

HJ

中华人民共和国国家生态环境标准

HJ 1353—2024

环境空气 颗粒物来源解析 正定矩阵 因子分解模型计算技术指南

Ambient air—Source apportionment on particulate matter—Technical guide
on positive matrix factorization model calculation

本电子版为正式标准文本，由生态环境部环境标准研究所审校排版。

2024-02-18 发布

2024-05-01 实施

生态环境部 发布

目 次

前 言.....	ii
1 适用范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 模型原理和计算流程.....	2
5 数据准备.....	3
6 基础计算.....	4
7 旋转计算.....	6
8 颗粒物源贡献计算.....	6
9 结果合理性判断.....	6
附录 A（资料性附录） 颗粒物源类识别方法及主要源类的标识组分.....	8
附录 B（资料性附录） PMF 模型计算记录表.....	9



前 言

为贯彻《中华人民共和国环境保护法》《中华人民共和国大气污染防治法》，加强大气污染防治，保护和改善生态环境，保障人体健康，规范环境空气颗粒物来源解析工作中正定矩阵因子分解（PMF）模型计算，制定本标准。

本标准属于环境空气颗粒物来源解析系列标准之一，规定了 PMF 模型计算方法。

本标准的附录 A 和附录 B 为资料性附录。

本标准首次发布。

本标准由生态环境部大气环境司、法规与标准司组织制订。

本标准主要起草单位：中国环境监测总站、南开大学、中国环境科学研究院。

本标准生态环境部 2024 年 2 月 18 日批准。

本标准自 2024 年 5 月 1 日起实施。

本标准由生态环境部解释。



环境空气 颗粒物来源解析 正定矩阵因子分解模型计算技术指南

1 适用范围

本标准规定了正定矩阵因子分解（PMF）模型计算的数据准备、基础计算、旋转计算、颗粒物源贡献计算、结果合理性判断等技术要求。

本标准适用于应用 PMF 模型对环境空气颗粒物进行来源解析相关的计算工作。

2 规范性引用文件

本标准引用了下列文件或其中的条款。凡是注明日期的引用文件，仅注日期的版本适用于本标准。凡是未注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

HJ 630	环境监测质量管理技术导则	
HJ 1352	环境空气 颗粒物来源解析技术规范	基于手工监测的受体模型法监测数据处理与检验技术规范
	《环境空气颗粒物来源解析监测技术方法指南》（监测函〔2020〕8号）	

3 术语和定义

下列术语和定义适用于本标准。

3.1

环境受体 ambient receptor

受到污染源排放影响的环境空气，简称受体。

3.2

正定矩阵因子分解模型 positive matrix factorization model (PMF)

将样本数据的 X 矩阵分解为因子贡献 G 矩阵和因子谱 F 矩阵，对因子谱 F 矩阵进行识别，并定量计算样本的因子贡献的一种多元因子分析类模型。

3.3

污染源化学成分谱 source chemical profile

特定污染源类排放颗粒物的相对稳定的化学组分信息（g/g），简称源成分谱。源成分谱应包含该类污染源的标识组分。

3.4

污染源贡献 source contribution

通过受体模型计算得到的污染源对环境空气颗粒物的贡献，用贡献浓度（ $\mu\text{g}/\text{m}^3$ ）或贡献占比表示。

3.5

污染源标识组分 source chemical tracer

污染源化学成分谱中对该源类有指示作用的一种或多种颗粒物化学组分，又称示踪组分。源标识组分是区别该源类与其他源类的重要标识物，每种污染源有各自的标识组分。

3.6

因子 factor

PMF 模型计算提取的 G 矩阵和 F 矩阵的每列信息，其中 F 矩阵的每列可识别为特定污染源类，G 矩阵的每列可计算得到特定污染源类的贡献。

3.7

Q 值函数 function Q value

反映 PMF 模型计算结果质量的一种目标函数。

3.8

残差 residual

颗粒物化学组分的实际测试值与 PMF 模型计算值之间的差值。

3.9

基础计算 base run

在计算过程中不加入任何数学、物理意义的约束条件，只根据模型 Q 值函数的收敛条件（Q 值最小）得到因子贡献 G 矩阵和因子谱 F 矩阵的一种 PMF 模型计算方法。

3.10

旋转计算 rotational run

对于基础计算得到的因子贡献 G 矩阵和因子谱 F 矩阵，加入一定的约束条件再次进行计算，转换成另一对矩阵（G*和 F*），使各因子谱的标识组分更加突出，从而解决因子谱难以识别问题的一种 PMF 模型计算方法。

4 模型原理和计算流程

4.1 模型原理

在环境空气颗粒物来源解析中，PMF 模型计算是将受体颗粒物化学组分浓度矩阵 $X_{(n \times m)}$ 因子化，分解为两个因子矩阵， $G_{(n \times p)}$ 和 $F_{(p \times m)}$ ，以及一个残差矩阵 $E_{(n \times m)}$ （见公式 1），在 G 矩阵和 F 矩阵中的数值都是正值（即非负限制）的限定下，通过定义一个目标 Q 值函数（见公式 2），选择 Q 值函数值最小时的 G 矩阵（因子贡献矩阵）和 F 矩阵（因子谱矩阵）计算结果，再利用污染源标识组分将 F 矩阵识别为不同的源类，对 G 矩阵进行多元线性回归计算，从而得出不同源类对环境样品的贡献。

$$X_{(n \times m)} = G_{(n \times p)} \times F_{(p \times m)} + E_{(n \times m)} \quad (1)$$

式中： $X_{(n \times m)}$ ——受体颗粒物化学组分浓度 X 矩阵；

$G_{(n \times p)}$ ——因子贡献 G 矩阵；

$F_{(p \times m)}$ ——因子谱 F 矩阵；

$E_{(n \times m)}$ ——残差 E 矩阵；

n ——样品个数；

m ——化学成分种数；

p ——解析的因子（污染源）数目。

$$Q = \sum_{i=1}^n \sum_{j=1}^m (E_{ij} / \sigma_{ij})^2 \quad (2)$$

式中： Q ——目标 Q 值函数；

E_{ij} ——第 i 个样品中第 j 个化学成分的残差；

σ_{ij} ——第 i 个样品中第 j 个化学成分的不确定度；

n ——样品个数；
 m ——化学成分种数。

4.2 计算流程

应用 PMF 模型进行颗粒物来源解析的计算流程包括数据处理和检验、基础计算及结果分析、旋转计算及结果分析、模型计算结果合理性判断等环节，见图 1。PMF 模型计算需要进行反复优化计算，调整计算数据和参数，以得到相对较优的计算结果。

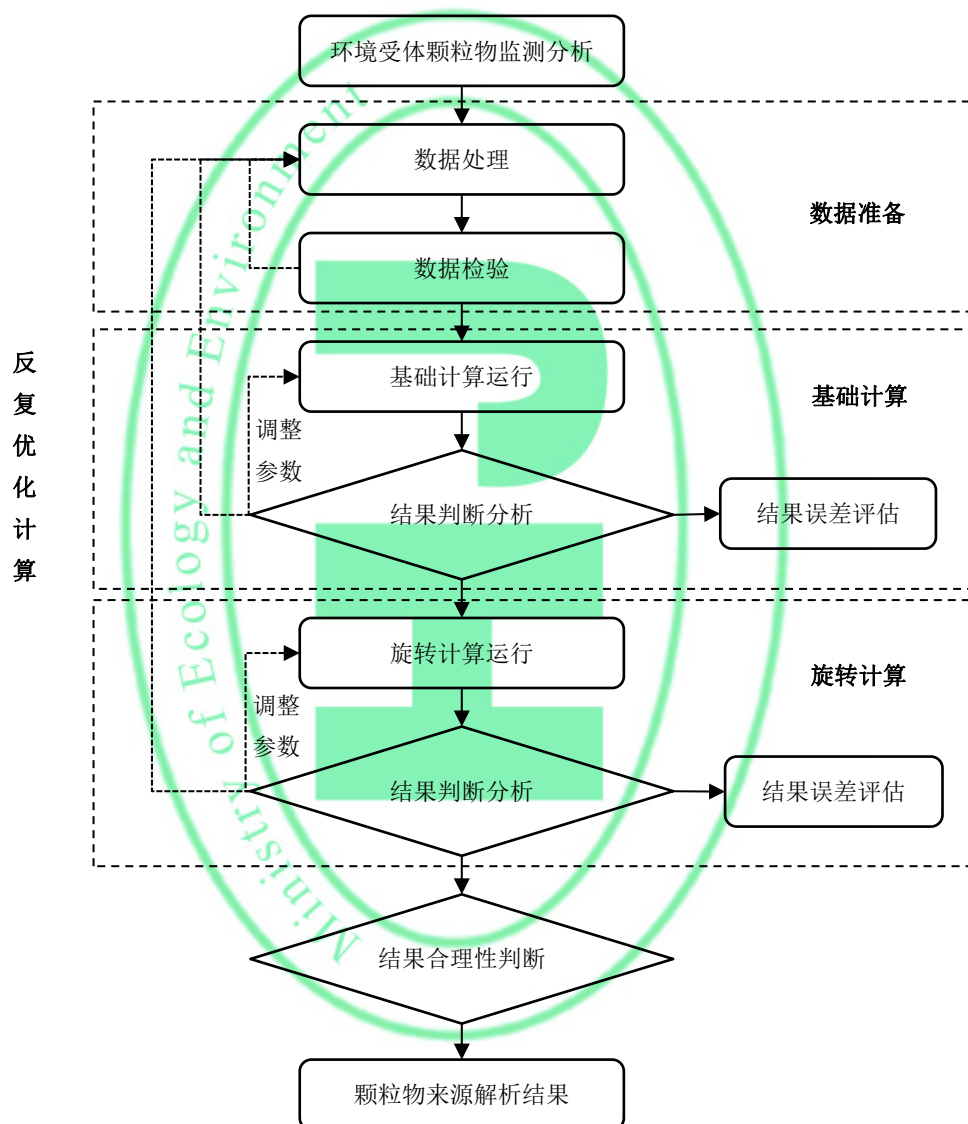


图 1 PMF 模型计算流程

5 数据准备

5.1 数据种类和数量要求。PMF 模型需要输入环境空气颗粒物质量浓度以及无机元素、碳组分、水溶性无机离子等化学组分浓度数据，应不少于 100 个有效样品的数据量。颗粒物化学组分需涵盖主要污染

源的标识组分，常用的标识组分包括有机碳（OC）、元素碳（EC）、 SO_4^{2-} 、 NO_3^- 、 Cl^- 、 NH_4^+ 、Si、Al、Na、Mg、K、Ca、Mn、Fe、Cu、Zn 等。为解析更多的源类，可增加多环芳烃、左旋葡聚糖、正构烷烃、脂肪酸、甾醇类、有机酸等一次和二次有机标识组分。

5.2 数据质量要求。用于颗粒物来源解析工作的监测数据，应按照采用国家标准、行业标准、地方标准的优先次序获得；无国家标准、行业标准、地方标准的，可采用国际标准、国外标准或研究建立的方法，经评估确认满足方法要求后使用，同时监测工作需按照 HJ 630 的要求纳入质量管理体系，开展质量保证和质量控制工作并符合相关要求。

5.3 数据格式要求。PMF 模型需要输入两个数据文件：（1）颗粒物及化学组分质量浓度的数据文件，（2）颗粒物及化学组分质量浓度不确定度的数据文件。两个数据文件内容需要对应，格式符合模型软件的要求。

5.4 数据检验。对于每个颗粒物样品的颗粒物及化学组分质量浓度数据，按照 HJ 1352，进行颗粒物化学组分质量平衡、阴阳离子电荷平衡等数据检验。

5.5 不确定度计算。不确定度是颗粒物来源解析工作中采样、实验分析等环节引起的误差，直接影响输入 PMF 模型中样品颗粒物及化学组分质量浓度数据的计算权重，根据以下公式进行计算：

$$Unc = \frac{5}{6} \times MDL \quad (3)$$

式中：Unc——颗粒物及化学组分的不确定度， $\mu\text{g}/\text{m}^3$ ；
MDL——颗粒物及化学组分的检出限， $\mu\text{g}/\text{m}^3$ 。

$$Unc = \sqrt{(EF \times Conc)^2 + (0.5 \times MDL)^2} \quad (4)$$

式中：Unc——颗粒物及化学组分的不确定度， $\mu\text{g}/\text{m}^3$ ；
MDL——颗粒物及化学组分的检出限， $\mu\text{g}/\text{m}^3$ ；
Conc——颗粒物及化学组分的质量浓度， $\mu\text{g}/\text{m}^3$ ；
EF——颗粒物及化学组分质量浓度的误差系数（Error fraction）。

当 Conc 小于等于 MDL 时，采用公式（3）计算；当 Conc 大于 MDL 时，采用公式（4）计算。EF 大小范围一般为 0.1~0.6，如颗粒物及化学组分质量浓度接近检出限或数据缺失较多时，则相应采用较大的数值。

6 基础计算

6.1 基础计算运行

6.1.1 数据的模型检验。通过 PMF 模型的浓度散点图、时间序列图等，检验不同化学组分之间关系、化学组分随时间变化情况。对于异常高或者低的化学组分数据，经核实没有特定污染过程（如沙尘天气等），可予以剔除。

6.1.2 数据计算权重的选择。颗粒物的关键标识组分（如 OC、EC、 SO_4^{2-} 、 NO_3^- 、 NH_4^+ 、Al、Si、Ca、K、Fe 等化学组分）的计算权重需设为“强”。其他非关键标识组分（如 Cl^- 、Na、Mg、Cu、Cd、Cr 等化学组分）的计算权重通常设为“强”，如测量不确定性较高，可设为“弱”计算权重。

6.1.3 因子个数的确定。因子个数需要基于对颗粒物可能来源、样品数量、采集时间等，通过模型反复计算确定。首先根据工作区域的基本情况（如当地排放源清单、污染源普查数据、工业布局、能源结构、机动车保有量等信息），初步确定因子个数的范围。然后通过设置不同的因子个数尝试计算，对各

计算情况进行分析，结合工作区域基本情况，验证和确定适合的因子个数，常用方法如下：

- 1) 根据 Q 值随因子个数变化关系判断。在设定合适的不确定度后，可逐渐增加因子数，观察 Q 值的变化。当因子数逐渐增加，而对应结果的 Q 值没有明显变化，则此时因子数增加为合理。当因子数增加到某一特定值（例如 p），Q 值开始有明显变化，则 p-1 则可能为合适的因子数。
- 2) 根据化学组分的加权残差大小判断。化学组分的加权残差（尤其是标识组分）应在正负 3 以内，如果加权残差有较大的分布，则应重新检查因子个数。
- 3) 根据多次计算结果的差异情况判断。多次运行不应该得到多个 Q 值或因子谱差异性比较大的结果（如 20 次运行的结果中，有多个结果之间差异较大），如有，则说明因子数的设定可能不合理。
- 4) 根据回归分析结果判断。将 G 矩阵与颗粒物浓度数据进行回归分析。如果回归系数出现负值，说明因子个数设置可能过多或存在共线性源；将 F 矩阵的每一个因子除以相应的回归系数后分别各自相加，如果大于 1.2，则说明选择的因子个数不合理。

6.2 基础计算的结果判断

6.2.1 对基础计算得到的多个计算结果，需进一步判断分析。首先，通过 Q 理论值（ Q_{Theo} ）和 Q 计算值（ Q_{True} ）的大小、是否收敛等，初步选择查看特定的计算结果。然后，从 Q 计算值与 Q 理论值差异分析、观测值/预测值差异分析、残差分析（Residual analysis）等对计算结果进一步判断。

6.2.2 Q 计算值与 Q 理论值越接近（85%~115%之间），表明计算结果更合理。

6.2.3 通过“观测值/预测值差异分析”，分析输入的观测值和计算得到的预测值的相关性。如果某化学组分的观测值和预测值之间有强相关性，则表明该化学组分计算较好；反之，则考虑是否应该降低该化学组分计算权重或从计算中排除。此外，还可以在时间序列上分析化学组分的观测值和预测值的差异。对于预测值明显高于观测值，需要进一步分析确定是否排除该数据。

6.2.4 通过“残差分析”，查看每一个化学组分的加权残差（通过不确定度加权）。比如，通过残差直方图中不同化学组分加权残差的百分比，判断各化学组分在基础计算的情况。如果残差直方图显示残差范围在（-3，+3）之间，则该化学组分的模型计算在数学意义上较好；如果化学组分具有许多较大的残差或显示非正常高斯分布曲线，则表示该化学组分的模型计算在数学意义上较差。

6.3 基础计算的结果分析

6.3.1 基础计算的结果主要包括因子谱 F 矩阵和因子贡献 G 矩阵。因子谱识别是 PMF 模型计算的关键，可通过分析不同因子谱中标识组分，将因子谱识别为具体的污染源类。颗粒物源类识别方法及主要源类的标识组分见附录 A。

6.3.2 当通过因子谱标识组分难以识别因子时，可分析因子贡献 G 矩阵的时间序列贡献情况，从而识别出在时间序列上有特征贡献变化的污染源类。

6.4 基础计算的误差评估

6.4.1 通过误差评估，分析计算结果的不确定性。结果误差评估方法主要有 Bootstrap (BS) 误差评估、Displacement (DISP) 误差评估、Bootstrap-Displacement (BS-DISP) 误差评估等。优先采用 BS 误差评估方法，评估模型计算结果的可靠性。

6.4.2 通过分析 BS 误差评估得到的因子和基础计算得到的因子匹配程度，评估因子中主要化学组分分配的合理性。如果两者匹配不到 80%，则这个因子的主要化学组分分配可能不当，需要通过调整因子数量或调整输入数据等提高匹配程度。

6.4.3 通过分析 BS 误差评估得到的 Q 修正值分布情况，评估基础计算中 Q 值的合理性。BS 计算可得到 Q 修正值的最小、最大、中位数，以及第 25 和第 75 百分位数值。基础计算得到的 Q 值一般要求

在第 25 和第 75 百分位数值之间。

6.4.4 通过分析 BS 误差评估得到的每个化学组分计算结果分布情况，评估化学组分计算结果的不确定性。比如一个化学组分的箱型图上下距（第 25 分位与第 75 百分位数）较宽（超过 20%），则表明该化学组分计算结果的不确定性可能较高；反之，则表明该计算结果的不确定性可能较低。

7 旋转计算

7.1 当模型基础计算得到的因子谱难以识别为实际源类的时候，可使用旋转计算（Rotational run），使因子谱中的标识组分更加突出，促进因子谱识别为具体污染源类。旋转计算包括 F 矩阵峰值模型计算（Fpeak model run）、约束模型计算（Constrained model run）。

7.2 F 矩阵峰值模型计算是常用的旋转计算方法，适用于标识组分的占比在各因子谱中的占比比较平均，不利于源类识别的情况。通过 F 矩阵峰值模型旋转计算，调整 F 矩阵中的各因子谱中化学组分占比，使其中的标识组分的占比更加突出。

7.3 F 矩阵峰值模型计算中，需要尝试不同的 Fpeak 参数。当 Fpeak 参数设为正值或负值时，F 矩阵中的化学组分在某些因子中占比会分别更加突出或趋向平均。查看不同 Fpeak 参数下 Q 值的变化情况，选择 Q 值变化在 5% 以内，并且 Q 值拐点之前 Fpeak 值对应的计算结果，参照“6.2、6.3、6.4”进行相应的结果判断、分析和误差评估。

7.4 PMF 共线性问题及解决方法。经过旋转计算后，如果提取的单个因子明显包含不同污染源类或子源类（如道路扬尘、施工扬尘等）的信息，则 PMF 模型计算出现了共线性问题，可进行二次解析。另外，可采用源成分谱约束的 PMF 模型进一步解析。

8 颗粒物源贡献计算

8.1 通过因子贡献 G 矩阵与颗粒物质量浓度进行多元线性拟合分析，将得到的回归系数乘以 G 矩阵各列，即为 F 矩阵识别的各污染源贡献。拟合公式如下：

$$PM_{2.5(j)} = \sum_{k=1}^p s_k g_{jk} \quad (5)$$

式中：PM_{2.5(j)}——第 j 个样品的颗粒物质量浓度；

s_k——回归系数；

g_{jk}——PMF 模型计算的 G 矩阵中第 k 源对第 j 个样品的归一化贡献；

p——污染源个数。

8.2 多个环境受体监测点位的模型计算。将所有受体监测点位数据一起输入 PMF 模型，计算得到该工作区域的总体源贡献结果；通过对其中特定点位全部样品的源贡献结果进行平均，得到该点位的源贡献结果。

8.3 填写 PMF 模型计算记录表，并附上参与计算的全部数据。PMF 模型计算记录表格式参见附录 B。

9 结果合理性判断

9.1 结果需符合模型计算要求，即各化学组分的模型计算结果与实测结果接近。可通过直接比较颗粒物及化学组分质量浓度计算值与实测值、分析主要化学组分或标识组分的残差、对比 Q 计算值与 Q 理论值等方法，判断模型计算结果的好坏。如果计算值与实测值越接近、化学组分残差越小、Q 计算值与 Q 理论值越接近，则表明计算结果更合理。

9.2 结果需符合当地实际情况，即经过识别的源成分谱特征及其源贡献要相对合理。可通过 PMF 模型计算得到的源成分谱与实测源成分谱相关性比较、PMF 模型计算的平均源贡献排序与 CMB 模型 (Chemical mass balance model) 计算结果一致性比较、源贡献变化趋势分析等方法来评估模型计算结果的合理性。此外，还可通过其他监测结果 (比如 PM_{2.5} 和 PM₁₀ 的浓度比值，自动监测的 SO₂ 和 NO_x 浓度比值，自动监测的 O₃、CO 等污染物浓度等) 或污染源信息等来辅助判断结果的合理性。



附录 A

(资料性附录)

颗粒物源类识别方法及主要源类的标识组分

可通过以下方法，将 PMF 模型计算得到的因子谱识别为具体的颗粒物源类：

A.1 根据各种源类的主要标识组分来识别。某个因子谱的标识组分与其他因子谱相比占比较高，则可识别为相应的源类。例如，某因子谱中 NH_4^+ 、 SO_4^{2-} 的占比分别大于 30%、50%，高于其他因子谱中 NH_4^+ 、 SO_4^{2-} 的占比，则该因子谱可识别为二次硫酸盐；某因子谱中 Ca、Si、Al 的占比最高，可识别为扬尘。常见源类的标识组分见表 A.1。

A.2 根据源类排放特有的时间变化特征来识别。某些源类排放具有显著时间变化特征，比如机动车源在早晚高峰时段贡献大，季节变化不明显；北方民用燃煤源存在明显的季节变化，冬季采暖期贡献大；扬尘在春季贡献大，在北方城市尤为突出；生物质燃烧源在农作物收割季节贡献大；烟花爆竹燃放源在特殊节假日（如春节期间）贡献大。如果在 PMF 模型计算中某个因子谱具有显著时间变化特征，则可根据该特征识别为相应的源类。

A.3 根据当地监测得到的主要源类源成分谱来识别。通过当地污染源采样和分析，构建当地主要的颗粒物源成分谱。将 PMF 模型计算得到的因子谱与实测源成分谱进行对照识别。

A.4 根据 PMF 模型进一步优化计算得到的因子谱来识别。当基础计算的结果难以识别为具体源类时，可使用源谱约束的 PMF 模型进行优化计算，突出因子谱中的标识组分，从而将因子谱识别为具体源类。

表 A.1 颗粒物主要源类的参考标识组分

一级源类	二级源类	无机特征组分	有机特征组分
燃煤源	电厂燃煤源、供热燃煤源、工业燃煤源、民用燃煤源等	EC、Al、As、Ca（或 Ca^{2+} ）、 SO_4^{2-} 、Cl ⁻	OC、藿烷、甾烷、烷基芘、多环芳烃
移动源	汽油车源、柴油车源等	EC、Cu、Zn	OC、藿烷、甾烷、晕苯、荧蒽、芘
扬尘	混合扬尘	Si、Al、Ca（或 Ca^{2+} ）	/
	施工扬尘	Ca（或 Ca^{2+} ）	/
	土壤扬尘	Si	/
工业工艺过程源	钢铁源	Fe、Mn、Cr、Ni	/
其他一次源	生物质燃烧源	K（或 K^+ ）、EC	OC、左旋葡聚糖、植物甾醇、萜类物质
	餐饮源	/	OC、胆固醇、十六烷酸、十八烷酸、豆甾醇、 β -谷甾醇、壬醛、9-十六烯酸
	海盐	Na（或 Na^+ ）、Mg（或 Mg^{2+} ）、Cl ⁻	/
	烟花爆竹燃放源	Sr、K（或 K^+ ）、Ba、Cu、Mg（或 Mg^{2+} ）	OC
二次源类	二次硫酸盐	SO_4^{2-} 、 NH_4^+	/
	二次硝酸盐	NO_3^- 、 NH_4^+	/
	二次有机物	/	OC、乙二酸、有机硫酸酯

附 录 B
(资料性附录)
PMF 模型计算记录表

表 B.1 PMF 模型计算记录表 (一)

项目名称:

模型版本:

方法依据:

输入数据 (见附件)	采样地点:	采样日期和时段:	颗粒物粒径:
	样品数量 (行):	化学组分数量 (列):	数据剔除情况:
	不确定度计算方法:		
	化学组分的计算权重:		
基础计算	因子个数:	计算次数:	种子数 (Seed) 值设定:
	结果误差评估:		
旋转计算	参数设定:		
	结果误差评估:		
输出记录	选择的计算结果:	Q 值:	是否收敛:
	源类识别		
	因子 1:	因子 4:	因子 7:
	因子 2:	因子 5:	因子 8:
	因子 3:	因子 6:	因子 9:
备注			

记录人:
年 月 日

校核人:
年 月 日

审核人:
年 月 日

表 B.3 PMF 模型计算记录表（三）

因子贡献（单位： ）										
时间	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6	因子 6	因子 8	因子 9	
备注										

记录人：
年 月 日

校核人：
年 月 日

审核人：
年 月 日