Journal of Water Resources Research 水资源研究, 2019, 8(3), 217-223
Published Online June 2019 in Hans. http://www.hanspub.org/journal/jwrr
https://doi.org/10.12677/jwrr.2019.83026

Hans汉斯

# Improvement of Geo-Accumulation Index Based on Maximum Entropy Principle

**Mingbo Zhang[1], Feng Chen[1], Jincheng Wang[1], Bao Qian[1], Xiao Xiao[1,2]**

[1]Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan Hubei
[2]Engineering Research Center of Eco-environment in Three Gorges Reservoir Region, Ministry of Education, Three Gorges University, Yichang Hubei
Email: zhangmb@cjh.com.cn

Received: May 6th, 2019; accepted: May 22nd, 2019; published: May 30th, 2019

## Abstract

**For the sampling error, the measuring error, the inhomogeneity of minerals, and the other uncertain factors, in the practical environmental investigation, the concentrations of heavy metals in the sediments are often represented informs of intervals instead of accurate values. However, the conventional geo-accumulation index (CGI) cannot deal with these uncertainties in heavy metal pollution evaluation. To solve this problem, an improved geo-accumulation index (IGI) model is established based on the maximum entropy principle. The heavy metal pollution in the sediment in the Dongting Lake is evaluated as an illustration to compare the effects of these two models. The result shows that IGI has a better capacity in dealing with the heavy metals, the concentrations of which are represented informs of intervals, and an obvious advantage in the hierarchical recognition and the pollution degree schedule. At last, the mathematical relationship between these two models is revealed. IGI is the generalization of CGI into uncertainty analysis, while CGI can be regarded as a special case when the width of the uncertainty interval in IGI approximates to zero.**

## Keywords

**Geo-Accumulation Index, Lake Sediments, Heavy Metal Pollution, Maximum Entropy Principle**

# 基于极大熵原理的改进型地累积指数研究

**张明波[1]，陈　峰[1]，汪金成[1]，钱　宝[1]，肖　潇[1,2]**

[1]长江水利委员会水文局，湖北 武汉
[2]三峡大学，三峡库区生态环境教育部工程研究中心，湖北 宜昌
Email: zhangmb@cjh.com.cn

收稿日期：2019年5月6日；录用日期：2019年5月22日；发布日期：2019年5月30日

作者简介：张明波(1966-)，男，教授级高级工程师，主要从事流域水资源管理与保护工作。

文章引用：张明波，陈峰，汪金成，钱宝，肖潇．基于极大熵原理的改进型地累积指数研究[J]．水资源研究，2019，8(3): 217-223. DOI: 10.12677/jwrr.2019.83026

*特邀论文

## 摘　要

**由于采样误差、测量误差以及沉积物时空分布的不均匀性等因素的影响，在实际的环境调查中，沉积物的重金属浓度往往具有一定的不确定性，而传统的地累积指数难以处理这类问题。为解决这一问题，本研究首先基于极大熵原理，建立了改进的地累积指数模型。而后以洞庭湖的沉积物重金属污染评价为例，对两种模型的评价效果进行了比较。结果显示改进的地累积指数模型能够更好地处理实测值以区间形式表示的重金属污染评价问题，而且相对于传统的地累积指数模型，改进的地累积指数模型在等级识别和污染程度排序方面均更有优势。最后，本研究通过数学证明，揭示了两种模型的联系与区别：改进的地累积指数是传统地累积指数向不确定性分析的拓展和深化，而传统地累积指数本质上是改进的地累积指数在不确定区间宽度趋于0时的特例。**

## 关键词

**地累积指数，湖泊沉积物，重金属污染，极大熵原理**



## 1. 引言

地累积指数是由 Muller 等学者提出的一种沉积物重金属污染评价模型[1]。它综合考虑了人为污染，环境地球化学背景值，以及自然成岩作用引起的背景值变动等因素的共同影响，通过构建简单而直观的污染指数，定量地反映沉积物中的重金属污染情况[2]。由于地累积指数能够区分自然环境和人类活动对沉积物重金属的影响，而且在参数选取中具有很高的灵活性，因此地累积指数模型近年来在国内外的河湖沉积物重金属污染评价中均得到了广泛的应用[1] [2] [3]。

然而，传统的地累积指数模型没有考虑不确定性因素的影响，在评价过程中，沉积物的重金属含量需要是精确的数值。但是在实际的沉积物环境调查中，采样误差、测量误差以及沉积物时空分布的不均匀性等不确定性因素往往是不可避免的，因此我们通常难以确定重金属浓度的精确值，而只能确定其可能的分布区间[4] [5]。为解决这一问题，本研究将基于概率论和极大熵原理，对传统的地累积指数进行改进，构建适用于区间型重金属浓度的地累积指数模型。

## 2. 评价模型

### 2.1. 传统地累积指数模型

传统地累积指数的表达式为[1] [2] [3]：

$$I_n = \log_2 \frac{x_n}{1.5 \cdot b_n}. \tag{1}$$

式中：为 $x_n$，$b_n$，$I_n$ 分别为第 $n$ 种重金属的现状浓度、背景值浓度和地累积指数，1.5 为考虑成岩作用引起背景值变动而引入的修正系数[1] [2] [3]。

根据 $I_n$ 的数值，重金属的污染状况被划分为 7 个等级：清洁($I_n < 0$)，轻度污染($0 \le I_n < 1$)，偏中度污染($1 \le I_n < 2$)，中度污染($2 \le I_n < 3$)，偏重度污染($3 \le I_n < 4$)，重度污染($4 \le I_n < 5$)，严重污染($I_n \ge 5$) [1] [2] [3]。

## 2.2. 改进的地累积指数模型

由于采样误差、测量误差以及沉积物时空分布的不均匀性等不确定性因素的影响，在实际的沉积物环境调查中，我们往往难以确定重金属浓度 $x_n$ 的精确数值，而只能确定一个可能的分布区间 $a_n \le x_n \le c_n$ [4] [5]。然而传统的地累积指数难以处理这类表达形式为分布区间的实测值，为了解决这一问题，我们基于概率论和极大熵原理对传统的地累积指数进行改进。

根据概率论原理，可以将第 $n$ 种重金属的现状浓度 $x_n$ 视为一个连续型随机变量 $X_n$。设 $X_n$ 的概率分布函数为 $F(x_n)$，对应的概率密度函数为 $f(x_n)$，那么此时地累积指数 $I_n^*$ 可以定义为：

$$I_n^* = \int_{x_n=a_n}^{c_n} \left( \log_2 \frac{x_n}{1.5 \cdot b_n} \right) \cdot f\left(x_n\right) \mathrm{d}x_n. \tag{2}$$

可以发现，求解式(2)的关键在于确定概率密度函数为 $f\left(x_n\right)$ 的解析式，本研究将基于极大熵原理解决这一问题。

在概率论与数理统计中，熵是随机变量不确定性的度量[6] [7]。连续型随机变量 $X_n$ 的熵值 $H_n$ 定义为：

$$H_n = -\int_{-\infty}^{+\infty} f\left(x_n\right) \cdot \ln f\left(x_n\right) \mathrm{d}x_n. \tag{3}$$

熵值 $H_n$ 越大，随机变量 $X_n$ 的不确定性越高。

根据热力学第二定律，孤立系统的熵值总是保持增大或者不变的趋势[6] [7]。基于该准则，Jaynes 于 1957 年提出了极大熵原理：在随机变量 $X_n$ 所有可能的分布函数 $f\left(x_n\right)$ 中，应该选择在一定约束下使得熵 $H_n$ 最大的分布作为其分布函数[6] [7]。

根据极大熵原理，$X_n$ 最可能的概率密度函数 $f\left(x_n\right)$ 可以通过求解以下最优化问题计算：

$$\begin{aligned} &\max : H_n = -\int_{-\infty}^{+\infty} f\left(x_n\right) \cdot \ln f\left(x_n\right) \mathrm{d}x_n \\ &\text{s.t.} \quad \int_{x_n=a_n}^{c_n} f\left(x_n\right) \mathrm{d}x_n = 1 \end{aligned} \tag{4}$$

根据泛函分析中的极值求解理论，可以构建如下目标函数 $L(f(x_n))$ [7]：

$$L\left(f\left(x_n\right)\right) = -f\left(x_n\right) \cdot \ln f\left(x_n\right) + \lambda \cdot f\left(x_n\right). \tag{5}$$

其中为 $\lambda$ 拉格朗日乘子。

根据拉格朗日对偶原理，最优解 $f\left(x_n\right)$ 满足 $\dfrac{\partial L\left(f\left(x_n\right)\right)}{\partial f\left(x_n\right)} = 0$ [7]，此时有：

$$f\left(x_n\right) = \mathrm{e}^{\lambda - 1}. \tag{6}$$

将式(6)代入约束条件 $\int_{x_n=a_n}^{c_n} f\left(x_n\right) \mathrm{d}x_n = 1$，可发现：

$$\lambda = \ln \frac{1}{c_n - a_n} + 1. \tag{7}$$

将式(7)代入式(6)可知，最优解 $f\left(x_n\right)$ 为：

$$f\left(x_n\right) = \frac{1}{c_n - a_n}. \tag{8}$$

将式(8)代入式(2)，可知当重金属浓度 $x_n$ 可能的分布区间为 $a_n \le x_n \le c_n$ 时，改进的地累积指数 $I_n^*$ 计算式为：

$$\begin{aligned} I_n^* &= \int_{x_n=a_n}^{c_n} \left(\log_2 \frac{x_n}{1.5 \cdot b_n}\right) \cdot \frac{1}{c_n - a_n} \mathrm{d}x_n \\ &= \int_{x_n=a_n}^{c_n} \left(\log_2 x_n - \log_2 1.5 - \log_2 b_n\right) \cdot \frac{1}{c_n - a_n} \mathrm{d}x_n \\ &= \int_{x_n=a_n}^{c_n} \left(\log_2 x_n\right) \cdot \frac{1}{c_n - a_n} \mathrm{d}x_n - \left(\log_2 1.5 + \log_2 b_n\right) \\ &= \frac{\left(c_n \cdot \ln c_n - c_n\right) - \left(a_n \cdot \ln a_n - a_n\right)}{\ln 2 \cdot \left(c_n - a_n\right)} - \left(\log_2 1.5 + \log_2 b_n\right) \end{aligned} \tag{9}$$

### 2.3. 两种地累积指数的数学关系

为更好地揭示两种地累积指数的数学关系，我们进一步结合数学证明进行论证。首先将不确定区间[$a_n$,$c_n$]的区间宽度定义为 $w_n$，即 $w_n = c_n - a_n$，那么此时改进的地累积指数 $I_n^*$ 计算式可以写成：

$$I_n^* = \frac{\left(a_n + w_n\right) \cdot \ln\left(a_n + w_n\right) - a_n \cdot \ln a_n - w_n}{\ln 2 \cdot w_n} - \left(\log_2 1.5 + \log_2 b_n\right). \tag{10}$$

当 $w_n$ 趋近于 0 时，对 $I_n^*$ 求极限：

$$\begin{aligned} \lim_{w_n \to 0} I_n^* &= \lim_{w_n \to 0} \left\{ \frac{\left(a_n + w_n\right) \cdot \ln\left(a_n + w_n\right) - a_n \cdot \ln a_n - w_n}{\ln 2 \cdot w_n} - \left(\log_2 1.5 + \log_2 b_n\right) \right\} \\ &= \left\{ \lim_{w_n \to 0} \frac{\left(a_n + w_n\right) \cdot \ln\left(a_n + w_n\right) - a_n \cdot \ln a_n - w_n}{\ln 2 \cdot w_n} \right\} - \left(\log_2 1.5 + \log_2 b_n\right) \end{aligned} \tag{11}$$

根据洛必达法则[8]有：

$$\lim_{w_n \to 0} \frac{\left(a_n + w_n\right) \cdot \ln\left(a_n + w_n\right) - a_n \cdot \ln a_n - w_n}{\ln 2 \cdot w_n} = \frac{\ln a_n}{\ln 2} = \log_2 a_n. \tag{12}$$

将式(12)代入式(11)：

$$\lim_{w_n \to 0} I_n^* = \log_2 a_n - \left(\log_2 1.5 + \log_2 b_n\right) = \log_2 \frac{a_n}{1.5 \cdot b_n}. \tag{13}$$

由于 $a_n \le x_n \le c_n$，因此当 $w_n$ 趋近于 0 时有：

$$\lim_{w_n \to 0} x_n = a_n. \tag{14}$$

结合式(1)、式(13)、式(14)可知：

$$\lim_{w_n \to 0} I_n^* = I_n. \tag{15}$$

由此可见，改进的地累积指数是传统地累积指数向不确定性分析的拓展和深化，而传统地累积指数实质上是改进的地累积指数在不确定区间宽度趋于 0 时的特例。

## 3. 结果与讨论

### 3.1. 算例

洞庭湖(28˚44'~29˚35'N, 111˚53'~113˚05'E)作为长江“双肾”之一的，接纳了上游金属矿产开采导致的大量重

金属污染物，加之每年汇集来自周边城市的工农业废水与生活污水，湖区重金属污染问题日趋严重，作为汇流终点的东洞庭湖环境问题更为严峻。本研究以洞庭湖的城陵矶站、鹿角站和樟树港站的沉积物镉、砷重金属评价为例(采样点见图 1)，对两种地累积指数的评价结果进行比较。具体数值见表 1 所示，其中各重金属的实测值与本底值均引自文献[4]。

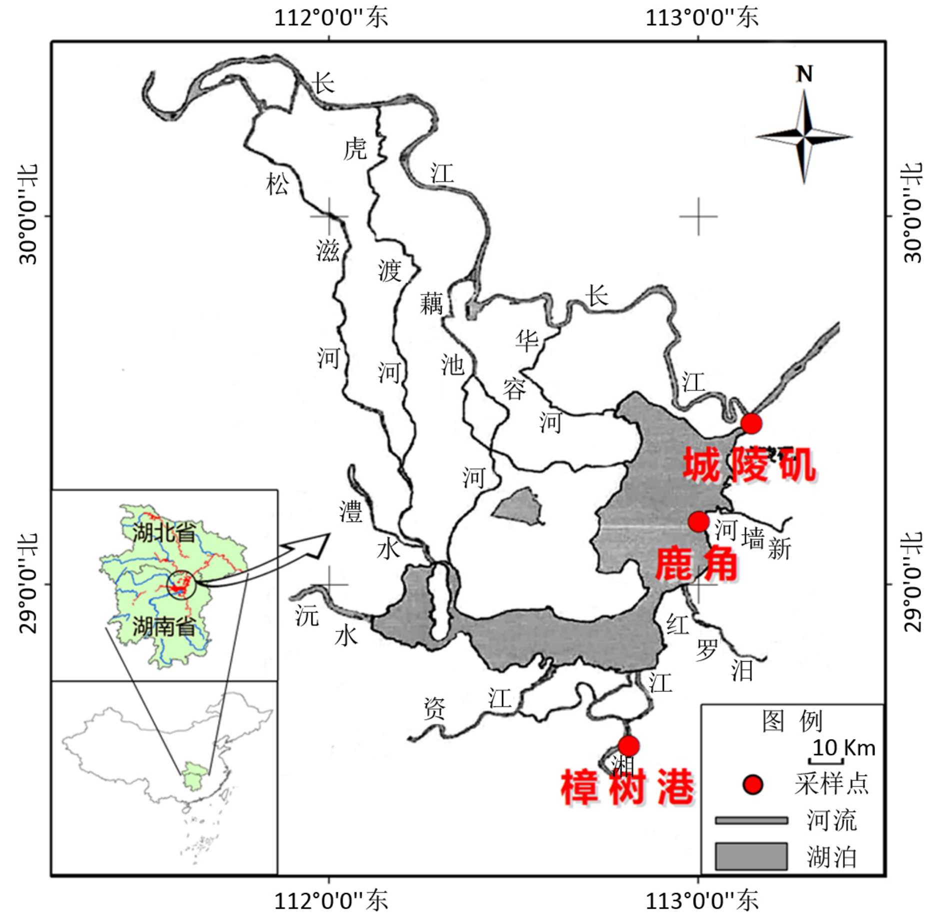


**Figure 1.** Distribution of sampling sites in Dongting Lake water system
**图 1.** 洞庭湖水系采样点布设示意图

**Table 1.** The measured value and background value of heavy metals
**表 1.** 各重金属的实测值与本底值(mg/kg)

| 采样点 | 镉 | 砷 |
|---|---|---|
| 城陵矶 | [0.248,6.35] | [7.44,57.1] |
| 鹿角 | [0.775,1.63] | [20.2,22.3] |
| 樟树港 | [5.14,79.9] | [21.3,289] |
| 本底值 | 0.23 | 13.41 |

通过表 1 可以发现，由于采样误差、测量误差以及沉积物时空分布的不均匀性等不确定性因素的影响，各沉积物中重金属的浓度均以区间的形式表示，而无法确定其精确的数值。根据式(1)与式(2)，对各沉积物的重金属污染状况进行评价如表 2。

**Table 2.** The evaluation results of the two methods
**表 2.** 两种地累积指数评价结果比较

| 重金属 | 采样点 | 传统地累积指数 | | 改进的地累积指数 | |
|---|---|---|---|---|---|
| | | 指数 | 等级 | 指数 | 等级 |
| 镉 | 城陵矶 | 无法计算 | 无法确定 | 2.95 | 中度污染 |
| | 鹿角 | 无法计算 | 无法确定 | 1.77 | 偏中度污染 |
| | 樟树港 | 无法计算 | 无法确定 | 6.68 | 严重污染 |
| 砷 | 城陵矶 | 无法计算 | 无法确定 | 0.50 | 轻度污染 |
| | 鹿角 | 无法计算 | 无法确定 | 0.08 | 轻度污染 |
| | 樟树港 | 无法计算 | 无法确定 | 2.70 | 中度污染 |

结合表 2 可发现，根据改进的地累积指数，城陵矶站、鹿角站和樟树港站的沉积物镉污染状况分别为中度污染、偏中度污染和严重污染；而沉积物砷污染状况分别为轻度污染、轻度污染和中度污染。两种重金属的评价区域污染程度排序均为：樟树港 > 城陵矶 > 鹿角；而三个测站的评价指标污染程度序均为：镉 > 砷。

## 3.2. 讨论

由于城陵矶站、鹿角站和樟树港站的沉积物镉、砷浓度均不是具体的数值，因此根据传统的地累积指数模型，很难确定相应的污染指数、污染等级，也无法对城陵矶站、鹿角站和樟树港站的污染程度进行排序。

综上所述，当沉积物中重金属的浓度具有不确定性时，传统的地累积指数很难识别它们的污染等级，也难以对城陵矶站、鹿角站和樟树港站，而改进的地累积指数则有效地解决了这一问题。因此，相比于传统的地累积指数模型，改进的地累积指数模型在等级识别和污染程度排序方面均更有优势。

# 4. 结论

改进的地累积指数模型能够更好地处理实测值以区间形式表示的重金属污染评价问题，而且相对于传统的地累积指数模型，改进的地累积指数模型在等级识别和污染程度排序方面均更有优势。

改进的地累积指数是传统地累积指数向不确定性分析的拓展和深化，而传统地累积指数本质上是改进的地累积指数在不确定区间宽度趋于 0 时的特例。

# 基金项目

国家重点研发计划(2016YFA0600901)；湖北省自然科学基金项目(2017CFB312)；三峡库区生态环境教育部工程研究中心开放基金(KF2018-05)；中央高校基本科研业务费(2017B20514)。

# 参考文献

[1] 田海涛, 张振克, 丁海燕, 等. 40 年来江苏石梁河水库重金属污染的沉积记录[J]. 湖泊科学, 2008, 20(5): 600-604.
TIAN Haitao, ZHANG Zhenke, DING Haiyan, et al. Recent 40-year sedimentary record of heavy metal pollution in the Shilianghe Reservoir, Jiangsu Province. Journal of Lake Science, 2008, 20(5): 600-604. (in Chinese)

[2] 魏荣菲, 庄舜尧, 杨浩, 等. 苏州河网区河道沉积物重金属的污染特征[J]. 湖泊科学, 2010, 22(4): 527-537.
WEI Rongfei, ZHUANG Shunyao, YANG Hao, et al. Pollution characteristics of heavy metals in sediments from the river network of Suzhou City. Journal of Lake Science, 2010, 22(4): 527-537. (in Chinese)

[3] MEN, C., LIU, R., XU, F., et al. Pollution characteristics, risk assessment, and source apportionment of heavy metals in road dust in Beijing, China. Science of the Total Environment, 2018, 612: 138-147. https://doi.org/10.1016/j.scitotenv.2017.08.123

[4] 祝慧娜, 李莹, 梁婕, 等. 基于区间数排序法的洞庭湖沉积物重金属生态风险分析[J]. 环境工程, 2014, 32(2): 114-117.
ZHU Huina, LI Ying, LIANG Jie, et al. Ecological risk assessment of heavy metals in sediment of Dongting Lake based on

ranking-method of interval numbers. Environmental Engineering, 2014, 32(2): 114-117. (in Chinese)

[5] 王晓飞, 邓超冰, 尹娟, 等. 基于区间数排序法的农田土壤重金属生态风险分析[J]. 中国环境监测, 2017, 33(3): 106-113.
WANG Xiaofei, DENG Chaobing, YIN Juan, et al. Ecological risk assessment of heavy metals in the contaminated farmland based on ranking-method of interval numbers. Environmental Monitoring in China, 2017, 33(3): 106-113. (in Chinese)

[6] PAPALEXIOU, S. M., KOUTSOYIANNIS, D. Entropy based derivation of probability distributions: A case study to daily rainfall. Advances in Water Resources, 2012, 45(45): 51-57. https://doi.org/10.1016/j.advwatres.2011.11.007

[7] 邱菀华. 管理决策与应用熵学[M]. 北京: 机械工业出版社, 2002.
QIU Kouhua. Management decision and applied entropy. Beijing: Machinery Industry Press, 2002. (in Chinese)

[8] 代恩华. 洛必达法则及斯铎兹定理的一种简便证法[J]. 高等数学研究, 2010, 13(5): 51-54.
DAI Enhua. Easy approach to L Hospitals rule and O. Stolz theorem. Studies in College Mathematics, 2010, 13(5): 51-54. (in Chinese)