

# 大气颗粒物源解析的遗传算法模型的改进

鲁磊<sup>1,2</sup>, 张卫东<sup>2</sup>, 蒋昌潭<sup>2</sup>, 赵琦<sup>1</sup> (1. 重庆大学城市建设与环境工程学院, 重庆 400044; 2. 重庆市环境科学研究院, 重庆 400020)

**摘要** 运用 MATLAB 遗传算法和直接搜索工具箱改进了应用于大气颗粒物源解析的遗传算法模型。引入数据进行算法程序探讨, 将模型计算结果用改进的二重源解析法进行了验证。结果显示, 改进的遗传算法源解析模型对数据处理快捷、简便, 在一定程度上能够解决严重共线性源类问题, 解析结果真实可靠。

**关键词** 源解析; 源类共线性; 遗传算法

中图分类号 S11 文献标识码 A 文章编号 0517-6611(2007)16-04823-03

## Improvement of Genetic Algorithm Model in Source Apportionment of Atmospheric Particulate

LU Lei et al (College of Urban Construction and Environmental Engineering, Chongqing University, Chongqing 400044)

**Abstract** The genetic algorithm model used in source apportionment of atmospheric particle was improved with the Genetic Algorithm and Direct Search Toolbox in MATLAB. The data were used to study algorithm procedures, and the calculating results were validated with the Improved Source Apportionment technique. The results showed that the improved genetic algorithm model used in source apportionment was quick and easy for data processing. And it could solve the problem of excessive source collinearity to a certain extent with its analytical results reliable.

**Key words** Source apportionment; Source collinearity; Genetic algorithm

大气颗粒物源解析技术是一种定性和定量研究大气颗粒物来源的技术。以化学质量平衡模型(CMB)为代表的受体模型被广泛使用。然而CMB模型要求各排放源相互独立, 即各排放源化学组成无共线性的存在, 在解析过程中严重共线性的源类不能同时纳入模型计算, 只能混合成一种源类, 或者舍去一些排放源类。这无疑降低了大气颗粒物治理的针对性<sup>[1-2]</sup>。冯银厂等在CMB模型的基础上将源解析过程分成单一尘源贡献对受体贡献求解、混合尘源对受体贡献求解、单一尘源对混合尘源贡献求解<sup>3</sup>步, 从中找出单一尘源和混合尘源之间贡献的定量关系, 从而解决了源类共线性问题<sup>[3-4]</sup>, 但是需要多步引入不同的排放源, 步骤较为繁琐。遗传算法应用于大气颗粒物的源解析是近年来国内学者为了简化模型算法而提出的一种模型方法<sup>[5]</sup>。该方法将计算机智能算法与源解析技术相结合, 基于全局优化并拥有极强的容错性, 不受条件的限制, 因此在一定程度上能够克服共线性源类不能同时纳入模型计算的问题, 但是算法过程不够完善。为此, 笔者将遗传算法过程加以改进, 并利用 MATLAB 语言编程进行实现, 同时引入实际数据进行试验, 用改进的二重源解析结果进行了验证。

### 1 模型简介

遗传算法是一类随机优化算法, 模拟了自然选择和遗传中发生的复制、交叉和变异等现象, 从任一初始种群(Population)出发, 通过停止准则判断种群是否为输出结果, 不满足准则的通过适应度评价, 高适应度个体被保留下来经过随机选择、交叉和变异操作, 产生一群更适应环境的个体, 使群体进化到搜索空间中越来越好的区域, 这样不断地繁衍进化, 最后收集到一群最适应环境的个体(Individual), 求得问题的最优解<sup>[6]</sup>。

遗传算法进行颗粒物源解析是由李祚泳等首先提出<sup>[5]</sup>。其基本思想是将CMB线性方程组的求解问题转化为一个非线性规划求解问题, 寻求目标函数的最优解。为了使模型解析结果更加准确, 将李祚泳模型优化准则的目标函数进行了

改进, 引入了受体和排放源元素测量值的标准偏差, 并且添加了约束条件:

$$MnF = \frac{\sum_{i=1}^m (C_i - \sum_{k=1}^p G_k \cdot X_{ik})^2}{\sum_{i=1}^m C_i^2 + \sum_{k=1}^p \sum_{i=1}^m X_{ik}^2 \cdot G_k^2} \quad (1)$$

$$[G_k] \in [0, ub] \quad (2)$$

式中,  $C_i$  为受体大气颗粒物中元素  $i$  的浓度测量值;  $X_{ik}$  为第  $k$  类源的颗粒物中元素  $i$  的含量测量值;  $G_k$  为第  $k$  类源贡献的浓度计算值;  $p$  为源类的数目,  $k=1, 2, \dots, p$ ;  $m$  为元素的数目,  $i=1, 2, \dots, m$ ;  $C_i$ 、 $X_{ik}$  分别为受体大气颗粒物的元素测量值和排放源的元素测量值的标准偏差;  $ub$  取值依据所测得的受体颗粒物浓度值  $M$ , 一般  $ub = M$ , 也可以适当缩小约束条件空间。

改进的优化准则采用 CMB 模型算法准则, 即使加权的元素测量值与计算值之差的平方和最小, 计算过程采用遗传算法模式。改进后的遗传算法模型是一种融入了 CMB 算法和遗传(GA)算法的混合模型。

### 2 遗传算法模型进行源解析的主要步骤

**2.1 优化准则目标函数的创建** 创建源码编辑文件(M-file), 编辑目标函数(1):

function y = fun(g, x, c, s, e)

y = sum(diag(diag(c - g \* x)^2 / diag(s + ...diag(diag(g)^2) \* e)));

上述命令中  $g$ 、 $x$ 、 $c$ 、 $s$ 、 $e$  均为矩阵的形式。其中,  $g$  表示  $1 \times p$  阶贡献值矩阵;  $x$  表示  $p \times m$  阶源成分谱矩阵;  $c$  表示  $1 \times m$  阶受体成分谱矩阵;  $s$  表示  $1 \times m$  阶受体成分方差(即标准偏差的平方)矩阵;  $e$  表示  $p \times m$  阶源成分方差矩阵。然后, 将编辑好的 M-file 以“fun.m”命名, 保存于 MATLAB 的 work 文件夹中。

**2.2 成分谱元素的选取和数据的输入** 在众多元素中, 一般选取表征各类源特点的且在迁移过程中变化不大的所谓标识性元素纳入模型计算。一些非标识性元素对结果可能产生干扰, 需要剔除。但是地区差异、采样分析误差以及排放源共线性等原因, 使得标识性元素不能仅靠直观或经验来选取, 因此成为源解析模型工作中最为繁琐的一环。遗传算

法基于全局搜索优化,对于个别元素的偏差有极强的容错性,因此大大简化剔除元素的工作。

以原点为初始解,利用 MATLAB 模式搜索目标函数(1)在约束条件(2)下的最小值  $f_{val}$  和对应点值  $g$ (排放源浓度值),计算出  $m \times 1$  阶有效方差矩阵  $Ve$ 。由于干扰元素以及严重的共线性源类的存在,对应点值势必达到约束条件的临界值,有效方差矩阵  $Ve$  显示了每个元素在模式搜索下加权的元素测量值与计算值之差所能达到的最小值,值越大表示该元素被优化的效果越差,即是干扰元素。

利用 MATLAB 语言编程,输入全部元素数据组成的源和受体成分谱及其方差矩阵  $x, s, c, e$ , 受体颗粒物浓度测量值  $M$ , 约束条件上限值  $ub$ , 运行如下程序(“%”后是程序的解释)即可自动找出并剔除干扰元素。

```
m = max(size(x)); p = min(size(x));
[g, fval] = patternsearch( @(g) fun(g, x, c, s, e), zeros(1, p),
[], [], [], [], zeros(1, p), ...ub * ones(1, p)),
Ve = diag(diag(c - g * x)^2 / diag(s + ...diag(diag(g)^2) *
e))'
% 模式搜索, 输出有效方差矩阵
x1 = []; c1 = []; s1 = []; e1 = []; limit = 1;
% 设置剔除元素有效方差限制值
for i = 1:length(Ve)
    if Ve(i) <= limit
        x1 = [x1, x(:, i)];
        c1 = [c1, c(:, i)];
        s1 = [s1, s(:, i)];
        e1 = [e1, e(:, i)];
    end
end
x = x1, c = c1, s = s1, e = e1, theremoved = find(Ve > limit) %
输出剔除元素后的成分谱及其方差, 并输出剔除的元素的
序数。
```

上述程序自动剔除了有效方差值大于1的元素。根据实际情况,修改有效方差限制值  $limit$  的大小,从而限制剔除的元素。

**2.3 算法参数的设定和运行** 遗传算法是一种计算机智能算法,其中有多参数的设定对结果产生一定的影响。MATLAB 软件中自带的遗传算法工具箱较好地实现了遗传算法,并且大多数参数的默认值均设置得较为恰当。利用 MATLAB 遗传算法工具箱进行来源解析,种群规模设置为100,初始种群范围设置同约束条件(2),终止代数和时间设置为无穷大,适应度目标值设定为0(其余参数设定均使用默认值)<sup>[6]</sup>。

遗传算法与 CMB 模型所用的有效方差最小二乘法最大的不同之处在于:有效方差最小二乘法是从单个初始值迭代求最优解的(假设初始贡献全为0)<sup>[7]</sup>,而遗传算法是从一个初始解集进行搜索,覆盖面大,利于全局寻优。每一次运行遗传算法,都是从一个随机产生的初始种群开始,因此每次得到的结果不同,同时由于严重共线性源类的存在,有时难免陷入无实际意义的负值结果(即贡献值出现负值)。遗传算法每次运算出的结果应当视为实际情况的一个概率值。

根据概率理论,多次遗传算法结果的累计平均值可以反映来源解析的实际结果,排除无意义的负值结果。

下面的程序调用遗传算法工具箱设置运行了100次,并自动将无意义的负值结果剔除,最后输出所有保留结果、保留运算次数、贡献浓度和贡献率的均值和标准差(“%”后为程序的解释)。

```
G = []; n = 100; % 遗传算法运行次数
options = gaoptimset( PopInitRange, [0;ub], ...
    PopulationSize, 100, Generations, inf, ...
    FitnessLimit, 0); % 设置遗传算法参数
for j = 1:n
    [g, fval] = ga( @(g) fun(g, x, c, s, e), p, options);
    if all(g >= 0) == 1
        G = [G; g];
    end
end % 剔除遗传算法运行出现的负值结果
G, num_G = max(size(G)), mean_G = mean(G),
std_G = std(G), rateg = mean_G / M * 100,
std_rateg = std_G / M * 100
```

% 输出遗传算法每次运行得出的贡献浓度矩阵,保留运算次数,贡献浓度和贡献率的累计均值以及标准差。

### 3 模型的验证

为了验证遗传算法进行大气颗粒物源解析的效果,引入某城市成分谱数据(表1、2)<sup>[4]</sup>。将引用数据输入编译好的程序,设置有效方差限制值( $limit$ )为1,遗传算法运行次数( $n$ )为100。上限值  $ub$  初始设置为  $ub = M = 248.3$ ,根据输出的累计平均贡献浓度分布情况缩小空间,最后确定  $ub$  为120,再次运行程序,输出结果中100次运算中保留了97次,97次运算输出贡献浓度的累计平均值见图1。保留运算次数为20次以前,扬尘和土壤尘的贡献浓度累计平均值变动最大,燃煤尘次之,并且三者曲线呈现互补,说明三者的源成分谱存在共线性,其中扬尘和土壤尘共线性最为严重,使得贡献浓度值在三者之间变动;其他源类贡献浓度累计平均值变动极小,可见基本不存在共线性。保留运算次数达到50次以后,发现各个源类的贡献浓度累计平均值变化极小。以97次保留运算结果累计平均值作为最终结果,分别得到各源类的贡献浓度,并用改进的二重源解析模型计算的结果进行了比较。表3表明,2种方法解析的7类尘源贡献的排序完全一致,仅在数值上有些偏差。

### 4 结论

(1) CMB 算法准则函数和遗传算法模式结合在一起形成的改进的遗传算法大气颗粒物源解析模型和二重来源解析模型方法一样,能够有效地将存在严重共线性源类的贡献值区分开来,并且2种方法所得各源类的贡献排序完全一致,仅在数值上有些偏差,解析结果真实可靠,运行简单,不需要多步输入不同源类进行计算。

(2) 利用 MATLAB 语言编程,设计自动剔除干扰元素的程序,结合遗传算法极强的容错性,大大减轻了来源解析的工作量。试验结果证明,该方法简便可靠。

表1 源成分谱数据 %

元素	土壤风沙尘		煤烟尘		建筑水泥尘		扬尘		尾气尘		海盐粒子		(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	
	x	s	x	s	x	s	x	s	x	s	x	s	x	s
Na	1.094	0.389	0.351	0.164	0.242	0.357	2.591	2.527	0.296	0.279	40.000	4.000	0.000	0.001
Mg	0.987	0.845	0.636	0.024	0.792	0.338	0.821	0.734	0.219	0.300	0.000	0.010	0.000	0.001
Al	4.182	1.037	16.422	0.542	5.132	0.616	6.566	1.398	0.266	0.153	0.000	0.010	0.000	0.001
Si	26.274	6.210	18.594	0.602	7.864	8.368	16.032	6.062	0.694	0.484	0.000	0.010	0.000	0.001
K	2.017	0.103	1.384	0.043	1.687	0.716	1.388	1.201	0.231	0.204	1.400	0.200	0.000	0.001
Ca	0.823	0.526	7.464	0.346	32.532	16.704	3.792	2.146	0.599	0.763	1.400	0.200	0.000	0.001
Ti	0.284	0.167	0.586	0.053	0.695	0.292	0.374	0.184	0.101	0.077	0.000	0.010	0.000	0.001
V	0.001	0.001	0.003	0.001	0.002	0.000	0.002	0.003	0.033	0.020	0.000	0.010	0.000	0.001
Cr	0.014	0.023	0.047	0.014	0.019	0.001	0.018	0.022	0.013	0.019	0.000	0.010	0.000	0.001
Mn	0.036	0.024	0.029	0.002	0.046	0.012	0.039	0.022	0.022	0.015	0.000	0.010	0.000	0.001
Fe	1.721	0.834	2.432	0.463	1.193	0.062	1.568	1.764	1.184	0.621	0.000	0.010	0.000	0.001
N	0.003	0.005	0.005	0.001	0.007	0.001	0.005	0.013	0.008	0.006	0.000	0.010	0.000	0.001
Cu	0.002	0.001	0.011	0.001	0.010	0.001	0.005	0.013	0.080	0.019	0.000	0.010	0.000	0.001
Zn	0.012	0.001	0.008	0.001	0.017	0.003	0.056	0.073	0.216	0.025	0.000	0.010	0.000	0.001
As	0.001	0.001	0.004	0.006	0.001	0.000	0.001	0.002	0.008	0.003	0.000	0.001	0.000	0.001
Pb	0.005	0.002	0.004	0.001	0.007	0.002	0.011	0.020	0.032	0.013	0.000	0.010	0.000	0.001
NO <sub>3</sub> <sup>-</sup>	0.034	0.024	0.057	0.058	0.006	0.016	0.177	2.090	0.774	1.079	0.005	0.002	0.000	0.001
SO <sub>4</sub> <sup>2-</sup>	0.427	0.423	2.247	0.525	7.162	0.014	5.945	4.665	3.872	3.010	10.000	4.000	72.697	7.270
Cl <sup>-</sup>	0.205	0.098	0.000	0.001	0.050	0.019	4.450	1.551	0.096	0.238	40.000	10.000	0.000	0.001
TC	1.205	0.154	13.457	12.437	1.660	1.099	7.702	1.814	89.870	8.987	0.000	0.010	0.000	0.001
OC	0.428	0.143	6.547	5.372	1.225	0.372	3.284	1.037	51.677	5.684	0.000	0.010	0.000	0.001

注: x 和 s 分别表示各排放源的元素含量值和标准偏差。

表2 受体成分谱数据 μg/m<sup>3</sup>

元素	贡献值	标准偏差	元素	贡献值	标准偏差
TOF	248.300 0	30.500 0	Fe	4.238 7	0.990 9
Na	9.394 8	0.954 3	N	0.095 0	0.042 2
Mg	1.405 2	0.842 9	Cu	0.044 2	0.047 0
Al	18.254 0	2.657 3	Zn	0.721 1	0.365 4
S	32.253 5	12.958 3	As	0.004 5	0.002 6
K	4.177 1	1.272 8	Pb	0.119 0	0.045 0
Ca	15.497 9	4.759 4	NO <sub>3</sub> <sup>-</sup>	1.095 1	11.531 5
Ti	0.497 9	0.491 9	SO <sub>4</sub> <sup>2-</sup>	22.211 5	10.403 7
V	0.006 4	0.004 0	Cl <sup>-</sup>	10.708 7	1.924 8
Cr	0.043 8	0.043 3	TC	30.950 4	8.469 5
Mn	0.098 6	0.049 5	OC	13.424 1	5.915 7

注: 受体成分谱贡献值和标准偏差单位。

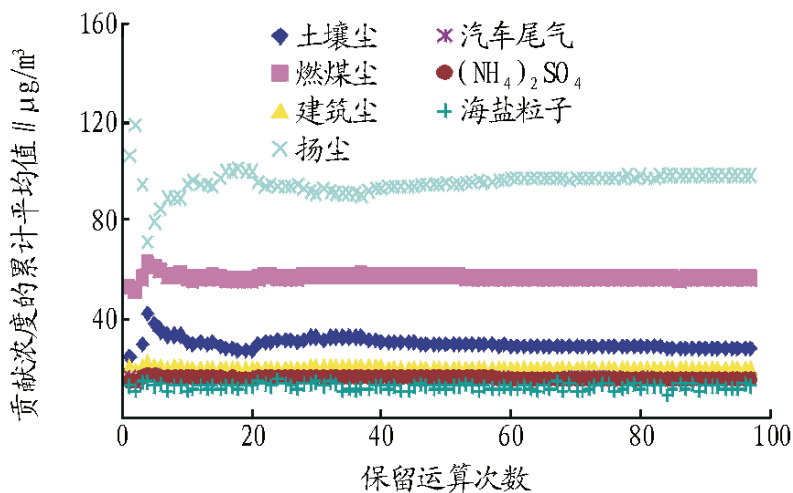


图1 97次遗传运算输出的贡献浓度累计平均值的散点趋势

表3 改进的遗传算法与改进的二重源解析法计算结果比较

源类	改进的遗传算法		改进的二重源解析方法	
	贡献值 μg/m <sup>3</sup>	分担率 %	贡献值 μg/m <sup>3</sup>	分担率 %
土壤尘	28.36	11.42	23.71	9.55
燃煤尘	56.84	22.89	50.10	20.18
建筑尘	20.02	8.06	23.49	9.46
扬尘	98.06	39.49	95.29	38.38
汽车尾气	15.79	6.36	19.42	7.82
海盐粒子	12.37	4.98	13.20	5.32
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	15.73	6.34	15.91	6.41

(3) 完善了颗粒物遗传算法解析模型的算法过程, 设计了多次运行遗传算法的程序, 用多次结果的累积平均值来反映实际情况下各源类的贡献值。试验证实, 该方法更加合理可靠。

参考文献

[1] 戴树桂, 朱坦, 白志鹏. 受体模型在大气颗粒物源解析中的应用和进展[J]. 中国环境科学, 1995, 15(4): 252 - 257.

[2] 冯银广, 白志鹏. 大气颗粒物二重源解析技术原理与应用[J]. 环境科学, 2002, 23(12): 106 - 108.

[3] 郝明途, 侯万国, 屈小辉, 等. 大气颗粒物二重源解析技术的方法改进[J]. 中国环境科学, 2005, 25(2): 138 - 141.

[4] 李祚泳, 彭茄红. 基于遗传算法的大气颗粒物的源解析[J]. 环境科学研究, 2000, 13(6): 19 - 21.

[5] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999.

[6] 雷英杰, 张善文, 李续武, 等. 遗传算法工具箱及应用[M]. 西安: 西安电子科技大学出版社, 2005.

[7] EPA. EPA CMB 2 uses manual[M]. Washington DC: Environmental Protection Agency & Air Quality Modeling Group, 2004.