

# 基于遗传算法的大气颗粒物的源解析

李祚泳<sup>1</sup>, 彭荔红<sup>2</sup>

( 1. 成都气象学院, 四川 成都 610041; 2. 厦门大学 环境科学研究中心, 福建 厦门 361005)

**摘要:** 遗传算法是一种全局优化的随机搜索方法, 适用于模型参数寻优。将遗传算法用于大气颗粒物源解析的 CMB 方程组中参数的优化, 得到各污染源对大气颗粒物的优化贡献率。该方法应用于成都市大气颗粒物的源解析结果与用其他多种源解析法得出的结果基本一致。理论分析和应用实践表明了遗传算法用于大气颗粒物源解析研究具有简便、准确和实用性强等特点。

**关键词:** 遗传算法; 参数优化; 大气颗粒物; 源解析

中图分类号: X 169 文献标识码: A 文章编号: 1001-6929(2000)06-0019-03

## Source Apportionment of Atmospheric Particulates Based on Genetic Algorithm

LI Zuo-yong<sup>1</sup>, PENG Li-hong<sup>2</sup>

( 1. Chengdu Institute of Meteorology, Chengdu 610041, China;

2. Research Center of Environmental Science, Xiamen University, Xiamen 361005, China)

**Abstract:** Genetic algorithm (GA) is a random search method which is suitable for the optimum of model parameters because of its global optimum. Optimal contribution rates of atmospheric particulates of each pollution source were calculated by using GA to optimize the parameters in CMB equation sets of sources apportionment of atmospheric particulates. Results of sources apportionment of atmospheric particulates in Chengdu using GA tallied with that of sources apportionment of other methods. It is shown that GA is simple, precise and practical for sources apportionment of atmospheric particulates by the theoretical analysis and practical application.

**Key words:** genetic algorithm; parameter optimum; atmospheric particulate; sources apportionment

大气颗粒物的源解析常用的受体模型有多种, 它们各有利弊<sup>[1~3]</sup>。笔者曾采用投影寻踪技术 (PP) 对大气污染物进行源解析, 在一定程度上克服了源解析的非线性问题<sup>[4]</sup>。不过, 当研究对象过于复杂时, 如何寻找初始优化方向, 有时会遇到困难。由 Holland 提出的遗传算法 (Genetic Algorithms, 简称 GA) 是一种处理模型参数优化估计的通用方法<sup>[5]</sup>。笔者将遗传算法用于化学质量平衡法 (CMB) 中的排放源的成分谱分布寻优, 从而获得各污染源对大气颗粒物的最优贡献率。

### 1 遗传算法 (GA) 的基本思想和优化准则函数

GA 的基本思想是基于 Darwin 的进化论和 Mendel 的遗传学说。其特点是群体搜索策略和群体中个体之间的随机信息交换, 搜索不依赖于梯度信

息。该算法将对问题的求解转化为对一群“染色体”的一系列操作, 通过群体进化, 最后收敛到一个最适应环境的“染色体”上, 从而求得问题的最优解<sup>[6]</sup>。

设模型参数估计为  $F$  取最小的优化准则函数:

$$\text{Min } F = \text{Min } \sum_{i=1}^m \|f(g_1, g_2, \dots, g_j, \dots, g_p; x_i) - y_i\|^q \quad (1)$$

式中, 待优化参数  $g_j \in [a_j, b_j]$ ,  $j = 1, 2, \dots, p$ ;  $a_j, b_j$  为  $g_j$  的初始变化区间;  $F$  为优化准则函数;  $f$  一般为非线性模型;  $\{(x_i, y_i) | i = 1, 2, \dots, m\}$  为模型的  $m$  对输入、输出观测数据;  $\|\cdot\|$  为取范数。  $q$  为视实际要求而定的实常数, 如范数的定义为欧氏距离, 则  $q$  可取 2, 也可取 1。

### 2 GA 用于大气颗粒物的源解析

#### 2.1 构造源解析的 GA 算法的优化准则函数

在 CMB 源解析法中, 大气中元素  $i$  的监测浓度  $c_i$  ( $\mu\text{g}/\text{m}^3$ ) 与排放源  $j$  排放的颗粒物中元素  $i$  的含量  $x_{ij}$  ( $\mu\text{g}/\text{mg}$ ) 之间满足如下关系:

收稿日期: 1999-11-09

基金项目: 国家“九五”重点科技攻关项目 (96-911-08-03)

作者简介: 李祚泳 (1944-), 男, 四川宜宾县人, 教授, 硕士。

$$c_i = \sum_{j=1}^p x_{ij}g_j \quad (i = 1, 2, \dots, m) \quad (2)$$

式中,  $g_j$  ( $\text{mg}/\text{m}^3$ ) 即为排放源  $j$  所排放的颗粒物在大气中的贡献量(可转化为源贡献率)。在污染源的个数  $p$  小于或等于元素个数  $m$  的情况下, 可以用最小二乘法等多种方法对式(2)求解而得到源贡献率  $g_j$ 。若采用 GA 算法求式(2)的最优解, 可构造模型优化准则的目标函数为:

$$\text{Min } F = \sqrt{\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p (x_{ij}g_j - c_i)^2} \quad (3)$$

## 2.2 成都市大气颗粒物源解析的优化求解

成都市是一个以尘污染为主的城市, 7 种主要尘污染源分别为土壤尘源( $x_1$ )、道路尘源( $x_2$ )、建筑尘源( $x_3$ )、工业燃煤尘源( $x_4$ )、民用燃煤尘源

( $x_5$ )、冶金尘源( $x_6$ )和汽车尾尘源( $x_7$ )。1992 年 1 月, 4 月, 7 月和 10 月 4 个季度, 在市区 10 个采样点, 以 5 日法进行采样。采样时间每日 8:00~20:00; 采样集于高密过氯乙烯有机纤维滤膜上; 采用重量法对样本进行分析。仪器选用青岛崂山电子仪器厂产的 KB-120 型采样泵, 配有 ZQ-10 型冲击式大气采样切割器分两级同时取样。采样高度距地面 12~15 m。综合市区 10 个采样点的大气颗粒物( $\text{PM}_{10}$ , 粒径  $d_j < 10 \mu\text{m}$ )采集样本和源解析有关的 12 种元素的含量分析, 得到成都市区大气颗粒物的 7 类主要尘排放源的元素组成和综合市区内 10 个采样点的大气样本元素的测试分析结果( $c_i$ ), 如表 1 所示。在目标函数式(3)下, 用 GA 对式(2)中参数  $g_j$  优化求解过程如下<sup>[7]</sup>:

表 1 成都市大气颗粒物的主要排放源元素组成及样本测试数据

Table 1 Element components of pollution sources of atmospheric particulates and monitoring data of samples in Chengdu Urban

元素	主要排放源的元素组成( $\mu\text{g}/\text{m}^3$ )							样品测试数据 $c_i$ ( $\mu\text{g}/\text{g}$ )
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	
Si	0.245 0	0.165 0	0.093 4	0.076 0	0.205 0	0.102 0	0.018 4	30.70
Ca	0.002 0	0.031 7	0.245 0	0.001 1	0.012 9	0.002 5	0.001 4	5.300
Mg	0.009 1	0.009 6	0.002 1	0.001 3	0.006 2	0.021 2	0.000 7	2.000
Fe	0.054 1	0.038 9	0.020 3	0.012 6	0.029 0	0.146 0	0.007 5	7.450
Mn	0.001 0	0.001 0	0.000 8	0.000 2	0.000 3	0.011 4	0.000 4	0.620
K	0.037 6	0.026 0	0.015 5	0.006 8	0.016 3	0.018 0	0.002 9	8.150
Al	0.060 4	0.049 2	0.022 4	0.023 4	0.027 3	0.016 5	0.003 4	5.110
Pb	0.001 1	0.000 8	0.000 2	0.000 4	0.000 3	0.001 1	0.001 3	0.590
Ti	0.000 5	0.000 2	0.001 5	0.000 5	0.001 0	0.000 2	0.000 1	1.830
As	0.002 3	0.001 0	0.000 6	0.001 3	0.001 4	0.009 9	0.000 3	1.620
Zn	0.001 5	0.000 9	0.000 3	0.001 4	0.000 6	0.009 8	0.000 1	0.910
Se	0.000 0	0.000 0	0.000 0	0.000 1	0.000 2	0.000 0	0.000 0	0.015
$g_k$	36.89	36.55	13.71	46.09	42.18	10.08	63.67	

a. 对参数  $g_j$  的二进制编码: 待估参数  $p = 7$  个; 凭经验设定  $g_j$  的取值范围为  $g_j \in [a_j, b_j] = [0, 200]$ , 并在  $[0, 200]$  内取  $2^{10} = 1024$  个离散点, 第  $k$  点取值为:

$$g_k = \frac{b_j - a_j}{2^{10} - 1} (k - 1), \quad k = 1, 2, \dots, 1024 \quad (4)$$

b. 初始父代个体群的随机生成: 从 1024 个解空间点中随机选取  $n = 300$  个点作为初始父代种群。

c. 父代个体的适应能力评价: 对 300 个父代个

体解码, 将表 1 中的各排放源的各种元素的测值  $x_{ij}$  及  $c_i$  值代入目标函数式(3), 求得相应的优化准则值  $F_k$ , 并根据下式计算第  $k$  个父代个体的适应度:

$$P'(k) = \frac{1/(F(k) \times F(k) + 0.001)}{\sum_{k=1}^n 1/(F(k) \times F(k) + 0.001)} \quad (5)$$

$$P(i) = \sum_{k=1}^i P'(k)$$

$$k = 1, 2, \dots, 300 \quad i = 1, 2, \dots, m$$

d. 父代个体按适应度的概率  $P_k$  大小进行选择: 从 300 个父代个体中选出两组各 300 个个体。

e. 父代个体杂交: 由两组各 300 个个体, 两两配对, 生成两组各 300 个新的子代个体。

f. 子代个体的变异: 任取上步中的一组 300 个子代个体, 将它们的二进制数的某段值依变异率 ( $C_a = 1$ ) 进行翻转操作, 即原二进制数码 0 与 1 互换。

g. 进化迭代把由上步得到的  $n$  个子代个体作为新的父代, 转入第  $c$  步操作, 进入下一次进化, 重新评价、选择、杂交和变异。

经验设置进化迭代次数, 在 586 微机经过 20 次迭代后, 把求得的优秀个体的参数区间作为 GA 的待估参数的新的变化区间, 再重新迭代 10 次后, 发现最优解连续保持不变, 输出此时最优解  $g_j$  列于表 1。把土壤尘源 ( $x_1$ ) 和道路尘源 ( $x_2$ ) 归并为土壤道路尘源; 把工业燃煤尘源 ( $x_4$ ) 和民用燃煤尘源 ( $x_5$ ) 归并为燃煤尘源, 归并后的 5 种尘源的归一化后的排放源贡献率  $g'_j$  如表 2 所示。

表 2 几种源解析法得到的成都市大气颗粒物排放源的贡献率  
Table 2 Contribution rates of pollution sources of atmospheric particulates in Chengdu Urban using some methods of sources apportionments

源解析法	$g'_1 + g'_2$	$g'_3$	$g'_4 + g'_5$	$g'_6$	$g'_7$	$\Sigma g'_j$
GA	29.5	5.5	35.4	4.0	25.6	100
PPR	41.6	4.3	44.5	3.6	6.2	100
TE	27.2	5.2	43.8	6.0	17.8	100
WLS	8.2	5.5	49.8	4.0	12.5	100
RR	24.3	3.4	52.2	6.3	13.8	100

### 2.3 结果分析与比较

用标识元素法 (TE)、普通加权最小二乘法 (WLS)、岭回归法 (RR)<sup>[3]</sup> 和投影寻踪回归 (PPR)<sup>[4]</sup> 等方法对成都市大气颗粒物源解析结果列于表 2。比较可见, 用 GA 法和其他几种源解析方法对成都市大气颗粒物源解析结果基本一致。GA 法用于大气颗粒物源解析的优点是不需要对排放源有特别的了解和限制, 只需按式 (3) 构造模型优化准则的目标函数, 并通过 GA 优化求解, 就能客观地求出各排放源对大气颗粒物的贡献率。

几种源解析法得出的排放源贡献率依从大到小排序的尘源依次是  $x_4 + x_5$  (燃煤尘)、 $x_1 + x_2$  (土壤道路尘) 和  $x_7$  (汽车尾尘)。上述 5 种尘源占全部尘源的 90% 以上, 因此, 应对这 5 类尘源采取措施加以控制和限制。

### 3 结论

a. GA 的处理对象是对参数集进行编码后的个体, 适用于各类优化问题。因此, 它用于大气颗粒物的源解析具有实际意义。

b. GA 作为一种新的全局优化搜索方法, 客观、

准确、简单、通用, 为大气颗粒物源解析提供了新的方法。其结果与实际情况相符合。

c. 应用 GA 的求解时间随问题规模增大而增长; 此外, 参数初始值范围的设定也会对收敛效率产生一定的影响。这些问题还有待于进一步加以研究解决。

### 参考文献:

- [1] Rauret G, Galceran M T, Rubio R, et al Factor analysis for assigning sources of ground water pollution [J]. International J Environ Anal Chem, 1990, 38: 389~ 397.
- [2] Hutter K M, Herrnberger V R D. Testing a source rate estimation method by application to a dispersion experiment [J]. Atmos Environ, 1997, 31(20): 3283~ 3290.
- [3] 周来东. 大气颗粒物来源解析数学模型评述 [J]. 城乡生态环境, 1994, 18(2): 11~ 20.
- [4] 李祚泳, 张欣莉, 丁晶. 成都市大气颗粒物源解析的 PPR 法 [J]. 环境科学研究, 2000, 13(5): 38~ 40.
- [5] Holland J H. Genetic algorithms [J]. Scientific American, 1992, (4): 44~ 50.
- [6] David B F. An introduction to simulated evolutionary optimization [J]. IEEE Trans. on NN, 1994, 5(1): 3~ 14.
- [7] 金菊良, 杨晓华, 储开凤. 基因算法的应用及改进 [J]. 河海大学学报, 1998, 26(2): 75~ 79.