

# 天津地区表土中蒽和芘的频率分布特征

段永红, 李本纲, 陶 澍\*, 王学军, 徐福留, 曹 军, 刘文新

(北京大学环境学院, 地表过程分析与模拟教育部重点实验室, 北京 100871)

**摘要:** 用频率分析、逐步聚类和最小二乘拟合等方法研究了天津地区表土中蒽和芘含量的频率分布特征。结果表明, 两者都呈双峰分布, 其中蒽的双峰特征尤为典型。可以用双高斯密度函数很好地拟合经对数变换的实际观测数据, 证明天津地区蒽和芘都包括两个相对独立的对数正态分布总体。根据拟合频率密度函数求得的蒽的两个子总体几何均值分别为  $2.73 \pm 2.17 \text{ ng/g}$  和  $26.52 \pm 1.55 \text{ ng/g}$ , 芘则为  $11.05 \pm 2.11 \text{ ng/g}$  和  $82.52 \pm 3.24 \text{ ng/g}$ 。芘的两个总体在空间上的分异比蒽更明显, 其高值总体主要集中在城区、近郊区和开发区, 而蒽的高值总体仅有局部聚集的一般趋势。芘与芘/苯并 [a] 芘之比的空间分布特征十分相似, 表明两个芘含量总体并存的原因与排放源特征有关。

**关键词:** 土壤; 多环芳烃; 蒽; 芘; 多态分布

**文章编号:** 1000-0585(2005)04-0559-06

多环芳烃是我国分布广、污染严重的持久性有机污染物。由于很强的大气迁移特性, 多环芳烃普遍存在于各种环境介质中。尽管如此, 一般情况下, 土壤仍是陆生生态系统中多环芳烃最重要的汇<sup>[1]</sup>。因此了解多环芳烃在土壤中的含量和分布具有重要意义。由于大量燃煤和其他来源, 天津地区遭受多环芳烃严重污染<sup>[2~5]</sup>, 且含量有继续增加的趋势。

污染物在特定介质中的频率分布可以从一个侧面反映其宏观统计特征。频率分布的另一重要意义是为经典、多元和空间统计分析提供必要的基础信息, 以确保数据满足各类统计分析的基本要求。对微量污染物而言, 由于含量很低, 且不可能为负值, 各类介质的区域性含量通常表现出典型的或近似的对数正态分布特点<sup>[6~8]</sup>, 其总体特征可以直接用几何均值等统计参数加以描述。少数情况下, 研究者也会发现多总体分布的特例。例如, Welp 等发现北莱茵-西伐利亚地区土壤中若干元素含量呈多态(峰)分布, 并据此可以区分这些元素的背景和污染来源<sup>[9]</sup>。目前有关土壤中 PAH 含量分布的报道, 如农业土壤、森林土壤、草地、城市土壤等不同土壤类型中测点 PAH 含量差异的比较<sup>[10]</sup>, 土壤剖面不同层次 PAH 含量的垂直分布<sup>[11]</sup>, PAH 在某种污染源周围的分布规律<sup>[12]</sup>, PAH 在不同粒径土壤颗粒上的分布规律<sup>[13]</sup>等一类测定样点不够多的研究中, 很难从宏观统计学角度分析数据的分布规律, 实质上谈的都是 PAH 在不同介质中分配比例问题, 结果表达多按正态分布处理。Nametal<sup>[14]</sup>和 Wang<sup>[15]</sup>等学者虽然分别在韩国和中国天津等较大的研究区域内测定了众多样点并使用 GIS 及空间插值等方法 and 手段探讨了 PAH 的区域分布规律, 但也未专门对众多 PAH 测定数据是否服从某种特殊的统计规律进行细致地研究; 对

收稿日期: 2004-07-14; 修订日期: 2004-10-15

基金项目: 国家自然科学基金(40332015, 40021101)、国家重点基础研究发展计划(2003CB415004)资助项目

作者简介: 段永红(1967-), 男, 山西榆社人, 副教授。主要从事环境科学及相关领域的研究。E-mail: yhduan-pku@sina.com

\* 通讯作者: 陶澍, 教授, 博士生导师。E-mail: taos@urban.pku.edu.cn, Tel: 010-62751938

结果的表示与分析,前者直接使用了测定值,后者则用了测定值的对数。

本文以天津地区表层土壤中具有特殊分布的蒽和芘为例,研究其频率分布特征以及分解为不同子总体的方法,为深入探讨不同多环芳烃分布差异及其成因提供方法学依据。

## 1 数据来源与研究方法

研究数据来自天津地区大致呈网格状均匀布点采样获得的 188 个表土样品中 2 种多环芳烃的 GC-MS 实测结果。详细的采样和分析方法参见相关文献<sup>[2~3]</sup>。在做频率分析前,尽管发现在全部 188 个数据服从正态分布假设的前提下可以检验出个别统计异常值,但考虑到多态分布的特点,剖分前暂不进行异常值剔除。

研究方法是首先对蒽和芘的含量测定值取对数,再用 spss 和 Matlab 对其进行频率分析,研究分组组数(8~99)对频率直方图分布形态的影响,发现蒽和芘分组数分别为 17 和 21 时可以得到最佳(平滑)结果,并以此为基础计算相应的组中值和频率密度;然后以组中值为自变量,相应的频率密度为因变量,采用 Matlab,直接调用其双高斯曲线模型和牛顿-高斯算法,在最小二乘条件下进行双峰分布的拟合,获得双峰均值、标准差和权重参数及拟合曲线的决定系数。以此为基础,采用二分法(即以拟合所得两高斯曲线的交点为界,把原总体分为左右两个总体)和三分法对双峰进行分离,并探讨分离总体的有关性质及双峰的成因。最后通过与用逐步聚类所得两类均值、样本量及各样本归属进行一致性比较,证实了对象的非单峰特征和双峰拟合分离结果的可靠性。

本文采用的三分法即以给定的标准差倍数为准则,将低值总体均值加上其标准差的某个倍数为右分界点,而高值总体均值减去其标准差的相同倍数为左分界点(右分界点取值应大于左分界点),把原总体全部样本分为三个不同的子总体,本文分别称之为低值组、混合组和高值组。其目的是想通过剔除位于两左右分界点间的样本,也即剔除与拟合双峰对应的低值总体高值部分和高值总体的低值部分,这样就有可能利用余下的样本在抛开不确定性较强的混合样的前提下探讨与拟合双峰对应的两独立总体的主要特征和成因。三分法混合样的多少可视研究的需要,通过调节标准差的倍数而变化。

## 2 结果与讨论

### 2.1 蒽和芘的频率分布特征

图 1 为分组统计得到的天津地区表层土壤中蒽和芘的频率分布,由图 1 可看出,天津

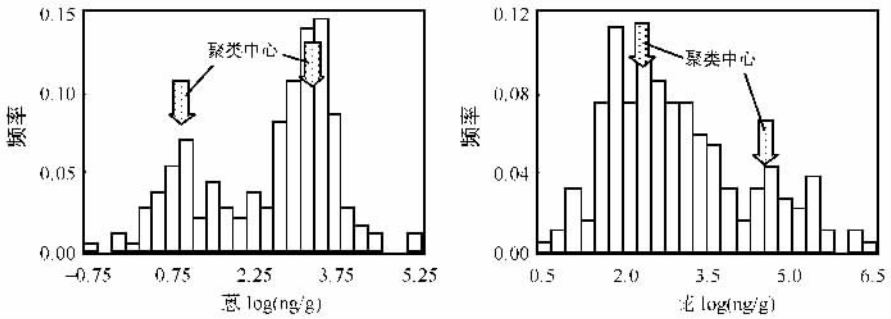


图 1 天津地区表土中蒽和芘的频率分布 (n=188)

Fig. 1 Histogram of anthracene and pyrene in topsoil from Tianjin

地区表土中蒽和芘均呈双峰分布特征, 其中蒽尤为典型。虽然蒽和芘含量的总体算术均值分别为 2.55 和 3.09 (ng/g, 经对数变换), 由于特定的分布形态, 它们并不反映真正的总体期望值。无论是蒽还是芘, 图 1 中所示分布形态基本是两个独立分布的叠加。从图 1 也可见, 蒽的两个独立峰的期望值分别在 1.0 和 3.4 左右, 芘的两个峰大约在 2.2 和 4.5 左右 (ng/g, 经对数变换)。由于图 1 显示结果已经对数变换, 以上结果显示天津地区这两种多环芳烃含量可能存在着两个相对独立的对数正态分布总体。尽管蒽和芘含量的频率分布均为双峰形态, 但两者差别十分显著, 蒽的高值总体样本数显著大于低值总体, 而芘的低值总体占绝对优势。

## 2.2 蒽与芘双峰分布的定量表征

**2.2.1 双峰拟合** 由 Matlab 拟合的图 1 中双峰分布频率密度函数的结果如图 2 所示, 图中实线为频率密度函数拟合结果, 虚线则为剖分后的两个独立对数正态分布函数。可见, 简单的双高斯曲线可以很好的拟合实测结果。对蒽和芘频率分布拟合的可决系数分别为 0.965 和 0.928。证实了它们在天津地区表层土壤中含量的双峰分布特征。根据拟合结果求得两个独立子总体的几何均值和几何标准差列于表 1 中拟合结果栏。为便于比较, 表中还给出了未剖分时全部数据的相应统计量。

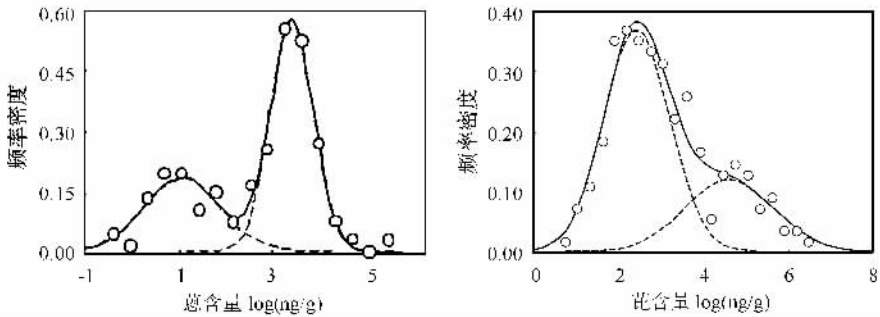


图 2 天津地区表土中蒽和芘的双峰分布拟合结果

Fig. 2 Least-square fitting of the bimodal distributions of anthracene and pyrene in surface soil from Tianjin with double Gauss curves

表 1 天津地区不同剖分方法剖分总体表土中蒽和芘含量的几何均值和几何标准差 (单位: ng/g)

Tab. 1 Geometric means and standard deviations of anthracene and pyrene in topsoil from Tianjin.

The results are based on different techniques of population classification (unit: ng/g)

剖分方法	拟合结果	拟合结果	未剖分	二分法	二分法	三分法	三分法	三分法
总体	第一总体	第二总体	全部数据	第一总体	第二总体	低值组	混合样	高值组
蒽	2.73±2.17	26.52±1.55	12.75±3.43	2.76±1.95	27.56±1.70	2.32±1.73	19.42±1.54	49.52±1.53
芘	11.05±2.11	82.52±3.24	22.08±3.45	11.40±1.93	109.36±2.14	5.50±1.41	18.14±1.63	136.17±1.94
蒽样本量	63	125	188	63	125	54	96	38
芘样本量	133	55	188	133	55	45	98	45

**2.2.2 双峰分离** 采用 Matlab 解非线性方程的方法求得蒽和芘两个总体分布密度函数的交点分别是 2.2906 和 3.6024, 以此为分界点, 用二分法可实现双峰的分离, 分离后各子总体的几何均值、几何标准差和样本量见表 1。

以表 1 中拟合获得的双峰几何均值和几何标准差的对数为准, 分别采用 3.29 倍和

1.96 倍标准差 (对应于 99.9% 和 95% 置信区间) 对蒽和芘的实测数据进行三分, 获得的三个子总体的几何均值和标准差结果见表 1, 图 3 则为由此得到的低值组和高值组的空间分布, 可见, 芘的高值组和低值组在空间上有比较明显的分布特征。虽然从图 1 和 2 的视觉上看, 芘的两个总体重叠似乎比蒽严重, 但其空间集聚和自相关程度更加明显。高值总体主要分布在城区、近郊及开发区。相比之下, 蒽的高值和低值样点在全市分布比较分散, 但是仍表现出相对集中的趋势, 即高值点有局部聚集的趋势。

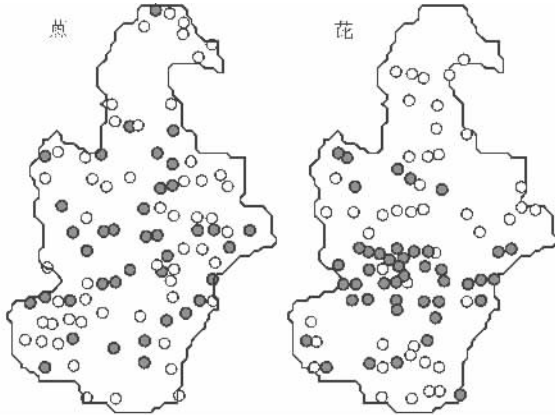


图 3 剖分得到的蒽 (3.29 倍标准差剖分) 和芘 (1.96 倍标准差剖分) 含量的高值和低值总体 (图中实心圆为高值总体, 空心圆为低值总体)

Fig. 3 Sub-populations of anthracene (3.29 std) and pyrene (1.96 std) based on standard deviations of their log-normal distribution (solid and open symbols represent high or low value population, respectively)

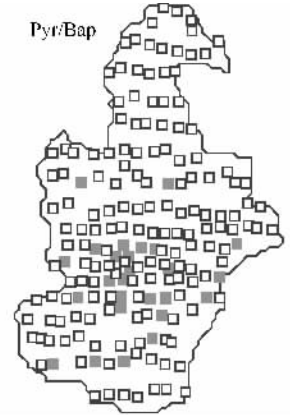


图 4 芘/苯并 [a] 芘比值的空间分布 (比值大于 2 和小于 2 者分别用实心和空心符号表示)

Fig. 4 Spatial distribution of pyrene/benzo [a] pyrene ratio (solid or open symbols represent samples with Pyr/Bap larger or less than 2, respectively)

**2.2.3 双峰拟合分离结果的验证** 采用逐步聚类法将蒽和芘的各 188 个对数含量数据聚为两类, 两个类别均在 0.001 以下水平显著, 且均符合对数正态分布。如果分别对两个类别在对数变换前提下进行异常值检验, 仅在蒽的高值组中发现 2 个异常。蒽双峰的聚类中心分别为 0.96 和 3.29, 样本量分别为 60 和 128 个; 而芘聚类中心则为 2.34 和 4.51, 样本量分别为 123 和 65。使用双峰拟合分离方法拟合获得的蒽的两个均值位置分别是 1.003 和 3.278, 对应的两类的样本量分别是 63 和 125; 芘的两个均值位置分别是 2.402 和 4.413, 对应的两类的样本量分别是 133 和 55。可见二者都比较接近, 这一结果也从另一侧面证实了其双峰分布的事实与拟合剖分的合理性。

### 2.3 蒽与芘双峰分布的成因分析

据有关文献, 可以用芘/苯并 [a] 芘之比判断 PAH 的污染来源。燃煤排放的该比值小于 2 而燃油排放的该比值大于 2<sup>[6]</sup>。根据天津地区 188 个样点的实测结果, 将该比值以 2 为界在图 4 表示 (图 4 也可作为蒽与芘 188 个样点的空间位置分布图)。图中结果表明, 燃煤污染是天津地区多环芳烃的主要来源, 其次是交通燃油。这与天津市环保局内部资料 (天津市重点污染源检测排放报告, 2001a) 显示的, 天津市工业污染能源结构中一次性能源以煤为主, 约占 90.02%; 其次是燃料油, 约占 2.93% 的数据是相符的。

PAH 的含量及其分布与污染源、迁移动力、土壤有机质含量、土壤中微生物降解、

淋溶、植物吸收、挥发和光解等众多因素、过程及土壤和 PAH 本身的性质密切相关, 是个非常复杂的问题。图 4 表现出来的空间分布特征与图 3 芘的两个总体的分布非常相似, 都是高值区多位于天津的市区及近郊区, 这可能与该区域的公路密度和交通流量都是全区最高的, 燃油量也较高有关, 可见天津市中心区及近郊区与其它以燃煤污染为主地区相比高的交通燃油污染及其空间上的集中分布很可能是造成芘呈双峰分布的主要原因。关于蒽含量统计学上呈双峰分布, 空间上整体呈集聚分布却不十分明显的原因可能较为复杂。总之, 对天津土壤中蒽和芘的对数含量呈双峰分布这个新发现特征的成因还有待进一步深入研究。

### 3 结 论

天津地区蒽与芘含量的对数值呈双峰分布, 用双高斯模型可拟合其密度函数, 获取双峰的均值位置、标准差及双峰分界点。以双峰密度函数曲线交点为分界点的二分法或以双峰均值加减适当倍数标准差构造的中间交叉区间的三分法都可合理剖分两个对数正态分布总体。蒽和芘, 尤其是后者的两个总体表现出特定的空间分布特征。其中芘的两个总体的空间分布直接反映了燃煤和交通燃油排放源的空间分异。

#### 参考文献:

- [1] Wang X L, Tao S, Xu F L, *et al.* Modeling the fate of benzo[a]pyrene in the wastewater-irrigated areas of Tianjin with a fugacity model. *Journal of Environmental Quality*, 2002, 31(3): 896~903.
- [2] 邢一, 王学军, 李本纲, 等. 天津地区表层土壤多环芳烃含量的中尺度空间结构特征. *环境科学学报*, 2003, 23(3): 311~316.
- [3] 王学军, 任丽然, 戴永宁, 等. 天津市不同土地利用类型土壤中多环芳烃的含量特征. *地理研究*, 2003, 22(3): 360~366.
- [4] 孙韧, 朱坦. 天津局部地区大气颗粒物上多环芳烃分布状态. *环境科学研究*, 2000, 13(4): 14~17.
- [5] 崔艳红. 土壤中多环芳烃对植物的生物有效性. 北京: 北京大学环境学院, 2003.
- [6] Ames M, Gullu G, Olmez I. Atmospheric mercury in the vapor phase, and in fine and coarse particulate matter at Perch River, New York. *Atmospheric Environment*, 1998, 32(5): 865~872.
- [7] Beier R. Calculation of expected air quality parameters based on monitoring results and modelling of planned additional sources. *Gefahrstoffe Reinhaltung Der Luft*, 1999, 59(4): 143~149.
- [8] Kishne A S, Bringmark E, Bringmark L, *et al.* Comparison of ordinary and lognormal kriging on skewed data of total cadmium in forest soils of Sweden. *Environmental Monitoring and Assessment*, 2003, 84(3): 243~263.
- [9] Welp G, Liebe F, Helfrich H P, *et al.* Inorganic toxicants in soils of North Rhine Westphalia. I. Evaluation of multi-modal frequency distributions. *Zeitschrift fur Pflanzenernahrung und Bodenkunde*, 1998, 161(3): 205~210.
- [10] Wolfgang Wilcke. Polycyclic aromatic hydrocarbons (PAHs) in soil—a review. *Journal of Plant Nutrition and Soil Science*, 2000, 163(3): 229~248.
- [11] Cousins I T, Gevao B, Jones K C. Measuring and modelling the vertical distribution of semivolatile organic compounds in soils. I: PCB and PAH soil core data. *Chemosphere*, 1999, 39(14): 2507~2518.
- [12] Brandt C A, Becker J M, Porta A. Distribution of polycyclic aromatic hydrocarbons in soils and terrestrial biota after a spill of crude oil in Trecate, Italy. *Environmental Toxicology and Chemistry*, 2002, 21(8): 1638~1643.
- [13] Muller S, Wilcke W, Kanchanakool N, Zech W. Polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBS) in particle-size separates of urban soils in Bangkok, Thailand. *Soil Science*, 2000, 165(5): 412~419.
- [14] Nam J J, Song B H, Eom K C, *et al.* Distribution of polycyclic aromatic hydrocarbons in agricultural soils in South

Korea. *Chemosphere*, 2003, 50: 1281~1289.

- [15] Wang X J, Zheng Y, Liu R M, *et al.* Kriging and PAH pollution assessment in the topsoil of Tianjin area. *Bulletin of Environmental Contamination and Toxicology*, 2003, 71(1): 189~195.
- [16] 赵振华, 等. 多环芳烃的环境健康化学. 北京: 中国科学技术出版社, 1993. 13~17.

## Frequency distributions of anthracene and pyrene in the topsoil of Tianjin

DUAN Yong-hong, LI Ben-gang, TAO Shu, WANG Xue-jun,  
XU Fu-liu, CAO Jun, LIU Wen-xin

( Laboratory for Earth Surface Processes, College of Environmental Sciences,  
Peking University, Beijing 100871, China)

**Abstract:** Frequency distributions of contaminants in special medium can reflect from one profile its macroscopic statistic characteristics, and can also supply necessary basic information for the classical, multivariate and spacial statistic analysis, so as to ensure basic needs to all kinds of methods for statistic analytical data. So for special studies on statistic distribution characteristics of PAH content with large numbers of specimen in a vast region have been rarely seen. Characteristics of frequency distributions of anthracene and pyrene in the topsoil of Tianjin, China, which was polluted seriously by PAHs, were studied based on frequency analysis, K-Means clustering and least-square fitting. It was demonstrated that both anthracene and pyrene, especially the former, are bimodally distributed. Double gaussian function can be used to fit the empirical distribution well after log-transformation, suggesting the presence of two independent lognormally distributed populations for either anthracene or pyrene. The geometric means of the two populations were  $2.73 \pm 2.17$  ng/g and  $26.52 \pm 1.55$  ng/g for anthracene and  $11.05 \pm 2.11$  ng/g and  $82.52 \pm 3.24$  ng/g for pyrene. Hereby, the research which adopts two-division-methods and three-division-methods indicates; that the spatial separation of the two pyrene populations is explicitly illustrated by symbol mapping, showing that the individuals of the high value population clustered in urban, near-suburb, and newly developed industrial areas. For anthracene, there is a tendency of local clustering. Besides, the similarity between the spatial distribution patterns of pyrene and pyrene/benzo[a]pyrene ratio suggests that the dual-modal distribution of pyrene was emission-source related. Fitting and separating methods of multi-modal distributions mentioned above provided methodological grounds for deep probing into their differences and origins of the distributive characteristics of different PAHs.

**Key words:** soil; PAH; anthracene; pyrene; multi-modal distributions