

《土壤有机物来源解析技术指南》

编制说明

《土壤有机物来源解析技术指南》编制组

2023年11月

目 录

目 录.....	I
1 项目背景.....	1
1.1 任务来源.....	1
1.2 标准现状.....	1
1.3 工作过程.....	1
2 方法指南制定的必要性分析.....	2
3 PCA 方法模型比较.....	3
4 制定方法指南的基本原则与技术路线.....	5
4.1 基本原则.....	5
4.2 技术路线.....	5
5 标准主要内容.....	6
6 主要条文.....	7
6.1 适用范围.....	7
6.2 规范性引用文件.....	7
6.3 术语和定义.....	8
6.4 土壤有机物污染样品采集.....	8
6.5 主成分分析模型方法.....	9
6.6 结果表达分析.....	14
7 主成分分析模型应用.....	15
7.1 导入数据.....	15
7.2 选择变量.....	15
7.3 统计量设置.....	16
7.4 主成分抽取设置.....	16
7.5 旋转方式设置.....	17
7.6 得分保存设置.....	17
7.7 选项设置.....	18
7.8 输出分析结果.....	18
7.9 分析结果解读.....	19

1 项目背景

1.1 任务来源

为全面贯彻《中共中央国务院关于加快推进生态文明建设的意见》，落实《土壤污染防治行动计划》，推进我国土壤有机物污染来源解析工作的进程，以便更有效地开展有机物污染溯源工作，依托山东省重点科研项目《矿区土壤生态修复与大宗固废高值利用关键技术及工程示范》中的《矿区土壤有机物重金属复合污染控制技术》课题，由山东大学起草编制了《土壤有机物来源解析技术指南》（以下简称“指南”）。

1.2 标准现状

土壤有机物污染是全球性环境问题，严重威胁人体健康和生态环境。目前对环境空气臭氧、大气颗粒物和地下水污染同位素等来源解析技术指南，尚未有土壤有机物来源解析技术指南。为深入了解土壤有机物污染来源，有效开展污染防治工作，需要建立科学有效的土壤有机物污染来源解析方法体系。

1.3 工作过程

(1) 方法调研

2019年-2020年编制单位开始进行广泛的调研工作，以了解当前土壤有机物来源解析方法的研究现状和应用情况。通过查阅相关文献和与领域内的专家进行交流，编制单位对各种方法的优缺点有了全面的了解。同时还参观了一些先进的实验室和研究机构，亲身体会他们的工作流程和技术设备，以为后续的工作提供了重要的参考和借鉴。

(2) 成立工作组

在充分了解现有方法的基础上，2020年山东大学环境科学与工程学院牵头组建了标准编制工作组。

(3) 标准研制

2021年由工作组对标准进行研制工作，讨论制定一套科学、可行的土壤有

机物来源解析方法标准。经过反复的修改和完善，最终形成了一份初步的标准草案。

(4) 应用验证

在标准草案完成后，编制单位开始了对该指南的应用验证工作。编制单位选择了一些有代表性的土壤样本进行分析测试，并严格按照标准中的要求进行操作。通过对实验结果的统计和分析，评估了该指南的准确性和可行性。同时，编制单位还与相关部门和企业进行了合作，将该指南应用于实际案例中，进一步验证其实用性和可靠性。

(5) 立项

在应用验证工作取得初步成果后，编制单位在 2023 年向山东省环境保护产业协会正式申请立项，并将该指南提交给专家组进行评审。专家组对该指南提出了一些宝贵的修改意见和建议。根据专家们的意见，编制单位对方法指南进行了进一步的修改和完善，确保其科学性和可操作性。

(6) 修改完善

在前期工作的基础上，编制单位对方法指南的初稿进行了全面的修改和完善，同时还增加了一些实例和案例分析，以便读者更好地理解和应用该方法指南。经过多次的修改和审校，最终形成了一份完整的《土壤有机物来源解析技术指南》。

2 方法指南制定的必要性分析

有机污染物是我国工业污染场地污染土壤中的主要污染物类型之一。工业场地中有机污染物具有来源广泛、种类繁多、迁移性强和污染持久等特点。常见的有机污染物包括总石油烃、氯代烃、苯系物、多环芳烃(PAHs)、农药、多氯联苯、二噁英类及其他有机污染物；场地有机污染类型包括有机污染物单一污染和有机污染物复合污染。2021 年专家骆永明等人以有机污染场地为研究对象，调查分析了全国 277 个有机污染场地的行业类别、污染类型以及有机污染物种类分布情况，发现主要贡献行业为化学原料及化学品制造行业，贡献率为 37.9%；污染类型中以各类有机污染物复合污染为主；除了不同有机污染物之间的相互复合，有机污染物与重金属复合污染特征明显，占总场地 58.5%。有机污染物种类分布中多环芳烃类污染场地最多，占 54.9%，与欧美有机污染场地中氯代烃类污染场地占比最多不同；其次为总石油烃和苯系物，占比分别为 49.5% 和 36.8%。为

解决土壤有机物污染问题，有效控制或消除有机物污染，需解析土壤有机物污染源并加以控制。

土壤有机污染物来源解析技术能够为相关防治工作提供关键的污染源组成信息，能够评估防治工作的有效性，使防治“有的放矢”、减少防治工作的盲目性，避免“一刀切”。但是土壤有机物来源解析的技术手段多样，来源复杂，污染源种类众多，受体模型法作为我国有机污染物来源解析监测技术的主要方法，仍存在一些问题，如采样、样品保存的规范问题；分析仪器统一管理质控的问题；数据检验与处理的规范化问题等。因此，有必要进一步规范源解析的方法和流程，提高其科学性和可行性。本指南的编制，可以有效提高土壤有机物来源解析数据成果的科学性、规范性和准确性，提升环境监测水平和能力，具有重要意义。

为完善和统一土壤有机物污染来源解析方法和标准，综合运用方法和技术形成土壤有机物来源解析技术规范。制定《土壤有机物来源解析技术指南》可以为相关管理部门提供支撑，帮助他们建立起科学、高效的管理体系。通过规范化的管理流程和方法，可以更好地监测、评估和控制土壤污染，保护生态环境和人类健康。该指南还可以为决策者提供科学依据，促进土壤环境保护政策的制定和实施。

3 PCA 方法模型比较

经过多年的发展和研究，许多源分配方法已经成功应用于污染源的分类和分布，如主成分分析（PCA）、正矩阵分解（PMF）、绝对主成分得分-多元线性回归法（APCS-MLR）和 UNMIX 模型。

PMF 即正交矩阵因子分解法（Positive Matrix Factorization）是由 Paatero 和 Tapper 在 1993 年提出的一种有效的数据分析方法。与其他源解析方法相比，具有不需要测量源成分谱，分解矩阵中元素分担率为非负值，可以利用数据标准偏差来进行优化，并且可处理遗漏数据和不精确数据等特点。但该方法没有直观的因字数显示，需要设置多个因字数进行多次运算以获得目标矩阵 Q 的最小值。

绝对因子得分—多元线性回归法（APCS-MLR），Thurston 和 Spengler 在 1985 年提出了绝对因子得分（APCS）的概念，将主成分分析与多元线性回归相结合，从而同时实现了定性和定量解析各个源类对受体的贡献值。APCS/MLR 模型分析的方法具有一定的局限性，结果可能会产生有歧义的情况，使因子分析问题无法正常解决。该模型需要基于大量的受体数据进行分析，经过方差最大正

交旋转，提取出诸多因子。但在强加限制条件下，如源成分谱和源贡献的非负处理，可以减少这种歧义。

UNMIX 是美国环境保护局推荐使用的一种先进的受体模型。其具有可视化的图形界面，同时结果中可以提供图形诊断工具。UNMIX 模型不需要事先测定源成分谱的数据信息，从而避免了成分谱采集工作难度高、工作量大的缺点；但 UNMIX 模型的缺点不容忽视，它需要大量的数据样本，在选取组分纳入模型计算时，要求尽量选取缺失值少的组分（缺失值小于三分之二的样本）或用检出限的一半来替代组分浓度值，数据的质量对模型结果有一定的影响。

然而上述方法需要大量的样品，并且它们仅有效地识别一般的污染源，例如农业源、工业源和自然源，在准确识别每个污染源的确切来源及其影响方面受到限制。主成分分析，也称为主分量分析，是由 Pearson 在 1901 年对非随机变量引入的一种统计方法，后由 Hotelling 将此方法推广到随机向量的情形。PCA 是一种常用的源解析方法，可以有效地对数据集维度进行降维，提高可解释性，并将信息丢失降到最低。此外，PCA 模型经其他单位应用验证，证明使用该模型得出的结果更优。因此，本指南使用 PCA 模型对土壤有机物来源进行解析，PCA 模型和其他方法模型比较如表 1 所示。

表 1 方法比较

模型名称	优势	局限性
主成分分析法 (PCA)	来源的定性识别结果相对可靠，操作简单	识别源数量有限，并非直接针对数据本身，而是针对数据产生的偏差进行分析；无法给出定量的源贡献
正定因子矩阵模型 (PMF)	添加不确定性数据对数据进行逐点分析；不需要详细的源成分谱信息，可较好处理缺失及不精确的数据；对因子分解矩阵进行非负约束，使得到的源成分谱和源贡献率不会出现负值，结果更加合理	没有直观的因字数显示；需要设置多个因字数进行多次运算以获得目标矩阵 Q 的最小值
APCS-MLR 模型	解决了因子分析的不足之处	不能准确预测污染源的指纹谱图；采样量大
UNMIX 模型	克服了模型给出负值源贡献的缺	样本量大；需要数据质量较高(高

模型名称	优势	局限性
	点；从受体出发快速获取因子识别污染源	于检出限)；缺失值较少($\leq 75\%$)的数据；难以识别有相似贡献的源

4 制定方法指南的基本原则与技术路线

4.1 基本原则

科学性原则：所制定的方法指南能够满足源解析工作的特定技术需求，所选择的技术方法要科学合理。

适用性原则：所制定的方法指南要能够在不同地区的源解析工作中使用，在技术上可行，具有可操作性。

规范性原则：所制定的方法指南对关键技术指标进行统一规定，提高源解析结果的可比性。

4.2 技术路线

主要技术路线如下图：

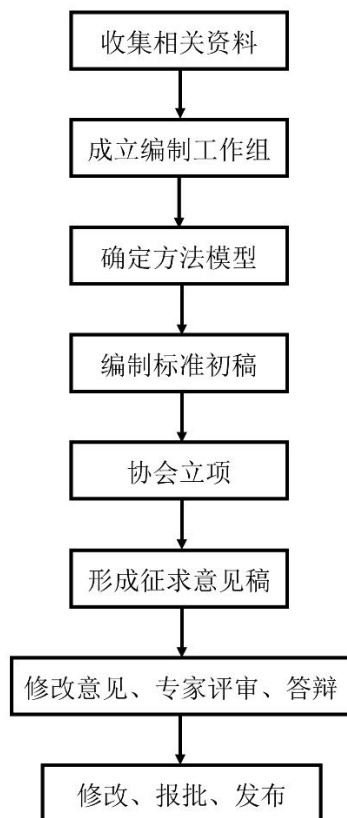


图 1 技术路线图

5 标准主要内容

本标准的主要技术内容及框架如下：

- 1) 适用范围
- 2) 规范性引用文件
- 3) 术语和定义
- 4) 土壤有机物污染样品采集
- 5) 主成分分析模型方法
- 6) 结果表达分析

6 主要条文

6.1 适用范围

本指南适用于指导耕地、林地、工矿用地、交通用地等开展土壤有机物来源解析工作。

6.2 规范性引用文件

本指南内容引用了下列文件中的条款。凡是未注明日期的引用文件，其有效版本适用于本指南。

GB/T 18834 土壤环境 词汇

GB/T 14550 土壤中六六六和滴滴涕测定的气相色谱法

GB/T 36197 土壤质量 土壤采样技术指南

GB 37822 挥发性有机物无组织排放控制标准

HJ/T 166 土壤环境监测技术规范

HJ 605 土壤和沉积物 挥发性有机物的测定 吹扫捕集气相色谱/质谱法

HJ 642 土壤和沉积物 挥发性有机物的测定 顶空/气相色谱/质谱法

HJ 25.1 场地环境调查技术导则

HJ 25.2 场地环境监测技术导则

HJ 703 土壤和沉积物 酚类化合物的测定 气相色谱法

HJ 743 土壤和沉积物 多氯联苯的测定 气相色谱-质谱法

HJ 784 土壤和沉积物 多环芳烃的测定 高效液相色谱法

HJ 805 土壤和沉积物 多环芳烃的测定 气相色谱-质谱法

HJ 834 土壤和沉积物 半挥发性有机物的测定气相色谱-质谱法

HJ 911 土壤和沉积物 有机物的提取 超声波萃取法

HJ 1019 地块土壤和地下水中挥发性有机物采样技术导则

6.3 术语和定义

土壤有机物 Soil Organic Matter

土壤有机物是指土壤中以碳水化合物、蛋白质、氨基酸以及脂肪等形式存在的天然有机物质及人工合成有机物质，包括挥发性有机物、多环芳烃类和石油烃类等。

主成分分析 Principal Component Analysis (PCA)

主成分分析是一种数据降维的方法，通常用于通过将数量很多的变量转换为仍包含集合中大部分信息的较少变量来降低数据集的维数。

土壤有机物方差贡献率 Variance Contribution Rate of organic matter in soil

土壤有机物方差贡献率是指在土壤有机物中，一个主成分所能够解释的方差占全部方差的比例，这个值越大，说明土壤有机物的主成分综合原始变量信息的能力越强。

6.4 土壤有机物污染样品采集

6.4.1 布点

宜采用 HJ/T 166 土壤环境监测技术规范进行布点。

6.4.2 样品采集

宜采用 GB/T 36197 土壤质量 土壤采样技术指南和 HJ 1019 地块土壤和地下水中挥发性有机物采样技术导则进行样品采集和贮存，土壤样品的保存和流转执行 HJ 25.1、HJ 25.2 和 HJ/T 166 的相关规定。

6.4.3 样品检测

土壤样品有机物测试项目及方法见表 2。

表 2 有机物检测方法及相关标准

测试项目	检测方法	参考标准
1,2-二氯乙烷	吹扫捕集/气相色谱质谱法	HJ 605

二氯甲烷	顶空/气相色谱-质谱法	HJ 642
四氯乙烯		
苯		
甲苯		
三氯氟甲烷		
2-甲基萘	气相色谱-质谱法	HJ 834
萘	高效液相色谱法 气相色谱-质谱法	HJ 784 HJ 805
芴		
菲		
蒽		
荧蒽		
芘		
苯并(a)蒽		
蒾		
苯并(k)荧蒽		
苯并(g,h,i)花		
苯酚		
石油烃 (C ₁₀ -C ₄₀)	气相色谱法	HJ 1021

6.5 主成分分析模型方法

6.5.1 模型原理

主成分分析（PCA）是一种考察多个变量间相关性的降维统计方法，其原理是设法将原来变量重新组合成一组新的互相无关的几个综合变量，同时根据实际需要从中可以取出几个较少的综合变量尽可能多地反映原来变量的信息的统计方法。

主成分分析法是一种数学变换的方法，把给定的一组相关变量通过线性变换转化为另一组不相关的变量，这些新的变量按照方差依次递减的顺序排列。假设对 p 个变量进行了 n 次测量/观测，那么原始测量数据的矩阵表示如下：

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

其中, x_i 表示如下:

$$x_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}, i = 1, 2, \dots, p$$

引入一个新变量 F , 是原来 P 个变量的线性组合, 表示如下:

$$F_1 = \alpha_{11}x_1 + \alpha_{21}x_2 + \dots + \alpha_{p1}x_p$$

$$F_2 = \alpha_{12}x_1 + \alpha_{22}x_2 + \dots + \alpha_{p2}x_p$$

.....

$$F_p = \alpha_{1p}x_1 + \alpha_{2p}x_2 + \dots + \alpha_{pp}x_p$$

或者 $F_i = \alpha_{1i}x_1 + \alpha_{2i}x_2 + \dots + \alpha_{pi}x_p (i = 1, 2, \dots, p)$

其中, $\alpha_{1k}^2 + \alpha_{2k}^2 + \alpha_{3k}^2 + \dots + \alpha_{pk}^2 = 1 (k = 1, 2, \dots, p)$

新变量 F 还需满足以下要求: (1) F_i 和 F_j ($i \neq j$) 不相关; (2) F_1 的方差大于 F_2 , 以此类推。这样 F_1 中就包含了原数据库中最多的信息, F_2 中包含的信息是除 F_1 外最多的, 以此类推。经过以上计算, 得到的 F_i 就是原数据信息中包含的第 i 个主成分。

6.5.2 主要计算流程

应用主成分分析 (PCA) 方法对有机物的来源进行解析, 主要计算流程包括数据标准化、计算相关系数矩阵、计算特征值和特征向量、特征值排序、选择主成分、构建变换矩阵和数据投影等环节, 见图 1。

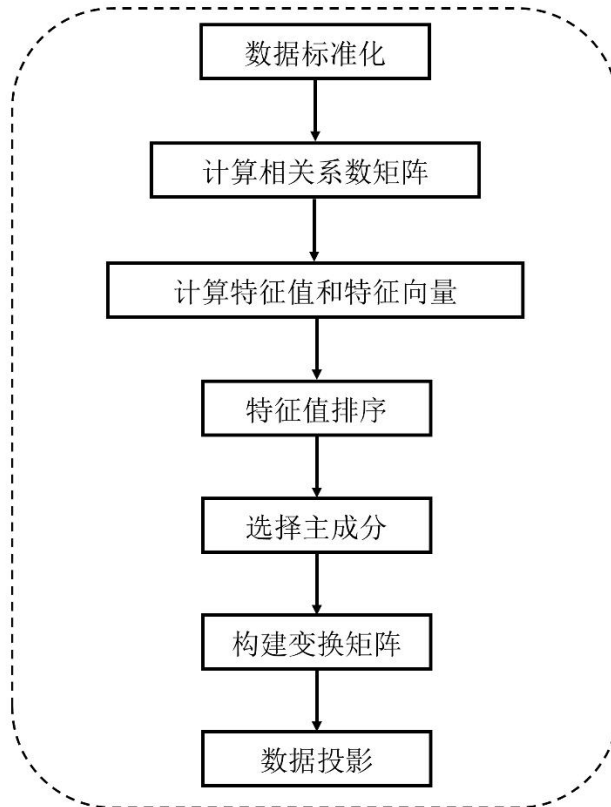


图 1 PCA 主要计算流程

6.5.3 数据标准化

对原始数据进行标准化处理，使得每个特征的均值为 0，方差为 1，以消除不同特征的量纲差异。

对于一组样本资料 X ，其中有 m 个观测值 x_1, x_2, \dots, x_m ，共 n 个样。

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

(1) 计算每一列的平均值 $\mu_i = \frac{1}{n} \sum_{j=1}^n x_{ji}$

(2) 每一列的方差 $\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ji} - \mu_i)^2$

(3) 对数据进行标准化处理， $Z_{ji} = \frac{x_{ji} - \mu_i}{\sigma_i}$

得到协方差矩阵 Z :

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}$$

6.5.4 计算相关系数矩阵

根据标准化后的数据，计算相关系数矩阵。计算相关系数， $r_{ij} = \frac{\sum_{k=1}^n z_{ki} * z_{kj}}{n-1}$ ， $(i, j=1, 2, \dots, m)$ ，得到相关系数矩阵 R

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix}$$

r_{ij} 即 X 矩阵的第 i 列的样本序列和第 j 列的样本序列之间的相关关系，其值为-1 到 1 之间，且 R 矩阵应该为对称矩阵，即 $r_{ij} = r_{ji}$ 。

相关系数程度区分如下表 1 和表 2 所示：

表 2 正负相关性

相关系数 r	相关性
$r > 0$	正线性相关
$r = 0$	线性无关
$r < 0$	负线性相关

表 3 相关性程度大小

相关系数绝对值 r	相关性程度
1	完全相关
[0.8, 1)	高度相关
[0.5, 0.8)	中度相关
[0.3, 0.5)	低度相关
[0, 0.3)	不相关

6.5.5 计算特征值和特征向量

协方差矩阵 Σ 是实对称阵，知其特征值为非负，不妨设其特征值 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$ ，它们对应的正交化后的单位特征向量如下：

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix}; a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix}; \dots; a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix}$$

若原先 X 的各个列代表的指标变量，合成向量，记为 $Var = [Var_1, Var_2, \dots, Var_p]^T$ ，则有 X 的第 i 个主成分为 $F_i = (a_i)^T Var = \alpha_{1i} * Var_1 + \alpha_{2i} * Var_2 + \dots + \alpha_{pi} * Var_p$ 。

特征值是衡量主成分影响力的重要指标，它代表引入该主成分可以解释平均多少原始标量的信息。特征值表示每个特征向量的重要程度，特征向量表示数据中的主要方向。

6.5.6 特征值排序

将特征值按照从大到小的顺序进行排序，同时对应的特征向量也进行相应的排序。求出特征值后要按大小予以排列： $\lambda_1 \geq \lambda_2 > \dots \geq \lambda_p \geq 0$ 。如果特征值小于 1，表示该主成分的解释力非常低，一般以特征值大于 1 位筛选主成分的标准。

6.5.7 选择主成分

方差贡献率，指的是一个主成分所能够解释的方差占全部方差的比例，这个值越大，说明主成分综合原始变量的信息的能力越强。方差贡献率的计算公式为：

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} (i = 1, 2, \dots, p)$$

相应的，主成分筛选中所确定的前 m 个主成分所能解释的全部方差占总方差的比例称为累计方差贡献率。其公式为：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i = 1, 2, \dots, p)$$

第一主成分的方差贡献率最大，它能解释原始变量 X_1, X_2, \dots, X_p 的能力最强，第 2，第 3， \dots ，第 p 个主成分的解释能力依次递减。

主成分数量的选取则是根据累积贡献率确定，一般要求累积贡献率达到 85% 以上，这样能保证新变量能包括原始变量的绝大多数信息。

此外，主成分的个数选取有 3 个主要的衡量标准：保留的主成分使得方差贡献率达到 85% 以上；保留的主成分的方差（特征值）大于 1；碎石图绘制了关于各主成分及其特征值的图形，只需要保留图形中变化最大之处以上的主成分即可。

6.5.8 构建变换矩阵

将选择的 k 个特征向量按列形成一个新的矩阵，称为变换矩阵。该矩阵可以对原始数据进行线性变换，将其映射到新的 k 维特征空间。

6.5.9 数据投影

将原始数据通过变换矩阵进行投影，得到降维后的新数据。投影的计算方法是将原始数据与变换矩阵相乘。

6.6 结果表达分析

(1) KMO 取样适合度检验和 Bartlett 球形检验

这两项检验是用来判断是否可以主成分分析。对于 KMO 值：0.8 上非常合适做主成分分析，0.7-0.8 之间一般适合，0.6-0.7 之间不太适合，0.5-0.6 之间表示差，0.5 下表示极不适合，对于 Bartlett 的检验 ($p < 0.05$, 严格来说 $p < 0.01$)，若显著性小于 0.05 或 0.01，拒绝原假设，则说明可以做土壤有机物污染主成分分析。

(2) 方差解释表格、成分矩阵表

特征根的值按照从大到小进行排序，一般选取方差（特征值）大于 1 的主成分，可以有效保留主成分。由成分矩阵表可得到主成分分析降维后的计算公式。

(3) 碎石图

碎石图是关于有机物各主成分及其特征值的图形，根据特征值下降的坡度来确认需要选择的主成分个数，当曲线变得平缓时，对应的因子数就可以作为参考因子数，只需要保留图形中变化最大之处以上的主成分即可。

(4) 因子载荷系数表、因子载荷矩阵热力图和因子载荷象限分析

因子载荷系数表可以分析到每个主成分中隐变量的重要性。因子载荷矩阵热力图可以分析到每个主成分中隐变量的重要性，热力图颜色越深说明相关性越大。因子载荷象限分析图通过将多因子降维成双主成分或者三主成分，通过象限图的方式呈现土壤有机物主成分的空间分布。

7 主成分分析模型应用

以位于山东省新泰市的盛泉煤矿土壤中有机物采样数据为例进行主成分分析。

7.1 导入数据

	甲基萘	萘	芴	菲	荧蒽	芘	苯并a萘	屈	苯并b荧蒽	苯并k荧蒽	苯并a比	苯并ghi花	石油烃
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	28.500
2	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	24.000
3	.300	.285	.000	.160	.000	.355	.135	.150	.650	.000	.230	.255	125.500
4	.000	.000	.000	.130	.000	.000	.000	.000	.000	.000	.000	.000	32.000
5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	52.000
6	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	19.000
7	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	44.000
8	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	28.500
9	1.100	.740	.000	.340	.000	.185	.000	.000	.205	.000	.205	.000	183.500
10	.650	.440	.000	.265	.000	.210	.000	.155	.295	.000	.000	.200	65.000
11	.000	.140	.000	.130	.000	.145	.000	.000	.000	.000	.000	.000	62.000
12	.500	.360	.000	.595	.200	.440	.215	.305	.465	.000	.210	.315	133.500
13	.300	.220	.330	.850	.000	.790	.585	.935	.900	.125	.375	.290	337.500
14	.200	.265	.000	.365	.150	.205	.125	.210	.285	.000	.000	.000	108.500
15	.300	.220	.000	.650	.280	.505	.295	.410	.605	.125	.270	.230	281.500
16	.300	.225	.000	.680	.285	.665	.355	.430	.645	.125	.325	.360	299.000
17	.300	.335	.130	.430	.170	.445	.205	.275	.380	.000	.185	.300	130.500
18	.350	.000	.000	.600	.205	.710	.315	.400	.585	.000	.305	.605	282.000
19	.000	.000	.000	.145	.000	.210	.000	.000	.000	.000	.000	.000	123.500
20	.150	.000	.000	.310	.000	.375	.150	.195	.270	.000	.000	.000	121.500
21	.300	.160	.000	.465	.155	.520	.235	.290	.375	.000	.200	.430	177.500
22	.200	.200	.000	.320	.000	.315	.000	.000	.325	.000	.000	.410	113.500
23	.250	.250	.000	.435	.150	.330	.070	.075	.570	.000	.230	.550	174.000
24	.350	.305	.000	.415	.000	.320	.000	.000	.680	.140	.325	.780	296.500
25	.400	.575	.130	.905	.300	.620	.180	.180	.755	.235	.465	.610	125.500
26	.400	.215	.000	.640	.185	.505	.175	.175	.610	.000	.340	.715	304.500
27	.200	.165	.000	.410	.000	.345	.000	.000	.390	.000	.235	.755	214.500
28	.250	.325	.000	.510	.000	.365	.000	.000	.485	.000	.260	.575	168.000

7.2 选择变量

选择“分析”、“降维”、“因子分析”，打开因子分析对话框。选择左侧框中变量，单击箭头，将其选入“变量”列表框。注意主成分分析和因子分析适用的是数值型变量，因此选入分类数据。

屈	苯并 b荧 蒽	苯并 k荧 蒽	苯并 a比 荧	苯并 ghi花	石油烃	变量	变量	变量	变量	变量	变量
.000	.000	.000	.000	.000	28.500						
.000	.000	.000	.000	.000	24.000						
.150	.650	.000	.230	.255	125.500						
.000	.000	.000	.000	.000	32.000						
.000	.000	.000	.000	.000	52.000						
.000	.000	.000	.000	.000	19.000						
.000	.000	.000	.000	.000	44.000						
.000	.000	.000	.000	.000	28.500						
.000	.205	.000	.205	.000	183.500						
.155	.295	.000	.000	.200	65.000						
.000	.000	.000	.000	.000	62.000						
.305	.465	.000	.210	.315	133.500						
.935	.900	.125	.375	.290	337.500						
.210	.285	.000	.000	.000	108.500						
.410	.605	.125	.270	.230	281.500						
.430	.645	.125	.325	.360	299.000						
.275	.380	.000	.185	.300	130.500						
.400	.585	.000	.305	.605	282.000						



7.3 统计量设置

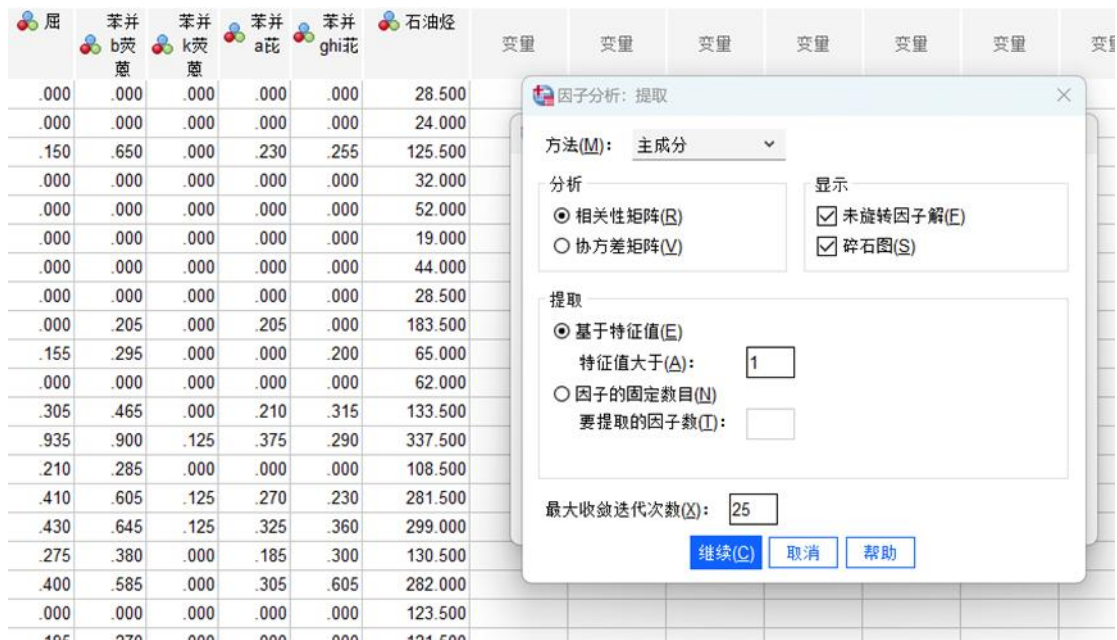
单击“描述”按钮，打开“因子分析：描述统计”对话框。该对话框包括“统计”和“相关性矩阵”两个选项组。本例子选择默认设置“原始分析结果”和“KMO和Bartlett球形检验”。单击“继续”按钮返回。

屈	苯并 b荧 蒽	苯并 k荧 蒽	苯并 a比 荧	苯并 ghi花	石油烃	变量	变量	变量	变量	变量	变量
.000	.000	.000	.000	.000	28.500						
.000	.000	.000	.000	.000	24.000						
.150	.650	.000	.230	.255	125.500						
.000	.000	.000	.000	.000	32.000						
.000	.000	.000	.000	.000	52.000						
.000	.000	.000	.000	.000	19.000						
.000	.000	.000	.000	.000	44.000						
.000	.000	.000	.000	.000	28.500						
.000	.205	.000	.205	.000	183.500						
.155	.295	.000	.000	.200	65.000						
.000	.000	.000	.000	.000	62.000						
.305	.465	.000	.210	.315	133.500						
.935	.900	.125	.375	.290	337.500						
.210	.285	.000	.000	.000	108.500						
.410	.605	.125	.270	.230	281.500						
.430	.645	.125	.325	.360	299.000						
.275	.380	.000	.185	.300	130.500						



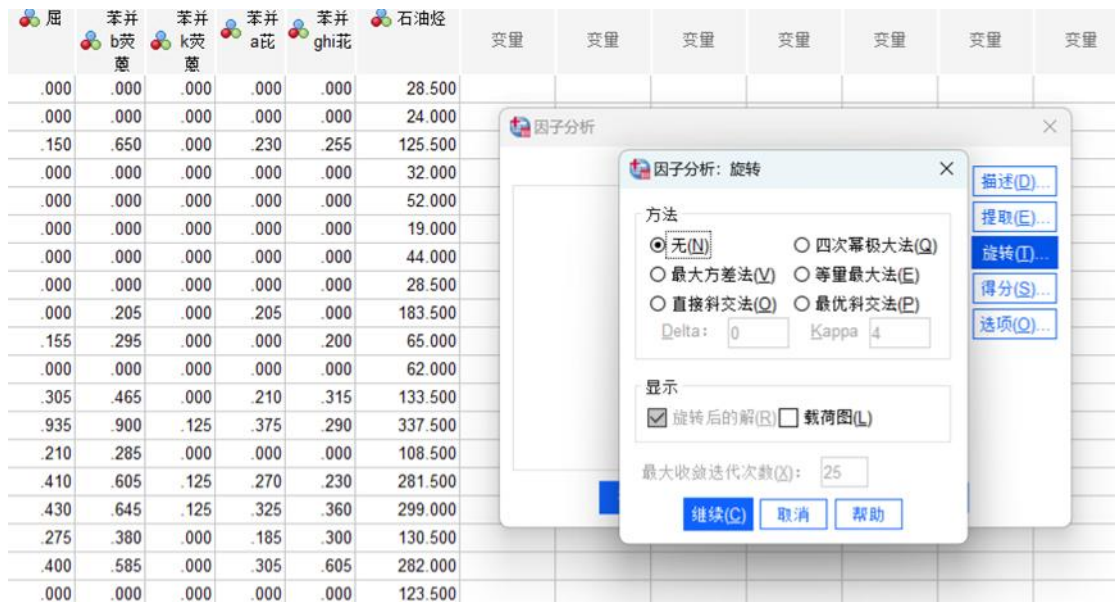
7.4 主成分抽取设置

单击“提取”按钮，打开“因子分析：提取”对话框，如下设置，然后返回。



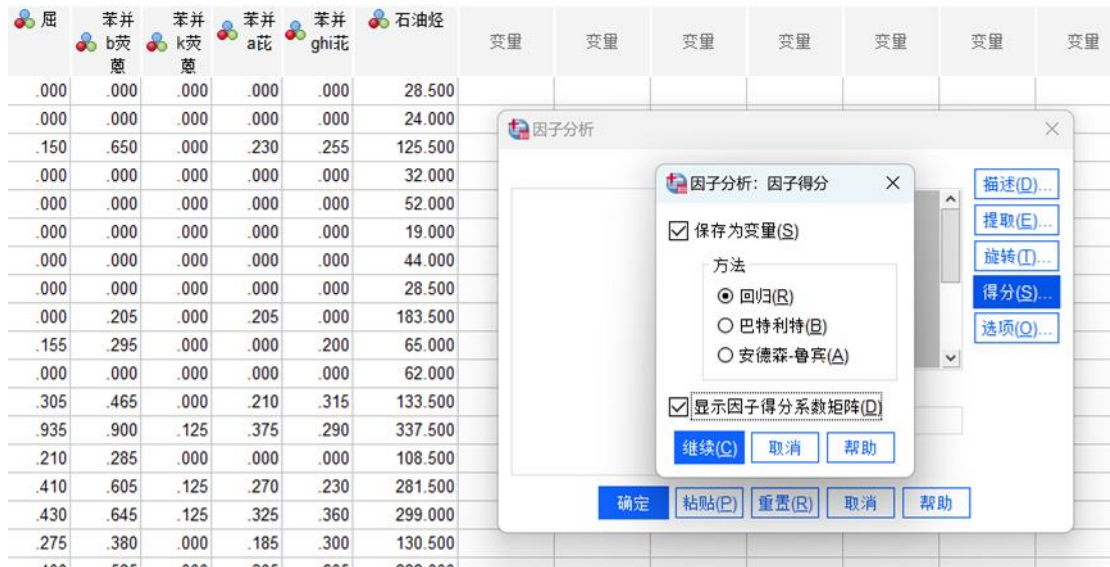
7.5 旋转方式设置

单击“旋转”按钮，打开“因子分析：旋转”对话框，该对话框用于因子分析时的旋转方式设置。因此本例子不做选择，保持默认选择无。



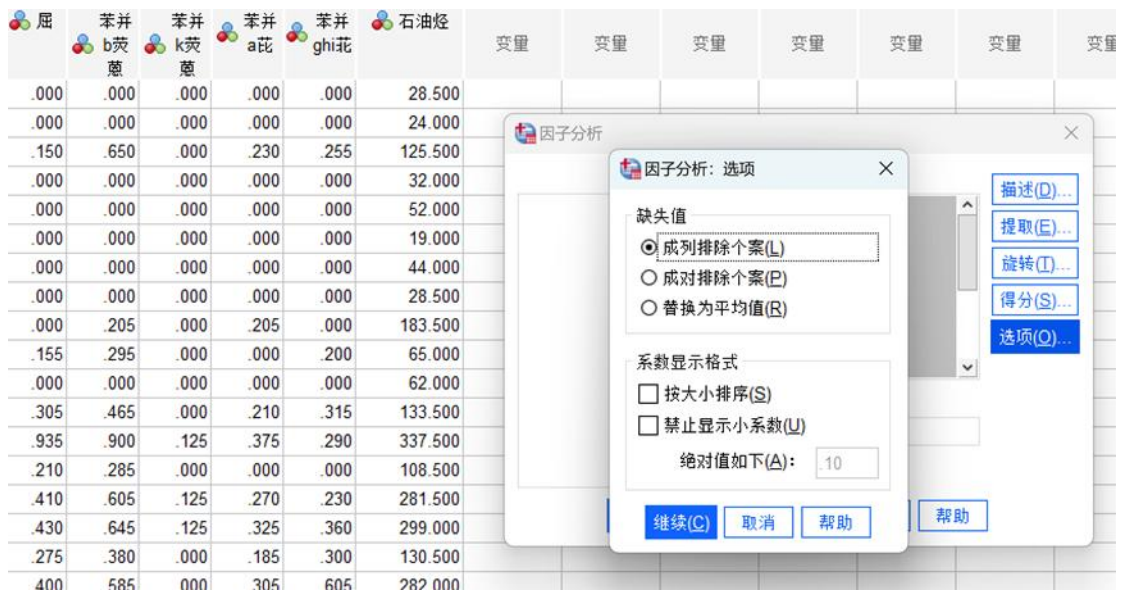
7.6 得分保存设置

单击“得分”按钮，打开“因子分析：因子得分”对话框。选中“保存为变量”复选框，将主成分分析得分保存于工作数据中，方法选择“回归”；然后选中“显示因子得分系数矩阵”复选框，单击“继续”返回。



7.7 选项设置

单击“选项”按钮，打开“因子分析:选项”对话框，该对话框用于对缺失值的处理和系数显示格式的设置。单击“继续”按钮，返回“因子分析”。



7.8 输出分析结果

对话框设置完毕，单击“确定”按钮，SPSS 输出分析结果。

7.9 分析结果解读

下表是 KMO 和 Bartlett 检验结果。从中可以看出，KMO 取样适合度检验统计量为 0.714 的高水平，说明案例中变量之间的信息有重叠，相关度大。Bartlett 检验卡方值 474.890，p 值小于 0.001，达到及其显著的水平。两种检验都表明，该例子适合主成分分析。

KMO 取样适切性量数		.714
巴特利特球形度检验	近似卡方	474.890
	自由度	78
	显著性	.000

公因子方差表示各变量所包含的信息能够被提取的主成分所表示的程度，也称为“共同度”。“初始值”表示每个变量演示信息均为 1，即 100%。而“提取”栏表示该变量的方差能被主成分所表示的程度，可以看出，变量的方差能被主成分解释在 70%以上。

	初始	提取
甲基萘	1.000	.883
萘	1.000	.959
芴	1.000	.780
菲	1.000	.935
荧蒽	1.000	.520
芘	1.000	.921
苯并 a 蒽	1.000	.931
屈	1.000	.925
苯并 b 荧蒽	1.000	.907
苯并 k 荧蒽	1.000	.455
苯并 a 芘	1.000	.900
苯并 ghi 芘	1.000	.834
石油烃	1.000	.782

提取方法：主成分分析法。

下表给出了提取的主成分的方差总解释量。可以看出，每个主成分能解释的方差的比例不同。第 1 主成分特征值为 7.8，解释原始变量的方差比例为 60%。

总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	7.800	60.000	60.000	7.800	60.000	60.000
2	1.798	13.829	73.829	1.798	13.829	73.829
3	1.132	8.709	82.538	1.132	8.709	82.538
4	.795	6.117	88.655			
5	.747	5.745	94.400			
6	.310	2.381	96.781			
7	.126	.972	97.753			
8	.112	.861	98.614			
9	.066	.508	99.123			
10	.060	.462	99.585			
11	.032	.249	99.834			
12	.018	.137	99.971			
13	.004	.029	100.000			

提取方法：主成分分析法。

下表给出了 3 个主成分的成分矩阵，也称为因子载荷，实质是指个主成分和个原始变量的相关系数。通过因子载荷矩阵可以得到原始指标变量的线性组合，如 $P_1 = a_{11} \times F_1 + a_{12} \times F_2$ ，其中 P_1 为指标变量 1， a_{11} 、 a_{12} 分别为与变量 P_1 在同一行的因子载荷， F_1 、 F_2 分别为提取的公因子。主成分表达式中相关系数值越大，表示该主成分对原始变量的代表性越大。可以看出，第一主成分 y1 与个原始变量的关系是 3 个主成分中较大的，说明它对原始变量的解释量最多。

成分矩阵^a

	成分 1	成分 2	成分 3
甲基萘	.540	.619	.457
萘	.529	.654	.501
芴	.550	-.502	.475
菲	.965	.048	-.031
荧蒽	.673	.019	-.258
芘	.948	-.103	-.106
苯并 a 蒽	.816	-.507	.085
屈	.771	-.535	.211
苯并 b 荧蒽	.944	.078	-.093
苯并 k 荧蒽	.669	-.079	.040
苯并 a 芘	.922	.179	-.134
苯并 ghi 芘	.677	.365	-.492
石油烃	.858	.034	-.213

提取方法：主成分分析法。

a. 提取了 3 个成分。

下表是 3 个主成分的得分系数。根据主成分得分系数，就可以计算出每个因素在 3 个主成分上的得分。假设用 X_1, X_2, \dots, X_n 表示所有标准化的原始变量，用 F_1, F_2 表示主成分得分。那么就可以得出如下的得分函数 $F_1 = 0.069X_1 + 0.068X_2 + \dots + 0.110X_{13}$ 。简单来说，因子载荷和得分的区别，前者使用主成分来表示原始变量，而后者是用原始变量表示主成分。

成分得分系数矩阵

	成分 1	成分 2	成分 3
甲基萘	.069	.344	.404
萘	.068	.364	.443
芴	.071	-.279	.420
菲	.124	.027	-.028
荧蒽	.086	.011	-.228
芘	.122	-.058	-.094
苯并 a 蒽	.105	-.282	.075
屈	.099	-.298	.186
苯并 b 荧蒽	.121	.043	-.082
苯并 k 荧蒽	.086	-.044	.035
苯并 a 芘	.118	.099	-.118
苯并 ghi 芘	.087	.203	-.434
石油烃	.110	.019	-.188

提取方法：主成分分析法。

组件得分。

下表是成分得分的协方差矩阵，由于 3 个主成分互不相关，经过标准化，因此成分得分的协方差矩阵为单位矩阵，即主成分之间的协方差均为 0，每个主成分的方差均为 1。

成分得分协方差矩阵

成分	1	2	3
1	1.000	.000	.000
2	.000	1.000	.000
3	.000	.000	1.000

提取方法：主成分分析法。

组件得分。

下图使碎石图，它显示了各个主成分的重要程度，横轴为主成分序号，纵轴为特征值大小。系统选取特征值大于 1 的前三个成分。

碎石图

