

用因子分析法研究大气气溶胶的来源

王 明 星

(中国科学院大气物理研究所)

提 要

本文简要介绍了因子分析法的统计学基础,提出了一种因子分析与经验正交变换相结合的分析方法。用这种方法研究了1980年3月在北京北郊用条式滤膜采样仪采样,用质子荧光法进行元素分析所得到的16种元素,40个样品的资料,得到了土壤尘加燃煤烟尘、汽车废气、石灰水泥粉尘、硫酸盐、草木燃烧、石油化工和海盐粒子等7种气溶胶来源对北京北郊大气气溶胶的相对贡献以及这些源的元素成分廓线。

一、引 言

大气气溶胶是我国许多城市,特别是北方干燥地区城市最主要的大气污染物。北京大气悬浮颗粒物的浓度经常超过国家环境质量标准。控制颗粒物排放是当前改善北京大气环境的首要任务之一。为了制定经济有效的污染控制措施,不仅需要综合评价环境质量现状,实测不同天气条件下大气气溶胶的浓度和化学组成,还需要对气溶胶的来源进行定量分析,从而做到有针对性的治理。过去几十年中,环境科学家主要依据源排放资料,利用扩散模式来计算污染物在所研究地区的分布。随着气溶胶采样和分析技术的提高,人们有可能积累大量有关气溶胶化学成分的资料。对这些资料进行统计分析可以建立接收点模式。所谓接收点模式就是由接收点(采样点)测量的气溶胶质量浓度和各种化学成分的浓度用统计学方法来推断各类污染源对测点气溶胶的贡献。这种方法不需对污染源进行普查,也无需考虑从污染源到接收点之间的化学变化,在实际应用上显示了很大的优越性。接收点模式大致可分为两大类,一类叫做化学元素质量平衡法,另一类叫做因子分析法。本文简要介绍因子分析法的统计学基础,把因子分析与经验正交变换相结合,分析了1980年3月在北京北郊取得的气溶胶元素浓度资料,得到了各类源对气溶胶的相对贡献和源化学成分廓线。

二、因子分析法的统计学基础

因子分析法的基本原理是利用资料点之间的相关性,把相关的变量用一个因子来代表,在维数减少了的空间中表征原来的样品集合。

1983年9月15日收到。1983年11月23日收到修改稿。

在大气气溶胶研究中,我们通常要从大气中抽取大量(设为 N 个)样品,现代分析技术可以从每一个样品分析出几十种(设为 M 种)化学成分的含量。这样,每一次实验就有 $M \times N$ 个资料,称为一个样品集合。这样的样品集合可以由 M 维空间的 N 个点完全表征。但是,每一个气溶胶样品的化学组分并不是各自独立、互不相干的,来自同一种源的那些成分之间存在较强的相关性。表征样品集合的空间的维数可以减少,有可能用 P 个($P < M$)因子来描述原来的样品集合。

在一点观测的大气气溶胶的各种元素的浓度可以认为是各种源的贡献的和,即

$$C_x = C_{x1} + C_{x2} + C_{x3} + \cdots + C_{xp}, \quad (1)$$

其中, C_x 是接收点某样品中 x 元素的质量浓度, $C_{xk}(k=1, 2, \cdots, p)$ 是第 k 种源对 x 元素浓度的贡献, p 是对接收点气溶胶有贡献的源的总数。每种源的贡献都可以写成两个因子乘积的形式,即

$$C_{xk} = a_{xk} F_k, \quad (2)$$

其中, a_{xk} 表示第 k 种源中 x 元素的相对含量, F_k 是第 k 种源对接收点气溶胶质量浓度的相对贡献。对所有 N 个样品 M 种元素我们有

$$C_{ij} = \sum_{k=1}^p a_{ik} F_{kj} \quad i=1, 2, \cdots, M, j=1, 2, \cdots, N \quad (3)$$

其中, C_{ij} 是第 j 个样品中第 i 种元素的测量浓度, a_{ik} 是第 k 种源的化学组成廓线, F_{kj} 是第 k 种源对第 j 个样品的贡献。写成矩阵形式,我们有

$$C = A \cdot F, \quad (4)$$

其中, C 是 $M \times N$ 阶资料矩阵, A 是 $M \times P$ 阶源化学组成廓线矩阵, F 是 $P \times N$ 阶源相对贡献矩阵。因子分析的目的就是要用适当的方法由资料矩阵 C 求出矩阵 A 和 F 。这在数学上有许多方法。一种较好的方法是经验正交展开^[4]。

设 A_1 为 P 个样准化正交向量构成的正交向量族,则样品向量在这些正交向量上的投影分量可以写成

$$F_1 = A_1^T C, \quad (5)$$

把投影向量投影到原来的 M 维空间的坐标轴上,就得到它在原空间的表示式,为区别起见记为 C' ,我们有:

$$C' = A_1 \cdot A_1^T \cdot C. \quad (6)$$

问题是要求出合适的 A_1 ,使得用 C' 代替 C 误差最小,并且能够通过适当的线性变换由 A_1 求出 A ,亦即使 A_1 的变换矩阵的每一列各与一种源的化学组成廓线联系起来。这就归结为求协方差矩阵的特征值和特征向量。若 A_1 为协方差矩阵的特征向量矩阵的前 P 个特征向量(对应于前 P 个最大特征值),则用 C' 代替 C 的误差可用 δ_p^2 来度量

$$\delta_p^2 = \sum_{m=p+1}^M \lambda_m, \quad (7)$$

λ_m 是协方差矩阵的特征值,按由大到小的顺序排列,更详细的论述可参见[4]。

求得 A_1 后就可通过适当的正交变换使变换后矩阵的每一列与一种源的化学组成廓线联系起来,即求得(4)式中的 A ,并进而求出 F 。

三、北京北郊 1980 年 3 月气溶胶样品的因子分析

1980 年 3 月 6 日—20 日我们在北京天文台兴隆观测站进行了一次气溶胶综合观测。作为对比观测,我们在北京北郊气象塔的实验楼顶用条式滤膜采样仪收集了气溶胶样品,所得样品用质子 X-射线荧光(简记为 PIXE)法分析了 16 种化学元素的浓度^[1]。北京气象塔位于北京城区和郊区的接壤处,周围比较开阔,局地大气污染源较少。因此,在这里收集的样品可以代表一个地区的平均状态,其中持续北风条件下的样品可代表城市郊区的大气,南风条件下的样品可代表城市烟羽。我们根据采样时气象塔提供的风向风速资料,把 84 个样品(每 4 小时 1 个)进行了分类,选出两组进行因子分析。第一组由稳定北风条件下的 17 个样品组成,风向都在西北至东北范围之内,风速为 1—5 米/秒;第二组为南风条件下的样品 23 个,风向在西南至东南,风速为 1—5 米/秒。表 1 列出了第一组样品分析得到的前 7 个特征向量组成的矩阵,即 A_1 , 及相应的特征值。可以看出,第一个

表 1 因子负荷 A_1

因子序数							
特 征 值	45.36	0.686	0.146	0.048	0.017	0.011	0.003
Mg	0.232	-0.118	0.676	-0.508	1.076	5.706	0.020
Al	2.761	-1.667	-2.278	2.894	0.162	0.694	-0.107
Si	5.546	-1.439	0.606	-0.400	2.398	-0.106	-0.008
S	2.038	6.100	-1.490	0.067	-0.395	0.301	-0.567
Cl	0.644	0.617	2.182	5.470	6.692	-0.103	-0.157
K	0.665	0.405	0.846	0.650	1.216	-0.027	5.110
Ca	1.920	0.567	6.124	1.344	-0.346	0.073	-0.053
Ti	0.156	-0.024	-0.117	0.125	-0.040	0.105	0.059
V	0.007	0.014	-0.006	-0.099	-0.079	-0.007	0.261
Mn	0.042	-0.007	0.131	0.180	0.482	0.028	-0.117
Fe	1.115	-0.023	0.730	0.562	-0.524	-0.449	0.335
Cu	0.005	0.017	-0.037	-0.104	-0.053	-0.020	0.145
Zn	0.093	0.106	0.291	0.548	1.317	1.234	0.322
Br	0.009	0.007	-0.008	-0.048	0.022	0.062	-0.012
Sr	0.028	0.019	-0.051	-0.005	-0.037	-0.156	-0.074
Pb	0.069	0.203	0.027	0.270	-0.009	0.018	2.019

因子含有丰富的 Al, Si, S, Ca 和 Fe, 而且大部分其它元素也有一定负荷, 它显然与土壤尘和燃煤烟尘有关;第二个因子只对 S 有较高负荷, 它应代表二次污染物硫酸盐;第三个因子与 Mg, Cl, Ca 等元素有关, 以 Ca 的负荷最大, 它可能与水泥、石灰粉尘有关;第四个因子和 Cl 的关系最大, 它可能代表海洋大气的影响。第五个因子对 Cl, Zn, Si 和 Mg 等元素有贡献而以 Cl 和 Zn 的负荷最高, 可能是草木燃烧和城市垃圾处理的排放物;第六个因子与 Mg 和 Zn 有关, 而以 Mg 的负荷为高, 这个因子所代表的源不太容易确定。根据文献 [1] 给出的资料, 这一因子可能代表重油燃烧的排放物, 然而这里 V (重油燃烧排放物的代表元素) 的负荷并不高。这可能部份地由于北京气溶胶中 V 含量较低, 许多样品中 V 的含

量仅略高于探测极限,有些样品中V含量低于探测极限而被取为零,就是说,在资料矩阵中V元素浓度误差较大;另外,气溶胶中的V容易在输送过程中变成挥发性物质而从颗粒物中失去,所以这第六个因子可能代表经过长距离输送的老化的重油燃烧排放物;第七个因子只对Pb有较高负荷,显然应代表老化了的汽车废气。由于上述同样的原因,这个因子的Br负荷不高。

根据上述分析,我们以文献[1]和[3]给出的各种源的元素组成廓线作为初始检验矢量,用迭代法求变换矩阵,并对 A_1 进行正交变换,求出了北京地区各类源的元素组成廓线,记为 A 列出表2, F_1 的相应变换后的矩阵也同时求出,记为 F 。样品集合在这组新的正交向量族上展开,记为 C_1 ,显然有

$$C_1 = A \cdot F - A_1 R R^{-1} F_1 = C' \quad (8)$$

表2 变换后的源成分廓线

因子序号	1	2	3	4	5	6	7
对 应 源	土壤尘和煤烟	硫酸盐	水泥石灰粉尘	海 洋	草木燃烧	石油化工	汽车
Mg	61.53	—	273.68	33.80	7.42	6.96	—
Al	2698.58	11.62	1034.34	1.11	4.14	1.74	0.21
Si	6938.42	21.70	697.60	—	2.21	—	0.039
S	396.98	2669.69	—	0.20	0.09	0.02	—
Cl	11.48	12.78	—	589.96	134.89	29.80	0.18
K	485.11	18.82	263.90	10.36	16.19	59.01	0.11
Ca	250.40	0.85	2122.19	13.83	92.06	1.47	0.19
Ti	152.20	16.39	50.20	1.57	2.46	—	0.39
V	16.01	0.67	6.18	—	—	1.69	0.08
Mn	55.18	—	—	34.47	7.97	0.71	0.18
Fe	1096.36	7.34	298.84	13.96	16.72	5.09	—
Cu	19.29	7.81	—	—	—	0.91	—
Zn	19.48	5.28	2.27	81.81	18.86	9.45	0.36
Br	15.80	5.46	—	—	—	—	—
Sr	36.52	13.69	—	—	—	—	0.08
Pb	8.80	4.10	89.70	—	10.19	14.72	1.89

其中 R 为转换矩阵。因此,用 C_1 来代表原来的资料矩阵,其相对误差为^[4]

$$\Sigma_p^2 = \sum_{m=p+1}^M \lambda_m / \sum_{m=1}^M \lambda_m, \quad (9)$$

这是我们因子分析质量的一个度量。对上述第一组样品集合的分析结果 $\Sigma_p = 0.061$ 。为了更清楚地说明分析误差,我们可以将由(8)式计算的 C_1 与原来的资料矩阵一一比较,限于篇幅,表3列出了一例。表3给出第一组第一个样品的计算元素浓度与原始资料的比较。可以看出,对于大多数元素来说两者非常接近,相对误差很小,只有那些浓度很低,原始资料误差较大的元素才有明显的相对误差。整个样品平均相对误差为0.089,与整个样品集合的平均相对误差 Σ_p 接近。对其它样品有完全类似的结果,限于篇幅,不能一一列出。这一结果表明,用7个因子能较好地表征北京大气气溶胶的元素浓度,就是说,北京北郊气溶胶主要有7种来源。

表 3 样品 1, 计算元素浓度与原始资料的比较

元 素	计算浓度	原始资料	相对误差
Mg	158.13	160.55	0.015
Al	1668.29	1667.85	0
Si	3540.44	3539.92	0
S	2861.48	2861.43	0
Cl	637.73	639.02	0.002
K	667.17	666.79	0.001
Ca	1451.85	1451.08	0
Ti	103.95	125.51	0.1718
V	10.06	10.79	0.068
Mn	24.43	26.07	0.063
Fe	782.63	813.66	0.038
Cu	9.05	11.84	0.236
Zn	102.88	58.11	0.435
Br	7.87	8.90	0.116
Sr	21.08	15.23	0.277
Pb	102.37	102.26	0

表 4 北风条件下各种源对元素浓度的贡献

浓度 (相对 单位) 元素	源	土壤尘	硝酸盐	水 泥 石灰粉尘	海洋	草木燃烧	石油化工	汽车	计算浓度	计算浓度 与实测浓 度之比
		加煤烟								
Mg		61.9	—	262.6	22.9	20.0	15.0	—	382.4	1.17
Al		2715.5	12.2	992.6	0.8	11.2	3.8	4.0	3740.1	1.00
Si		6981.9	22.8	669.4	0	6.0	0	0.7	7680.8	1.00
S		399.5	2807.3	0	0	0	0	0	3206.8	1.00
Cl		11.5	13.4	—	399.6	364.2	64.4	3.4	856.5	1.09
K		488.1	19.8	253.2	7.0	43.7	127.5	2.2	941.5	1.01
Ca		252.0	0.9	2036.5	9.4	248.6	3.2	3.6	2554.2	1.00
Ti		153.1	17.2	48.2	1.1	6.6	—	7.3	233.5	1.08
V		16.1	0.7	5.9	—	—	3.6	1.5	27.8	2.11
Mn		55.5	—	—	23.3	21.5	1.5	3.5	105.3	2.04
Fe		1103.2	7.7	286.8	9.5	45.2	11.0	0	1463.4	0.97
Cu		19.4	8.2	—	—	—	2.0	—	29.5	2.76
Zn		19.6	5.5	2.2	55.4	50.9	20.4	6.8	160.8	1.31
Br		15.9	5.7	—	—	—	—	—	21.6	1.52
Sr		36.8	14.4	—	—	—	—	1.6	52.8	1.30
Pb		8.8	4.3	86.1	—	27.5	31.8	36.0	194.5	1.86

在特征向量矩阵中有些元素是负值, 表示相应的源很少或不排放这类元素。当这些源作用时, 其它元素浓度上升, 这些负值负荷的元素浓度相对下降, 可能低于该元素浓度的本底值。表 4 列出了第一组样品元素浓度在各种源上的分配结果, 表 4 最后一栏给出了计算元素浓度与原始资料之比, 这是我们对元素浓度分配结果的质量的直接衡量。可以看出, 除少数几个浓度较低的元素外这一比值都很接近 1。差别较大的几个元素皆因

表 5 北风条件下各种源对元素浓度的相对贡献

元素 \ 源	源						
	土壤尘	硫酸盐	水泥石灰粉 尘	海 洋	草木燃烧	石油化工	汽 车
Mg	16.2	0	68.7	6.0	5.2	3.9	0
Al	72.6	0.3	26.5	0	0.3	0.1	0.1
Si	90.9	0.3	8.7	0	0.1	0	0
S	12.5	87.5	0	0	0	0	0
Cl	1.3	1.6	46.7	42.5	7.5	0.4	0
K	50.8	2.1	26.9	0.7	4.6	13.5	0.2
Ca	9.9	0	79.7	0.4	9.7	0.1	0.1
Ti	65.6	7.4	20.6	0.5	2.8	—	3.1
V	57.9	2.5	21.2	—	—	12.9	5.4
Mn	52.7	—	—	22.1	20.4	1.4	3.3
Fe	75.4	0.5	19.6	0.6	3.1	0.8	0
Cu	65.8	27.8	—	—	—	6.8	—
Zn	12.2	3.4	1.4	34.4	31.7	12.7	4.2
Br	73.6	26.4	—	—	—	—	—
Sr	69.7	27.2	—	—	—	—	3.0
Pb	4.5	2.2	44.3	—	14.1	16.3	18.5

表 6 南风条件下各种源对元素浓度的相对贡献

元素 \ 源	源						
	土壤尘加煤 烟	硫酸盐	水泥石灰粉 尘	海 洋	草木燃烧	石油化工	汽 车
Mg	27.7	0.5	40.7	21.1	3.7	2.3	3.9
Al	59.5	0.2	39.7	0	0.1	0.5	0
Si	83.9	0.3	15.5	0	0	0	0.2
S	10.6	89.3	0	0	0	0	0.1
Cl	0.5	0.5	—	83.6	6.3	9.0	0
K	30.2	1.5	38.4	2.6	0.3	26.9	0
Ca	5.5	0	90.9	0.9	2.2	0.4	0
Ti	35.5	20.6	40.6	3.2	0	0	0
V	22	42.1	20.2	25.7	2.7	—	7.1
Mn	43.3	—	—	16.5	—	40.2	—
Fe	82.6	1.6	1.4	0.4	1.9	0	12.1
Cu	53.9	0.7	—	—	—	29.4	16.0
Zn	—	15.7	43.4	37.4	3.2	—	—
Br	—	55.9	17.7	24.7	1.6	—	—
Sr	41.6	18.8	—	13.2	2.6	—	23.8
Pb	—	24.2	49.8	—	10.4	15.5	—

原始资料测量误差较大,在分配结果中有负值出现。

对第二组南风条件下的样品进行分析得到了完全类似的结果,限于篇幅不再将中间结果一一列出,只给出元素浓度在各种源上的分配,与第一组样品进行一些分析比较。为更清楚地说明各种源对北京郊区气溶胶(第一组北风条件下样品)和城区(第二组南风条

件下样品)气溶胶元素浓度的贡献大小,我们分别对两组样品计算了各种源对元素浓度的百分比贡献,如表 5 和表 6 所示。由表 5 和表 6 看出,Al, Si, Ti, Fe 等在气溶胶中含量丰富的元素主要来自土壤尘和燃煤烟尘,这与一般规律一致,但北京气溶胶中的 Al 和 Fe 还有相当一部份来自水泥、石灰粉尘。特别值得注意的是北京气溶胶中的 Ca 主要不是来自土壤尘而是来自水泥、石灰粉尘,这是北京气溶胶中的 Ca 富集度很高,远高于地壳平均值的缘故。北京气溶胶中的 Pb 来源比较复杂,贡献最大的不是汽车废气而是水泥、石灰粉尘,草木燃烧和石油化工均有一定贡献。这种来源分布与欧美城市完全不同。其原因是北京汽车流量较少,所用汽油的含铅量也比较少。这里需要指出的是北京气溶胶中铅元素的浓度较低,在 1980 年 3 月采样的许多样品中铅的浓度接近 PIXE 的探测极限,原始资料误差较大,所以,关于 Pb 来源的分析不如其它元素可靠,但可以肯定北京汽车铅污染不严重。

四、气溶胶总质量浓度的分析

如果用于分析的所有样品的气溶胶总质量浓度已知,则因子分析能够直接给出各种源对气溶胶总质量浓度的贡献。事实上,(3)式可以改写成

$$C_{ij} = \sum_{k=1}^p a_{ik} F_{ki} = \sum_{k=1}^p \frac{a_{ik}}{b_k} (b_k F_{ki}), \quad (10)$$

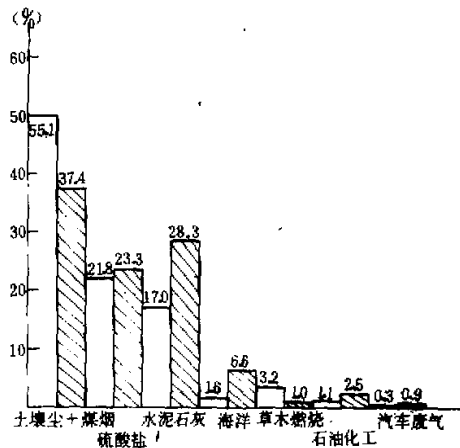


图 1 各种源对气溶胶总质量浓度的贡献

▨ 代表南风条件下样品 □ 代表北风条件下样品

其中, $\frac{a_{ik}}{b_k}$ 表示第 k 种源中第 i 种元素的相对浓度, 而 $b_k F_{ki}$ 则表示第 i 个样品中由第 k 种源贡献的质量浓度, 所以第 i 个样品的总质量浓度就由下式给出

$$M_i = \sum_{k=1}^p b_k F_{ki}. \quad (11)$$

因此,由第三节所得到的 F 矩阵和各个样品的质量浓度,很容易用最小二乘法求出 b_i ,从而求出各种源贡献的质量浓度。我们于 1980 年 3 月在北京北郊所取气溶胶样品未能测量总质量浓度,不能用上述方法计算各源对质量浓度的贡献。我们可以根据各种源的排放物中某些特定元素的浓度与该源的排放物之质量浓度的关系^[3]由表 4 所列数据估算出各种源对总质量浓度的贡献。第二组样品可作同样处理,所得结果一并绘于图 1。

图 1 清楚地表明,北京气溶胶最主要的来源是土壤尘加煤烟,硫酸盐和水泥石灰粉尘。这四种源的贡献占气溶胶总质量的 90% 以上。图 1 还表明,在北京气象塔处,北风条件下和南风条件下各种源相对贡献的主要差别是土壤尘加煤烟的贡献。北风条件下土壤尘加煤烟的相对贡献要大得多。这显然是由于城区其它污染源要比郊区严重。另外在南风条件下海洋大气的影响较大,这是由于东南风能将海洋大气带到测点。

五、讨 论

本文结果表明用经验正交展开因子分析法能够从一点测量的大气气溶胶元素浓度推断气溶胶的来源,这种方法的优越性是显而易见的,它将在环境科学上发挥作用。

对于本文所用实验资料中所包含的元素来说,土壤尘和燃煤烟尘具有非常类似的元素相对组成,所以因子分析未能把土壤尘和燃煤烟尘分开,而是作为一个因子。对于环境科学研究来说,把土壤尘和燃煤烟尘定量地分开无疑是非常重要的。为了做到这一点,用于分析的资料中应当包括两者中相对含量差别较大的化学成分,如元素 As 和多环芳烃。但是,这类化学成分在气溶胶中含量极低,需要经常测量以保证所需要的精度,并且要在因子分析时对它们作适当加权处理。这是因子分析法的一个困难所在,也是一个重要的研究课题。

为了能准确地确定各种源对气溶胶总质量浓度的贡献,在将来的观测中要在测量元素浓度的同时,测量每个样品的总质量浓度。

本文结果表明,北京气溶胶的最主要成分是土壤尘加燃煤烟尘(约占 50%),其次为硫酸盐(约占 20%)和水泥石灰粉尘(约占 20%)。值得注意的是,尽管北京气溶胶中硫酸盐的相对含量比欧美城市中(一般占 50% 以上)低得多,但硫酸盐的绝对浓度已远远超过欧美城市。因为北京气溶胶的总质量浓度一般要比纽约市高 5 倍以上,冬季可高一个量级^[2]。

致谢: 在研究过程中得到曾庆存教授的热情指导,作者深表谢意。

参 考 文 献

- [1] Gordon, G. E., *Ann. of N. Y. A. S.*, Vol. 338, pp. 93—102, 1980.
- [2] 王明星,科学探索,卷 3, No. 2, pp. 13, 1983.
- [3] Hopke, P. K., *ACS Symposium Series*, 167, 21—49, 1981.
- [4] 曾庆存,大气红外遥测原理,附录 3,科学出版社,1974.
- [5] 王明星等,环境科学丛刊,卷 2, No. 2, 1—10, 1981.
- [6] 任丽新等,大气科学,卷 6, 11—17, 1982.
- [7] Winchester J. W. et al., *Nuclear Instruments and Method*, Vol. 181, 391—398, 1981.
- [8] 王明星等,大气科学,卷 5, 136—144, 1981.

SOURCE IDENTIFICATION AND APPORTIONMENT FOR ATMOSPHERIC AEROSOL THROUGH THE APPLICATION OF FACTOR ANALYSIS

Wang Mingxing

(Institute of atmospheric physics, Academia Sinica)

Abstract

The statistical basis for factor analysis is briefly discussed. A method combining common factor analysis with orthogonal transformation has been proposed and discussed in detail. The method has been tested with a data set consisted of 16 elemental concentrations in 40 samples collected at the Beijing meteorological tower, March 1980. 7 sources, soil+coal fly ash, sulfate, cement and lime dust, oil, plant burning, motor, and marine aerosol, corresponding to 7 factors have been identified and the source profiles obtained.