1	0.05	2.50	0.15	7.50	0.25	12.50	0.50	25.0
NO _x	0.015	1	0.05	3.33	0.10	6.67	0.15	10.0	0.30	20.0
NO ₂	0.015	1	0.04	2.67	0.08	5.33	0.12	8.00	0.25	16.7
PM ₁₀	0.02	1	0.05	2.50	0.10	5.00	0.25	12.50	0.50	25.0
TSP	0.05	1	0.12	2.40	0.30	6.00	0.50	10.00	1.20	24.0
CO	0.5	1	1.50	3.00	4.00	8.00	6.00	12.00	10.00	20.0
降尘	2.0	1	5.0	2.75	12.00	6.00	24.00	12.00	40.00	20.0

0、1、2、3、4 代表级别; 降尘单位为 $\text{km}^2/\text{月}$, 其余为 mg/m^3 。

通过式 (6)、式 (7) 对普适参数线性优化, 结果为 $a = 75.256$, $b = 0.390$, 由此得到适用于表 1 的 7 种大气污染物的大气质量评价污染损害普适公式：

$$R_j = 1 / (1 + 75.256e^{-0.390x_j}) \tag{8}$$

$$I_j = 10 / (1 + 75.256e^{-0.390x_j}) \tag{9}$$

通过 GA 对普适参数的优化结果: $a = 38.0$, $b = 0.265$, 对应普适公式为^[11]：

$$R_j^* = 1 / (1 + 38.0e^{-0.265x_j}) \tag{10}$$

$$I_j^* = 10 / (1 + 38.0e^{-0.265x_j}) \tag{11}$$

通过 GA 优化参数 a, b 是全局性的搜索, 优化结果精确^[6,7], 而通过线性回归优化参数 a, b 得到的结果只是对线性化公式 (6) 的最佳优化, 并不是原污染损害率公式 (4) 的最佳优化, 故优化精度不如 GA 高。虽然两种方法所得参数 a, b 值的差异较大, 但是, 通过线性回归得到的普适公式同样达到了极显著水平 ($R^2 = 0.998$), 完全可以适用于评价过程, 并且该方法简单可行。

3.2 危害指数的取值范围与评价级别之间的对应关系

由式 (9)、式 (11) 分别求出表 1 中 7 种污染物各级标准的污染损害指数平均值 I_k 和 I_k^* , 得到大气环境质量评价级别、级与污染损害指数 I_k 和 I_k^* 之间的对应关系, 见表 2。

表 2 大气质量标准与 I_k 和 I_k^* 之间的对应关系

大气级别 (k)	()	()	()
I_k 的范围	0.37	1.44	3.89
I_k^* 的范围	0.52	1.25	3.35

4 通过广义模糊对比加权确定综合污染损害指数

受 m 种污染物污染的大气污染损害综合指数计算公式为:

$$I = \sum_{j=1}^m W_j I_j = \sum_{j=1}^m W_j 10 / (1 + 75.256 e^{-0.390 x_j}) \quad (12)$$

式中 W_j 与损害分指数 I_j 的关系并不是简单的线性关系^[5], 而是呈现 S 型的变化规律。因此, 确定权重的时候必须做适当调整, 当分指数为 0.5 的因子权重不改变, 分指数 > 0.5 , 特别是接近 1 时, 应该适当削弱; 而分指数 < 0.5 , 特别是接近 0 时, 应该适当加强。而满足上述原则的权重值与分指数的关系应该用广义对比形式:

$$W_j = \begin{cases} (\frac{\mu_j}{2})^{1/2} & 0 < \mu_j < 0.5 \\ 1 - (\frac{1 - \mu_j}{2})^{1/2} & 0.5 < \mu_j < 1 \end{cases} \quad (13)$$

式中 W_j 代表了 j 污染物的权重大小, 将 W_j 归一化后, 即可得到 j 污染物的广义模糊对比权值 W_{j0} 。其中:

$$\mu_j = C_j / C_{j\mu} \quad (14)$$

式中: C_j —— j 污染物的实际监测值;

$C_{j\mu}$ —— j 污染物的明显危害浓度限值 (一般而言 $\mu_j < 1$)。

5 实例

某地区 5 个监测点的 SO_2 、 NO_x 和 TSP 3 项污染物的浓度监测值通过公式 (3)、(12)、(13)、(14) 计算, 得出各监测点污染物的权重值 W_j 和大气污染损害综合指数 I , 以及采用基于线性回归 (LG) 和 GA 法的评价结果见表 3。

表 3 各监测点中 SO_2 、 NO_x 和 TSP 检测及评价结果

监测点	(SO_2) / W_j	(NO_x) / W_j	(TSP) / W_j	I	LG	GA
1	0.055/0.30	0.036/0.31	0.239/0.39	0.53		
2	0.022/0.26	0.013/0.26	0.188/0.48	0.34		
3	0.097/0.33	0.027/0.23	0.417/0.44	1.45		
4	0.158/0.36	0.036/0.22	0.512/0.42	2.64		
5	0.027/0.19	0.035/0.29	0.457/0.52	1.79		

SO_2 、 NO_x 、TSP 单位为 mg/m^3 。

表 3 表明, 基于线性回归的普适公式评价结果与基于 GA 优化的评价结果完全一致, 通过线性优化可以大幅简化普适参数的优化过程, 可以方便地应用于实际评价过程中。

6 结语

通过线性回归方法对普适指数公式中的参数优化作了简化, 克服了 GA 优化过程的缺点, 实例表明了该优化方法的可靠性和正确性, 为污染损害指数法在环境质量评价中的应用提供了便利。

[参考文献]

- [1] 李祚勇, 丁晶, 彭荔红. 环境质量评价原理与方法 [M]. 北京: 化学工业出版社, 2004.
- [2] 刘硕. 对几种环境质量综合指数评价方法的探讨 [J]. 中国环境监测, 1999, 15 (5): 33 - 37.
- [3] 曾现来, 刘晓红, 张增强. 应用欧氏距离聚类法综合评价环境质量 [J]. 中国给水排水, 2003, 19 (12): 99 - 100.
- [4] 杨土建, 赵秀兰, 张润玲, 等. 环境质量评价中关键污染因子的确定方法 [J]. 环境监测管理与技术, 2002, 14 (2): 20 - 23.
- [5] 袁志发, 周静芋. 多元统计分析 [M]. 北京: 科学出版社, 2002.
- [6] 曾现来, 张增强, 刘晓红. 城市生活垃圾中各成分权重模型的建立及验证 [J]. 农业环境科学学报, 2004, 23 (4): 774 - 776.
- [7] HOLLAD J H. Genetic algorithms [J]. Scientific American, 1992 (4): 44 - 55.