

# 遗传算法在点源扩散浓度反演 排放源强中的应用<sup>①</sup>

陈军明 徐大海 朱 蓉

(中国气象科学研究院,北京 100081)

## 提 要

污染源清单的建立是目前城市大气污染浓度预报中的一大障碍。使用遗传算法和多点源扩散模式,从控制点浓度成功地反演了点源排放清单。源强反演数值的统计结果表明,该方法简单易行,效果较好,为空气污染预报系统提供了一种反演排放清单的途径。

**关键词:** 空气污染 源强 反演 遗传算法

## 引 言

近 20 年来,城市大气污染预报模式的研究得到了很大的发展。从过去的统计预报模式,已发展到今天的中尺度气象预报模式、污染扩散模式和光化学模式相结合的城市大气污染预报系统。如加拿大国家研究委员会的 MC2-CALGRID 模式系统<sup>[1]</sup>,是由一个非静力平衡的半显式半拉格朗日中尺度气象模式和一个综合(comprehensive)光化学模式结合而成的模式系统;德国 Cologne 大学的 EURAD 大气污染预报系统<sup>[2]</sup>由 MM5、EEM 污染扩散模式和 CTM2 化学传输模式组成。但绝大多数的大气污染物传输扩散模型都不可避免的依赖于污染源资料。然而,污染源调查是一个具有内在的、经常性的和相当大不确定性的复杂过程<sup>[3]</sup>,成为城市大气污染预报中的一大障碍。

空气污染预报的最终目标在于建立污染事件警报系统,在可能出现污染事件的时期可以采取自愿和强制性的减排措施。要求工业部门减少排放量,甚至完全停产,发电厂转为污染事件期间储备的低硫燃料。要求某些工厂在烟雾期间暂时只使用高烟囱排放,禁止低于 40m 的烟囱工业排放。要做到这一点,必须要了解污染浓度分布与排放源之间的关系<sup>[4]</sup>,为此,首先需要建立城市的排放

源清单。建立排放清单往往需要耗费大量人力物力对城市的每个污染源进行调查。长期以来,作为对源清单建立的补助方法,人们还使用接受器模型和源解析的方法。对观测地点得到的气溶胶化学成分的资料进行统计分析,可以直接判断各类源对气溶胶的化学成分浓度和总质量浓度的贡献,这种数学模型称为受体模型<sup>[5]</sup>。1991、1995 年徐大海和朱蓉<sup>[6,7]</sup>在 ATDL 模式的基础上导出了大气污染物面源源强分布到浓度分布的扩散稀释矩阵,将源强矩阵与浓度矩阵线性地联系起来,由当地的气候数据、地理条件、各类源高、排放气体的干湿沉降和衰变参数计算出扩散,直接由浓度矩阵求取源强矩阵,从而提出了由数值模式反演源强的方法。2001 年刘品高等将遗传算法引入大气污染总量控制中,应用它来进行面源源强的反演,用以确定地区的大气排放总量并且和稀释矩阵反演方法作了对比<sup>[8]</sup>。本文将遗传算法应用到点源源强的反演中,可以在浓度和点源源强之间建立一种直接的关系,为空气污染预报和总量控制提供基础。

## 1 遗传算法的基本原理

遗传算法是组合最优化算法的一种,要求解的数学问题模型为:

$$\min f(x)$$

<sup>①</sup> 《国家重点基础研究规划项目》:首都北京及周边地区大气、水、土环境污染机理与调控原理(G1999-045700)资助

$$\begin{aligned} s.t. & g(x) \geq 0 \\ & x \in D \end{aligned}$$

其中  $f(x)$  为目标函数,  $g(x)$  为约束函数  $x$  为决策变量,  $D$  为决策变量所属的有限域。

求解该问题的遗传算法是 70 年代初由美国 Michigan 大学的 Holland 发展起来的, 也是一种仿生算法, 主要借用生物进化中“适者生存”的规律, 利用某种编码技术作用于称为染色体的二进制数串, 从一个初始种群出发, 不断重复执行选择、杂交、和变异的过程, 使种群进化越来越接近某一目标, 从而求得某一实际问题的最优解。遗传算法利用简单的编码技术和繁殖机制来表现复杂的现象, 从而解决非常困难的问题。特别是由于它不受空间限制性假设的约束, 不要求诸如连续性、倒数存在和单峰等假设。与其他传统的相比, 它主要有以下几个不同之处:

(1) 它不是直接作用在参变量集上, 而是利用参变量集的某种编码。

(2) 它的操作对象是一组可行解, 因而具有良好的并行性。

(3) 它只需利用目标的取值信息, 无需梯度等高价信息, 具有通用性。

(4) 它的择优机制是一种“软”选择, 加上其良好的并行性, 使它具有良好的全局优化性和稳健性。

(5) 它操作的可行解集是经过编码化的(通常采用二进制编码), 目标函数解释为编码化个体(可行解)的适应值, 因而具有良好的可操作性与简单性。

由于具有上述特点, 因此从 20 世纪 70 年代以来, 该技术已经发展成为一种自组织、自适应的综合技术。适合数值求解那些带有多参数、多变量、多目标和在多区域但连通性较差的优化问题, 这类问题通过解析求解或是计算求最优解的可能性很小, 主要依赖于数值求解。遗传算法也是一种有普适性的数值求解方法, 它对目标函数的性质几乎没有要求, 甚至都不一定要显式地写出目标函数, 同求解问题的其他启发式算法有较好的兼容性。

### 1.1 遗传算法的基本步骤

实现最基本的遗传算法大致有以下的几个步骤:

(1) 对问题的参变量进行编码; 随机给出一个有  $N$  个染色体的初始群体  $\text{pop}(t)$ ,

$t := 1$ ;

(2) 对群体  $\text{pop}(t)$  中的每一个染色体  $\text{pop}_i(t)$  计算它的适应函数

$$f_i = \text{fitness}(\text{pop}_i(t));$$

(3) 给出一定的停止规则, 评价每一个染色体的适应度, 如果满足停止规则, 则算法停止; 否则计算概率

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}, i = 1, 2, \dots, N \quad (1)$$

并以概率分布式(1)从  $\text{pop}(t)$  中随机选一些染色体构成一个种群

$$\begin{aligned} \text{newpop}(t+1) &= \{\text{pop}_j(t) \mid \\ & j = 1, 2, \dots, N \}; \end{aligned}$$

(4) 通过交配, 交配概率为  $P_c$  得到一个有  $N$  个染色体的  $\text{crosspop}(t+1)$ ;

(5) 以较小的概率  $P_m$ , 使得一个染色体的基因发生变异, 形成  $\text{mutpop}(t+1)$ ;  $t = t + 1$ , 一个新的群体  $\text{pop}(t) = \text{mutpop}(t)$ ; 返回第二步。

上述的步骤不断循环, 最终将会逼近最优解<sup>[9~11]</sup>。

### 1.2 遗传算法解决源强反演的可行性

对污染源强的分布进行反演, 实际上就是根据控制点的浓度寻找一个最佳的源强分布场, 使得计算的控制点浓度尽可能的接近观测值。这样可以认为源强分布是合理的。因为各个污染源的排放强度都在一个有限的区域内, 故各污染源的源强组合是有限的。因此可以把污染源的源强反演归结为一个组合最优化问题, 其中计算的控制点浓度与观测值的误差限制可以构成目标函数, 而源到浓度的扩散模式构成约束函数。当计算的控制点浓度接近于观测浓度时, 目标函数  $f(x)$  达到最小, 即此优化问题有解。有定理可以证明<sup>[11]</sup>, 源强组合最优化问题在一定条件下, 其可行解集合是有限的, 该问题的最优解一定存在和可以得到, 就是说源强分布的最

优解也一定存在。

## 2 遗传算法应用于点源源强反演的计算方案

为了研究遗传算法应用于源强反演的实际运用效果,我们设计了一个遗传算法的软件,并对它进行了测试。在应用遗传算法求解该问题时,先完成以下几个步骤:

(1)确定表示方案:采用二进制编码,将污染源的排放强度编码成二进制,如果每个源强用  $L$  位二进制,有  $N$  点源,则每个染色体的长度为  $N * L$ ;

(2)构造适应度函数,确定适应值度量;

(3)确定控制算法的参数和变量;主要参数有群体规模  $N$ ,次要参数有复制概率  $P_r$ 、杂交概率  $P_c$  和变异概率  $P_m$  等参数;

(4)确定指定结果的方法和停止运行的准则。

基于遗传算法的源强反演方案的流程图如图 1 所示。

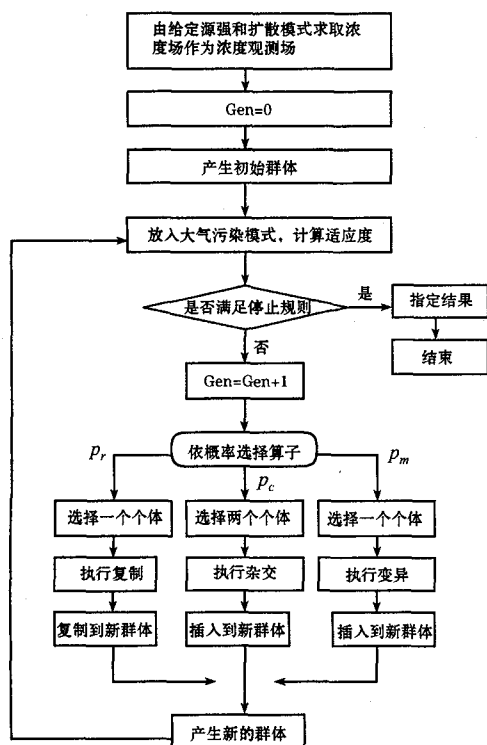


图 1 遗传算法的源强反演流程图

## 3 计算实例

### 3.1 计算方案设计

我们选取了美国环保局的具有地形调整的多点源模式(MPTER)<sup>[10]</sup>来嵌套到我们的算法模块中作为遗传算法中的约束函数。MPTER 是美国环保局的大气污染模式。它主要应用于乡村地形中计算空气污染物的浓度,同时它提供了可以选择的轻微的地形调整参数。它基于高斯型扩散模式,稳定度的分类采用 Pasquill-Gifford (P-G) 分类法。MPTER 对于点源的计算,是在 RAMR 基础上发展起来的,地形调整系数作为稳定度函数输入模式。最大的计算范围为 250 个点源和 180 个接受器<sup>[12]</sup>。首先使用美国环保局提供的每小时的气象资料和 30 个 TSP 源头的排放强度资料作为“源强实测值”,计算出 50 个接受器的浓度数据并以此作为“实测浓度”,并且由其中 30 个接受器的浓度数据反算上述 30 个排放数据,除了将反演出的源强与“实测源强”对比外,还使用反演源强和每小时的气象资料计算的 50 个接受器的浓度和“实测浓度”作对比。

在对源强进行编码时,为了达到足够的精度,我们采用 15 位二进制字符来表示一个十进制数。我们把初始资料代入扩散模式,求得各个控制点的浓度,假定这些浓度值没有误差(理想的),将其作为控制点的标准浓度。我们用浓度的计算值与控制点浓度的差值的倒数作为适应度函数,来评价染色体的性能。适应度值越大,表明计算的浓度值与控制点的浓度就越接近,染色体的性能就越好,被复制到新一代的概率越大。同时确定一个度量,当适应度值达到了一定的标准时结束进化,输出最优的源强分布场。

### 3.2 计算结果与分析

根据上述方案,我们作了一些数值试验。试验范围如图 2 所示。在不同气象条件下反演了 30 个点源的源强分布。结果反映随着迭代次数的增加,计算值和观测值的相关系数在增大。反演的效果就越好。迭代到 1500 次时平均的相关系数可以达到 0.59,迭代到 2500 次时,相关系数可以达到 0.65 以上,迭代到 4000 次时相关系数可以达到

0.80 以上,迭代 4000 次在 PⅢ550 的微机上要运算 8 个小时。结果如图 3~5 所示。反演的结果和控制点的位置有较大关系,如果控制点的位置有代表性,反演的效果就好一些,反之则会差一些。同时如果点源对控制点浓度的贡献率较大的话,则反演的效果会很好。反之则会较差。我们统计了贡献率最大的 10 个点源,反演的源强和观测值的相关系数均达到 0.80。使用反演的源强资料进行污染浓度的计算,各控制点的浓度的平均相对误差为 0.14。所以计算结果是可信的。但是我们从图 3 可以看到点源 5 和点源 8 误

差比较大,主要是它们的位置离得较近,比较难通过计算反演得出。点源 11 和点源 13 的反演值较小,是因为它们对选择的控制点浓度贡献较小,所以反演效果较差。

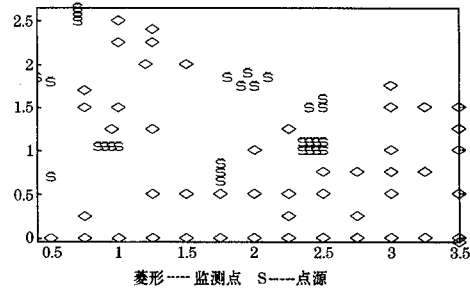


图2 点源的位置和监测点的位置

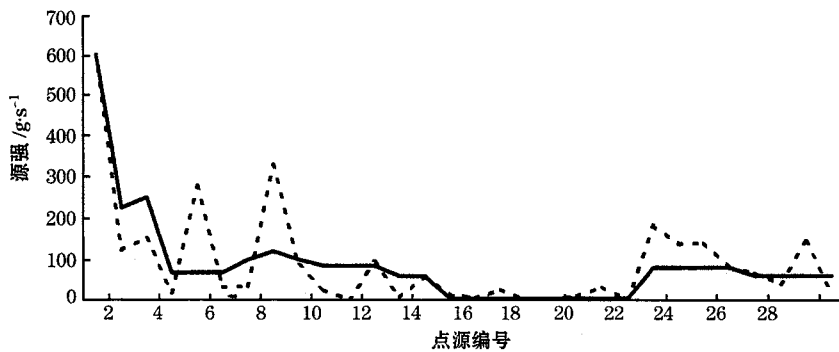


图3 迭代 4000 次,反演计算值与观测值的对比  
实线:实测值 虚线:反演计算值

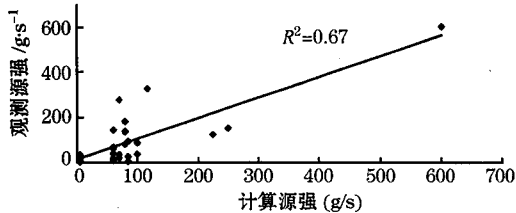


图4 计算值与观测值的相关(30 个点源)

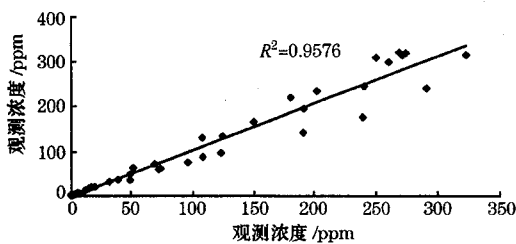


图5 观测浓度与计算浓度的相关(50 个监测点)

### 3.3 算法存在的问题和改进方向

目前算法的代码测试运行情况良好,但是还存在一些问题。遗传算法也是一种枚举

的方法,枚举是以时间为代价的,当要反演的源强规模较小时,枚举的时间还可以接受,但是当源排放清单的规模很大时,搜索的空间就会变得很大,这样计算时间也会增加非常的快,最后无法接受。我们通过对问题的深入了解,找出描述问题的所有参数的特性,以及最优解所要满足的条件等等,使问题具体化。例如每个排放源的源强都是有一定范围的,我们可以根据排放标准,大致确定排放强度的范围,这样就可以缩小搜索的空间,使得收敛的时间缩短。对于某些特定的点源,具有确定标准的排放值,这样就可以在编码中固定它们所对应的二进制位,使其不进行运算,这样也可以减少运算量。同时,还可以构造更好的适应度函数,使其能更科学地判断个体的优劣,引导进化向着正确的方向进行,减少冗长的搜索时间。

此外,反演的结果和选取的扩散模式有

很大的关系,模式的准确度越高,反演的效果就会越好。在多点源模式中,只考虑了物理过程,将来可以选取更好的模式和算法耦合,加入化学过程,以取得更好的反演效果。

#### 4 结论

本文尝试性地将遗传算法应用在污染源源强的反演方案中,通过计算结果可以得出以下结论:

(1)用遗传算法来反演污染源源强分布完全是可行的,计算结果也是可信的。

(2)将遗传算法和扩散模式结合进行反演,对于粒子的处理主要依赖于选取的扩散模式,与算法本身无关,这和传统的受体模式不同。

(3)传统的反演方法在源强和浓度的关系上,只能给出一种虚拟的联系,而遗传算法可以给出较为直接的关系,可以在很大的程度上,减少对大量的小源和不确定性污染源调查的工作量,从而为污染预报和污染控制提供了基础。

(4)用遗传算法来反演,可以简单而且直接地和其他扩散模式结合,遗传算法的模块和扩散模式的模块是相对独立的,二者结合的时候代码的编写工作量很小,而且不必对所解决的问题有太深的了解,所以通用性很强。

#### 参考文献

- 1 Hedley. M. A. et al. Evaluation of An Air Quality Simulation of the Lower Fraser Valley, — I meteorology, — II Photochemistry, Atmospheric Environment, 1997, 31:1605—1630.
- 2 Jakobs. H. j. et al. The Use of Nested Models for Air Pollution Studies: An Application of the EURAD Model to a SANA Episode. J. Appl. Meteor, 1995, 34: 1301—1319.
- 3 Report of the Meeting of Experts on Atmospheric Urban Pollution And the Role of National Meteorological Services, Geneva, 7—11, October 1996. No. 115: 10.
- 4 徐大海,朱蓉.城市空气污染数值预报系统 CAPPS 及其下一代的构想.全国城市空气污染预报及污染防治学术会议论文集.贵阳:中国环境科学学会大气环境分会, 2001.
- 5 王明星著.大气化学(第二版).北京:气象出版社,1999: 205~211.
- 6 徐大海,朱蓉.城市大气污染源强反演.大气环境和环境影响评价.北京:气象出版社,1998.
- 7 赵德山,徐大海.城市大气污染总量控制方法手册.北京:中国环境科学出版社,1991:244~299.
- 8 刘品高,余瑶等.遗传算法在大气污染总量控制中的应用研究.全国城市空气污染预报及污染防治学术会议论文集.贵阳:中国环境科学学会大气环境分会,2001.
- 9 张文修,梁怡等.遗传算法的数学基础.西安:西安交通大学出版社,2000:10~52.
- 10 刘勇,康立山,陈毓屏.非数值并行算法(第二册)——遗传算法.北京:科学出版社,2000:1~50.
- 11 邢文训,谢金星.现代优化计算方法.北京:清华大学出版社,1999:140~180.
- 12 User's Guide for MPTER A Multiple Point Gaussian Dispersion Algorithm with Optional Terrain Adjustment. 1980.

## Application of Genetic Algorithms to Point-source Inversion

Chen Junming Xu Dahai Zhu Rong

(Chinese Academy of Meteorological Sciences, Beijing 100081)

#### Abstract

The making an investigation of air pollution sources is a difficult thing for prediction of urban air quality. A new method is introduced into the inversion of air pollution point-source distribution. The direct relationship between concentration and source's discharge can be established by this method. Data of TSP from EPA were used to simulation a case.

**Key Words:** air pollution intensity of the source inversion genetic algorithm