

# T/SDEPI

团 体 标 准

T/SDEPI 044—2024

## 土壤重金属来源解析 正定矩阵因子分解法 方法指南

Guide for source apportionment of soil heavy metals by Positive  
Matrix Factorization method

2024-04-19 发布

2024-04-19 实施

山东省环境保护产业协会 发布

## 目 次

前言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语及定义.....	1
4 土壤重金属来源解析工作流程.....	2
5 土壤重金属样品采集.....	2
6 样品分析方法选择.....	2
7 模型原理和计算流程.....	3
8 分析计算.....	4

## 前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由山东省环境保护产业协会提出并归口。

本文件起草单位：山东大学、山东建筑大学、新汶矿业集团有限责任公司、济南实华工程咨询有限公司、山东泉涌环境科技有限公司、山东鲁金环境工程有限公司。

本文件主要起草人：崔兆杰、崔晓玮、任丽军、蔺琨、何竞宇、薛文秀、刘雷、潘哲、单绍磊、田伟、张程伟、刘佳、潘汝东、张琦、张建、贺波。

# 土壤重金属来源解析 正定矩阵因子分解法方法指南

## 1 范围

本文件规定了具体方法模型原理、数据准备与计算、结果合理性判断等内容。  
本文件适用于土壤重金属来源解析，包括农业用地和建设用地土壤等。

## 2 规范性引用文件

本文件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方法体系；凡是未注明日期的引用文件，其最新版本（包括所有的修改单）适用于本方法体系。

GB/T 17136	土壤质量 总汞的测定 冷原子吸收分光光度法
GB/T 17138	土壤质量 铜、锌的测定 火焰原子吸收分光光度法
GB/T 17139	土壤质量 镍的测定 火焰原子吸收分光光度法
GB/T 17141	土壤质量 铅、镉的测定 石墨炉原子吸收分光光度法
GB/T 22105	土壤质量 总汞、总砷、总铅的测定 原子荧光法
GB/T 36197	土壤质量 土壤采样技术指南
HJ 25.1	建设用地土壤污染状况调查技术导则
HJ 25.2	建设用地土壤污染风险管控和修复监测技术导则
HJ/T 166	土壤环境检测技术规范
HJ 491	土壤和沉积物 铜、锌、铅、镍、铬的测定 火焰原子吸收分光光度法
HJ 680	土壤和沉积物 汞、砷、硒、铋、锑的测定 微波消解/原子荧光法
HJ 780	土壤和沉积物 无机元素的测定 波长色散 X 射线荧光谱法
HJ 784	土壤和沉积物 多环芳烃的测定 高效液相色谱法
HJ 803	土壤和沉积物 12 种金属元素的测定 王水提取-电感耦合等离子体质谱法
HJ 923	土壤和沉积物 总汞的测定 催化热解-冷原子吸收分光光度法

## 3 术语及定义

下列术语和定义适用于本文件。

### 3.1

**正定矩阵因子分解模型** Positive Matrix Factorization (PMF) Model

是一种用于分析样品中各种组分来源以及组成的受体模型。它从样品数据中提取若干因子，利用标识组分将因子识别为不同的源类，再通过多元线性回归计算不同源类对样品的贡献。

### 3.2

**土壤重金属源成分谱** Source Chemical Profile of heavy metals in soil

简称土壤重金属源谱，它是指各类污染土壤中各种重金属元素所占的比例，通常以质量百分比的形式表示。

### 3.3

**土壤重金属源贡献** Source Contribution of heavy metals in soil

是指土壤样品中各种重金属元素由不同污染源所贡献的量，可用贡献分担率（%）表示。

### 3.4

#### 土壤重金属源标识组分 Tracer Species of heavy metals in soil

是指土壤重金属源谱中具有指示作用的一种或多种重金属元素，又称示踪组分。

## 4 土壤重金属来源解析工作流程

土壤重金属来源解析工作流程见下图。

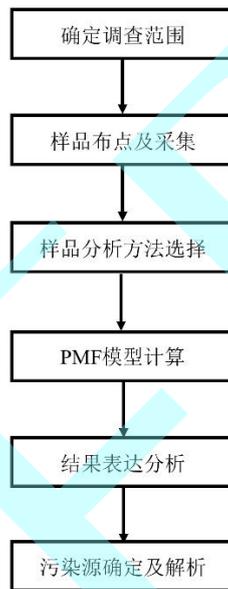


图 1 土壤重金属来源解析工作流程图

## 5 土壤重金属样品采集

土壤重金属样品的采集工作参照 GB/T 36197-2018 和 HJ/T 166-2004 进行，样品采集数量不少于 16 个点。土壤样品调查与监测执行 HJ 25.1 和 HJ 25.2 的相关规定。

## 6 样品分析方法选择

土壤重金属的检测方法见表 1。

表 1 土壤重金属元素检测方法

重金属元素	标准编号
镉	GB/T 17141
	HJ 1315
汞	HJ 680
	GB/T 22105.1

重金属元素	标准编号
	GB/T 17136
	HJ 923
砷	HJ 1315
	HJ 680
	GB/T 22105.2
铜	GB/T 17138
	HJ 491
	HJ 780
铅	GB/T 17141
	HJ 491
	HJ 780
铬	HJ 491
	HJ 780
锌	GB/T 17138
	HJ 491
	HJ 780
镍	GB/T 17139
	HJ 491
	HJ 780
钴	HJ 1315

## 7 模型原理和计算流程

### 7.1 模型原理

正定矩阵因子分解法（以下简称 PMF）是一种基于统计学原理的土壤重金属来源解析方法。该方法假设污染土壤中重金属的含量由多种来源混合而成，并根据样品中重金属含量的相关性，将其分解为多个因子，每个因子代表一种重金属源。

应用 PMF 模型进行来源解析的原理是将受体重金属元素浓度矩阵  $X_{(n \times m)}$  因子化，分解为两个矩阵  $F_{(p \times m)}$  和  $G_{(n \times p)}$ ，以及一个残差矩阵  $E_{(n \times m)}$ ，具体如下式所示：

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

式中：

$G_{(n \times p)}$ ——因子贡献矩阵；

$F_{(p \times m)}$ ——因子源谱矩阵；

$n$ ——样品个数；

$m$ ——重金属种类数；

p——解析出的因子（污染源）数目。

PMF 模型限定矩阵 G 和 F 中的组分都是正值，即非负限制。PMF 模型解析上述矩阵的方法是通过定义一个“目标函数（Object function）” Q，并使这个目标函数的值最小；当目标函数 Q 值最小时，模型将受体浓度矩阵 X 分解成 G 矩阵（因子贡献矩阵）和 F 矩阵（因子源谱矩阵）。

$$Q_{(E)} = \sum_{i=1}^m \sum_{j=1}^n \left( \frac{E_{ij}}{\sigma_{ij}} \right)^2 \dots\dots\dots (2)$$

式中：

$\sigma_{ij}$ ——第 j 个样品中第 i 个重金属的不确定度。

## 7.2 计算流程

应用 PMF 模型进行来源解析的计算流程包括数据准备、数据导入和数据检验、基础计算及结果分析、旋转计算及结果分析、模型计算结果评估等环节。开展 PMF 模型计算时，需要通过反复优化计算参数，得到相对较优结果的过程。在数据检验中，优化输入数据。在基础计算和旋转计算中，需要经过不断尝试计算，优化调整计算参数，才能得到相对较优的计算结果。

## 8 分析计算

### 8.1 数据准备

#### 8.1.1 数据质量要求

用于土壤重金属源解析工作的监测数据，应采用国家、行业或地方方法标准获得；没有国家、行业或地方方法标准，可采用国际标准、国外标准或研究建立的统一监测方法。

监测数据应采用法定计量单位，有效数字位数根据计量仪器的精度和分析仪器的示值确定，不得随意添加或删除。对于采样、运输、存储、分析失误所造成的离群数据应予以剔除。低于方法检出限的数据以“未检出”报出，并注明检出限值数值。

#### 8.1.2 数据格式要求

PMF 模型需要输入两个数据文件：（1）重金属元素质量浓度的数据文件，（2）重金属元素质量浓度不确定度的数据文件。以 EPA PMF 5.0 软件为例，导入数据文件格式可以为 CSV、XLS、XLSX 等，两个数据文件内容需要对应，第一行为组分名称，组分名称必须唯一，中间不能含有逗号；其余内容为具体数据，数据格式应符合模型软件的要求。

#### 8.1.3 不确定度的计算

不确定度直接影响输入 PMF 模型中样品重金属质量浓度及组分浓度数据的计算权重，可根据以下公式进行计算。

$$U_{nc} = \frac{5}{6} MDL \dots\dots\dots (3)$$

$$U_{nc} = \sqrt{(EF \times Conc)^2 + (0.5 \times MDL)^2} \dots\dots\dots (4)$$

式中：

$U_{nc}$ ——重金属元素的不确定度，mg/kg；

*MDL*——重金属元素方法检出限, *mg/kg*;

*Conc*——重金属元素的质量浓度, *mg/kg*;

*EF*(*Error fraction*)——误差系数。

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。误差系数的大小范围一般为 0.1-0.6。对于重金属质量浓度或者组分质量浓度不稳定或者接近检出限, 采用较大的误差系数。当重金属质量浓度或者组分质量浓度数据缺失较多时, 可设定较大的误差系数。

## 8.2 基础计算数据准备

### 8.2.1 基础计算运行

#### (1) 数据模型检验

通过 PMF 模型的浓度散点图、时间序列图等, 检验不同组分之间关系、组分随时间变化情况。对于异常高或者低的组分数据, 经核实没有特定污染过程 (如附近有特定污染源等), 可予以剔除。

#### (2) 数据计算权重的选择

重金属关键标识组分的计算权重须设置为“强”(Strong), 其它非关键组分通常也设置为“强”, 但如果测量的不确定度较高, 可设置为“弱”(Weak) 计算权重。

#### (3) 因子个数的确定

因子个数需要基于对重金属可能来源、样品数量、采集时间等基本认识以及模型反复计算来综合确定。一般会根据采样场地周边可能存在的排放源清单、污染源普查数据、工业布局、能源结构等信息, 初步确定因子的个数范围。

#### (4) 尝试性计算

通过设置不同的因子个数进行尝试性计算, 对各计算情况进行分析, 结合采样场地基本情况, 验证和确定适合的因子个数。根据模型计算确认因子个数的常用方法如下:

a) 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, 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 可能为适合的因子数。

b) 根据组分的残差大小判断。组分的残差 (尤其是标识组分) 应在  $\pm 3$  以内, 如果残差有较大的分布, 应重新检查因子个数。

c) 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 个结果中, 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。

d) 根据回归检验结果判断。将 G 矩阵与浓度数据进行回归检验。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源; F 矩阵的每一个因子除以相应的回归系数后, 如果加和大于 1.2, 说明选择的因子过少。

### 8.2.2 基础计算结果判断

Q 值有三种类型: Q 理论值 ( $Q_{\text{Theo}}$ )、Q 计算值 ( $Q_{\text{True}}$ )、Q 修正值 ( $Q_{\text{Robust}}$ )。

对基础计算得到的多个计算结果, 需进一步判断分析。通过 Q 理论值 ( $Q_{\text{Theo}}$ )、Q 计算值 ( $Q_{\text{True}}$ )、是否收敛 (Converge) 等, 初步选择查看特定的计算结果。从 Q 计算值 ( $Q_{\text{True}}$ ) 与 Q 理论值 ( $Q_{\text{Theo}}$ ) 差异分析、观测值预测值差异分析、残差分析 (Residual analysis) 等对计算结果进行判断。Q 计算值 ( $Q_{\text{True}}$ ) 与 Q 理论值 ( $Q_{\text{Theo}}$ ) 越接近 (85%-115%之间), 表明模拟计算结果更合理。

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。如果某组分的观测值和预测值之间有强相关性, 则表明该组分计算较好; 反之, 则考虑是否应该降低该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组分的输入值和预测值的差异。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。

通过“Residual Analysis（残差分析）”，查看每一个组分的加权残差（通过不确定度加权）。比如残差直方图中不同组分加权残差的百分比，判断各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，则该组分的模型计算在数学意义上较好；如果组分具有许多较大的残差或显示非正常曲线，则表示该组分在数学意义上计算较差。

### 8.2.3 基础计算结果分析

基础计算的结果主要包括因子源谱矩阵（F）和因子贡献矩阵（G）。因子源谱展示了不同化学组分在各因子中的占比及浓度。因子贡献显示了不同因子对各组分的平均贡献及时间序列贡献（点线图）。

因子的识别是 PMF 模型计算的关键，可以通过分析不同因子谱中标识组分来将因子识别为具体的污染源类。

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

### 8.2.4 基础计算误差评估

通过误差评估，确定计算结果的不确定度，从而判断源解析结果的准确性。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。推荐采用 BS 误差评估方法评估模型计算结果的可靠性。通过分析多次 BS 计算得到的因子和基础计算得到的因子匹配程度、Q 值的分布情况、各化学组分计算结果分布情况等，评估结果的可靠性。

(1) 在 BS 误差评估中，通过分析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%，则这个因子的主要组分分配可能不当，需要通过调整因子数量或提高数据质量等提高匹配程度。

(2) 在 BS 误差评估中，通过 Q 修正值的分布情况，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。

(3) 在 BS 误差评估中，通过分析每个化学组分计算结果的分布情况，评估化学组分计算结果的不确定度。比如一个化学组分的箱型图上下距（25 分位数与 75 分位数）较宽（超过 20%），则表明该化学组分计算结果的不确定度可能较高；如果一个化学组分的箱型图上下距较窄（小于 20%），则说明该化学组分模型计算结果的不确定度可能较低，模型计算结果较好。

## 8.3 旋转计算

当模型基础计算得到的因子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），其中前者是常用的旋转计算方法。

### 8.3.1 F 矩阵峰值模型计算

F 矩阵峰值模型计算适用于标识组分的占比不够突出（某一标识组分在各因子中的占比比较平均），不利于源类识别的情况。通过 F 矩阵峰值旋转计算，调整 F 矩阵中的各因子中化学组分占比，从而使因子中的标识组分的占比更加突出。

F 矩阵峰值模型计算中，需要尝试不同的 Fpeak 参数。当 Fpeak 参数设为正值时，F 矩阵中的化学组分在某些因子中占比会更加突出；Fpeak 参数设为负值时，F 矩阵中的化学组分在各因子中的占比则相对较平均。查看 Q 值在不同 Fpeak 参数下的变化情况，要求 Q 值的变化在 5% 以内，并且选择 Q 值拐点之前的 Fpeak 值对应的解析结果，参照前文进行相应的结果判断、分析和误差评估。

### 8.3.2 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的信息，则 PMF 模型计算出现了共线性问题。可采用一些约束的 PMF 模型（如：多元线性引擎 2-组分比值约束模型（ME2-SR）、偏目标转换-正定矩阵因子分析模型（PTT-PMF）等）进一步解析。

### 8.4 重金属源贡献计算

在 PMF 模型计算中输入数据包含重金属元素质量浓度，可直接计算得到污染源对重金属的源贡献结果。还可以将计算得到的 G 矩阵（源贡献矩阵）与重金属元素质量浓度做多元线性拟合分析，计算出 G 矩阵各列（源）的拟合系数。将各列乘以拟合系数，即为各污染源对重金属的贡献，拟合公式如下：

$$C_{(j)} = \sum_{k=1}^p s_k g_{jk} \dots\dots\dots (5)$$

式中：

$C_{(j)}$ ——第j个样品的重金属质量浓度；

p——源个数；

$s_k$ ——回归系数；

$g_{jk}$ ——PMF模型计算的G矩阵中第k个源对第j个样品的贡献；

### 8.5 结果合理性判断

源解析结果既要符合模型计算要求，还需要符合实际情况。

符合模型计算要求是模型模拟计算出的各组分浓度与实测测试结果接近。可通过直接比较重金属和化学组分模拟值与实测值、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，判断模型模拟结果的好坏。如果模拟值与实测值越接近、化学组分残差越小、Q计算值与Q理论值越接近，则表明模拟计算结果更合理。

符合实际情况要求是经过识别的源谱特征及其源贡献合理。可通过PMF模型计算得到的源谱与实测源谱相关性比较、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、源贡献变化趋势及其日变化趋势分析等方法来评估模型计算结果的合理性。此外，还需要通过当地其它监测结果或污染源信息等来辅助判断结果的合理性。

### 8.6 结果表达

#### （1）计算记录表

完成 PMF 模型的计算后，应将计算结果参照表 2 进行记录，表中应详细列出各个污染源的贡献率、源成分谱等信息，并附上参与计算的全部数据。

表 2 PMF 模型计算记录表

输入 数据	采样地点：	采样日期和时间：	土壤类型：
	样品数量：	重金属元素数量：	数据剔除情况：
	不确定度计算方法：		
	重金属元素权重：		
基础 计算	因子个数：	计算次数：	Seed 值：
	结果误差评定：		

旋转 计算	参数设定:		
	结果误差评定:		
输出 记录	Q 值:	是否收敛:	
	源类识别		
	因子 1:	因子 2:	因子 3:
	因子 4:	因子 5:	因子 6:
	因子 7:	因子 8:	因子 9:
备注			

### (2) 堆叠柱状图、弦图等

除计算记录表外，还可以通过堆叠柱状图、弦图等展示各个污染源对重金属浓度的贡献，或者通过柱状图或饼图展示污染源中各种重金属的相对比例。