

团 体 标 准

T/ACEF □□□—2025

环境空气 挥发性有机物来源解析 正定矩 阵因子分解（PMF）模型应用技术指南

Ambient air—Source apportionment on volatile organic
compounds—Technical guide on the calculation of positive matrix
factorization model

（征求意见稿）

2025-□□-□□发布

2025-□□-□□实施

中华环保联合会 发布

目 次

前 言 III

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 3

5 数据准备 4

6 基础计算 5

7 旋转计算 6

8 VOCs 源贡献计算 7

9 结果合理性判断 7

附 录 A 8



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。

请注意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。

本文件由中华环保联合会归口。

本文件主编单位：

本文件副主编单位：

本文件参编单位：

本文件主要起草人：



环境空气 挥发性有机物来源解析 正定矩阵因子分解（PMF）模型应用技术指南

1 适用范围

本文件规定了环境空气挥发性有机物来源解析正定矩阵因子分解模型计算流程及结果判断方法，包括模型原理、基础计算、旋转计算、资源贡献计算、结果合理性判断等内容。

本文件适用于应用正定矩阵因子分解模型对环境空气挥发性有机物的手工、在线采样的分析数据进行来源解析。

本文件适用于城市、区县、园区等小尺度、且不涉及大气化学反应过程的挥发性有机物来源解析。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

HJ 630 环境监测质量管理技术导则

HJ 683 环境空气 醛、酮类化合物的测定 高效液相色谱法

HJ 759 环境空气 65 种挥发性有机物的测定 罐采样/气相色谱-质谱法

HJ 1010 环境空气挥发性有机物气相色谱连续监测系统技术要求及检测方法

HJ 1353 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

3 术语和定义

3.1

环境空气 Ambient air

是指人群、植物、动物和建筑物所暴露的室外空气。

[来源：GB 3095-2012，3.1]

3.2

挥发性有机物 Volatile Organic Compounds (VOCs)

参与大气光化学反应的有机化合物，或者根据有关规定确定的有机化合物。在表征 VOCs 总体排放情况时，根据行业特征和环境管理要求，可采用总挥发性有机物（以 TVOC 表示）、非甲烷总烃（以 NMHC 表示）作为污染物的评价项目。

[来源：GB 37822-2019，3.1]

3.3

环境受体 Ambient Receptor

受到污染源排放影响的环境空气，简称受体。

3.4

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

一种多元因子分析类模型，将样本数据的矩阵（X）分解为因子贡献矩阵（G）和因子谱矩阵（F），对因子谱矩阵进行识别，并定量计算样本的因子贡献。在环境空气挥发性有机物来源解析中，应用正定矩阵因子分解模型（PMF）计算是通过对环境样品（受体）的化学组分数据进行分析，提取若干因子，利用标识组分将因子识别为不同的源类，再通过多元回归计算，结合源成分谱和排放源等信息，综合判断不同因子（源类）对环境样品的贡献。

3.5

受体化学组分 Ambient Receptor Chemical Composition

环境空气 VOCs 样品化学组分的浓度或占比信息。

3.6

VOCs源成分谱 VOCs source Profile of pollution source

特定污染源类所排放 VOCs 的化学组分的浓度或占比信息，以各化学组分的质量浓度占比计。

3.7

标识组分 Tracer Species

污染源 VOCs 成分谱中对该源类有指示作用的一种或多种有机物组分，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。标识组分在该源成分谱中的含量，一般比在其他源类中高。

3.8

因子 Factor

PMF 模型计算提取的因子谱和因子贡献信息。因子谱是因子中各化学组分浓度（ $\mu\text{g}/\text{m}^3$ ）或占比（%）信息，可以被识别为污染源类；在因子谱识别结果的基础上，因子贡献信息可以计算相应的污染源类贡献浓度及占比。

3.9

不确定度 Uncertainty

源解析工作中采样、实验分析等环节引起的误差。

3.10

Q值 Function Q

反映 PMF 模型计算结果好坏程度的一种目标函数值。Q 值有三种类型：Q 理论值（Q_{Theo}）、Q 计算值（Q_{True}）、Q 修正值（Q_{Robust}）。Q 理论值是 PMF 模型在理想条件计算时的数值，可以通过 PMF 模型输入的数据个数减去可用于因子计算的数据个数得到。Q 计算值是 PMF 模型在实际条件计算时的数值，理论上越接近 Q 理论值，表明结果越理想。Q 修正值是在 PMF 模型中对无法参与计算或异常的

数据（不确定度一缩放残差大于 4）修正后进行实际计算时的 Q 值。

3.11

残差 Residual

VOCs 组分的实际测试值与 PMF 模型计算值之间的差值。

3.12

基础计算 Base Run

在 PMF 模型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（要求得到最小 Q 值）得到因子贡献矩阵（G）和因子谱矩阵（F）的计算方法。通过基础计算，可以得到基于输入的 VOCs 相关浓度及不确定度数据的初步结果。

3.13

旋转计算 Rotational Run

对于基础计算得到的因子贡献矩阵（G）和因子谱矩阵（F）结果，在一定的约束条件下再次进行的计算，将因子贡献矩阵（G）和因子谱矩阵（F）可以被转换成另一对矩阵（G*和 F*），使因子中标识组分更加突出，解决基础计算中因子难以识别的问题。

4 模型原理和计算流程

4.1 模型原理

PMF 的原理是利用权重计算出环境受体中各 VOCs 组分的误差，再通过最小二乘法确定 VOCs 的主要污染源及其贡献率。应用 PMF 模型进行 VOCs 来源解析计算的原理是将受体 VOCs 组分浓度矩阵（ $n \times m$ ）因子化，分解为两个因子矩阵，（ $p \times m$ ）和（ $n \times p$ ），以及一个“残差矩阵”（ $n \times m$ ），如下式表示：

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

式中：（ $n \times p$ ）——因子贡献矩阵；
 （ $p \times m$ ）——因子谱矩阵；
 n——样品个数；
 m——VOCs 组分种数；
 p——解析出来的污染源的数目。

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 矩阵和 F 矩阵。

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

式中： σ_{ij} ——第 j 个样品中第 i 个 VOCs 成分的不确定度。

4.2 工作程序

应用 PMF 模型进行 VOCs 来源解析的计算流程包括数据准备及导入，基础计算及结果分析、旋转计算及结果分析和结合当地主要排放源信息进行合理性判断等环节，见图 1。进行 PMF 模型时，需要根据 VOCs 浓度和不确定数据、基础运算结果和旋转运算结果，不断对 VOCs 浓度数据进行剔除和参数设定，以得到相对较优的计算结果。在导入 VOCs 组分浓度数据时，特征组分应优先选择环境浓度高、源指示性强和监测数据相对完整的组分进行 PMF 运算。

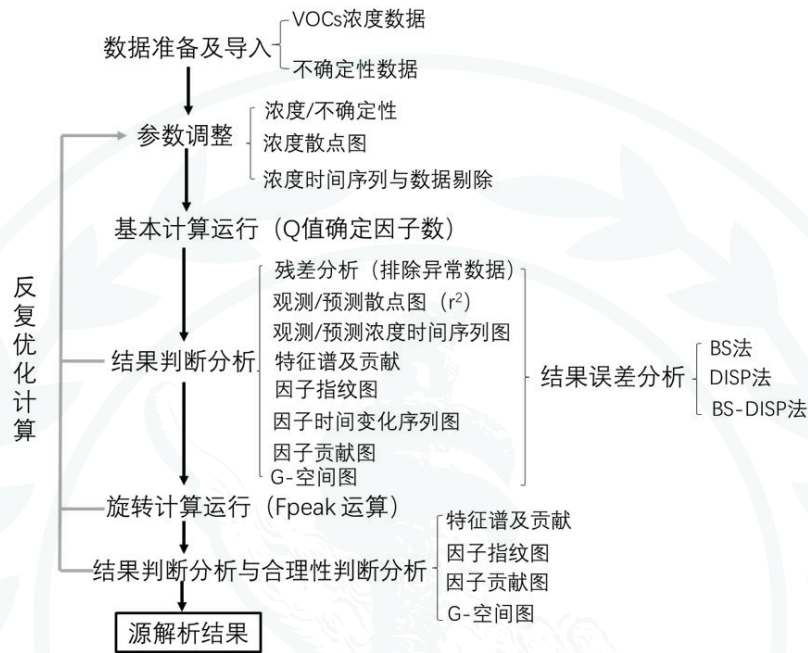


图 1 PMF 模型的主要计算流程

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 VOCs 质量浓度，以及烷烃、烯烃、芳香烃等化学组分浓度数据，至少需要 56 种 VOCs 组分(PAMS 组分)，每个受体采样点位要求不少于 100 个有效样品的数据量。VOCs 组分需涵盖主要污染源的标识组分，常用的标识组分包括苯、甲苯、乙苯、二甲苯、苯乙烯、乙烯、乙烷、乙炔等。为解析更多的源类，推荐增加醇类、酯类、酮类、醚类、卤代烃等有机标识组分。增加输入的 VOCs 组分和样品数量，将使 PMF 模型计算结果更稳定、更具指向性。

5.2 数据质量要求

用于 VOCs 源解析工作的质量浓度数据，应采用国家、行业或地方方法文件获得；没有国家、行业或地方方法文件，可采用国际文件、国外文件或研究建立的方法，经确认后使用，同时监测需按照 HJ630 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。VOCs 组分的分析需参照相关国家文件方法 HJ 683、HJ 759、HJ 1010 等。

5.3 不确定度计算

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

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

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

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

当 VOCs 组分的质量浓度小于等于 MDL 时，采用公式（3）计算；当 VOCs 组分的质量浓度大于 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 VOCs 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 VOCs 组分质量浓度数据缺失较多时，可设定较大的误差系数。

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，检验不同组分之间关系、组分随时间变化情况。对于异常高或者低、种类明显偏少的组分数据，经核实没有特定污染过程（如事故排放等），可予以剔除。

6.1.2 数据计算权重的选择。VOCs 的关键标识组分（如苯、甲苯、乙苯、二甲苯、苯乙烯、乙烯、乙烷、乙炔等组分）的计算权重需设为“强”（Strong）；其他非关键标识组分的计算权重一般设为一般设为“中”（Medium）；接近最低检测限，或波动过大的组分，可设为“弱”（Weak）计算权重。

6.1.3 因子（Factor）个数的确定。因子个数需要基于对 VOCs 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。

6.1.4 根据城市基本情况（比如根据当地排放源清单、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。

6.1.5 根据模型计算中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，结合城市基本情况，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：

（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 则可能为适合的因子数。

（2）根据组分的残差大小判断。组分的残差（尤其是标识组分）应在正负 3 以内，如果残差有较大的分布，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 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。

（4）根据回归检验结果判断。将 G 矩阵与 VOCs 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，则说明选择的因子可能过少。

6.2 基础计算结果判断

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。通过 Q 理论值（QTheo）和 Q 计算值（QTrue）的大小、是否收敛（Converge）等，初步选择查看特定的计算结果。从 Q 计算值（QTrue）与 Q 理论值

（QTheo）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（QTrue）与 Q 理论值（QTheo）越接近（85 %~115 %之间），表明模拟计算结果更合理。

6.2.2 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；如果某组分的观测值和预测值之间无强相关性，则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。

6.2.3 通过“残差分析（Residual Analysis）”，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。

6.3 基础计算结果分析

6.3.1 基础计算的结果主要包括因子谱矩阵（F）和因子贡献矩阵（G）。因子谱展示了不同VOCs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。

6.3.2 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。

6.3.3 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。

6.4 基础计算的误差评估

6.4.1 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。

6.4.2 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。

6.4.3 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。

6.4.4 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。

7 旋转计算

7.1 旋转计算方法

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）。

7.2 旋转计算运行

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“6.2、6.3、6.4”进行相应的结果判断、分析和误差评估。

7.3 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 模型进一步解析。

8 VOCs 源贡献计算

8.1 在 PMF 模型计算中输入数据包含 VOCs 质量浓度及各组分质量浓度，将计算得到的 G 矩阵（源贡献矩阵）与 VOCs 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 VOCs 的贡献，拟合公式如下：

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

式中：VOCs_j——第 j 个样品的 VOCs 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——源个数；

s_kg_{jk}——计算得到的 k 源对第 j 个样品的贡献。

8.2 多个环境受体监测点位数据的 PMF 模型计算。将城市多个点位的源解析结果进行平均，则得到城市的解析结果。

9 结果合理性判断

9.1 源解析结果既要符合模型计算要求，还需要符合实际情况。

9.2 符合模型计算要求是模型模拟计算出的各 VOCs 组分浓度与实测测试结果接近。可通过直接比较 VOCs 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。

9.3 符合实际情况要求是经过识别的因子谱及其源贡献合理。可通过 PMF 模型计算得到的因子谱与实测源成分谱匹配性比较、源贡献与当地主要排放源信息比较分析等方法来评估模型结算结果的合理性。当地主要排放源信息主要来自排污许可、大气污染物排放清单、大气污染物与温室气体融合排放清单、环境统计等。

附 录 A
(资料性)

VOCs 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识别为具体的 VOCs 源类：

- A.1 根据资料调研收集的各源类主要的标识组分来识别。某个因子的标识组分与其他因子相比占比
较高，则可识别为相应的源类。例如，某因子中异戊二烯占比最高，则该因子可识别为天然源；因子中
乙烷、乙烯、乙炔和丙烷等 C2~C3 组分的占比高，可识别为燃烧源等。常见源类的标识组分见表 A.1。
- A.2 根据当地实际监测得到主要污染源的源成分谱来识别。通过当地污染源 VOCs 采样和组分定性
定量分析，构建当地主要 VOCs 污染源的源成分谱。将 PMF 模型计算得到的因子谱与源成分谱进行对
照识别。
- A.3 根据 PMF 模型进一步优化计算得到的因子来识别。当基于基础计算的结果无法有效识别具体
源类时，可使用约束模型计算进行优化，突出因子中的标识组分，进而识别因子。
- A.4 VOCs 主要源的常用标识组分表（表 A.1）中，石油化工包括石油炼制和化学原料和化学制品
制造，有机化工包括煤制合成气生产、煤制液体燃料生产、氮肥制造、涂料油墨胶黏剂制造、化学纤维
制造、医药制造、化学农药制造、焦化等。

表 2 VOCs 主要源的常用标识组分

污染源	组分
固定燃烧源	苯
	乙炔
	乙烯
	甲苯
	苯乙烯
	间, 对-二甲苯
	丙酮
	乙烷
生物质燃烧源	乙烯
	苯
	乙醛
	甲醛
	乙炔
汽油车	甲醛
	异戊烷
	甲苯
	乙烯
	乙烷
	正己烷
	间, 对-二甲苯
	乙炔
柴油车/机	3-甲基壬烷
	正十二烷
	甲醛
	乙炔
	丙酮

	十一烷
	癸烷
	丙烯
	乙醛
	乙烯
	正壬烷
	环己酮
	苯乙烯
石油化工	正丙烷
	环己烷
	正丁烷
	异丁烷
	丙酮
	正己烷
	癸烷
	正壬烷
有机化工	丙酮
	异丙醇
	二氯甲烷
	氯丙烯
	丙烯腈
	氯甲烷
	二硫化碳
	丙烯醛
	正己烷
	环己烷
	丙烷
	间, 对-二甲苯
工业涂装	乙酸丁酯
	乙苯
	1-甲基-乙酸丙酯
	邻-二甲苯
	苯乙烯
	甲苯
	丙烯醛
	乙酸甲酯
工业涂装-低 VOCs 含量	1-甲基-乙酸丙酯
	乙酸丁酯
	乙醇
	乙酸乙酯
	4-甲基-2-戊酮
	苯
	甲苯
包装印刷	乙酸甲酯
	乙苯
	1, 2-二氯丙烷
	2-乙基-1-己醇

包装印刷-低 VOCs 含量	乙酸乙酯
	异丙醇
	乙醇
	异戊烷
	2-甲基己烷
	3-甲基己烷
塑料和橡胶制品制造	苯乙烯
	甲苯
	乙酸乙酯
	乙苯
	邻-二甲苯
	间,对-二甲苯
木材加工	2-丁酮
	异戊烷
	癸烷
	正戊烷
	丙烯
	正丁烯
其他溶剂	甲苯
	苯乙烯
	甲醛
	四氯乙烯
	乙二醇
	二甲苯
制鞋	乙醇
	甲醇
	丁酮
	丙二醇
	丙酮
	甲苯
食品制造	乙酸乙酯
	2-丁酮
	3-甲基戊烷
	环戊烷
	甲基环戊烷
	2-甲基戊烷
废弃物处理源	3-甲基己烷
	正己烷
	异丙醇
	乙醇
	正戊烷
	正辛烷
	甲苯
	氯苯
	乙烷
	全氯乙烯
	正丁烷
	乙烯

	甲苯
	丙烷
	苯
油品储运销	异戊烷
	正丁烷
	正戊烷
	丙烷
	甲苯
	异丁烷
	甲基叔丁基醚
	3-甲基戊烷
沥青铺路	十一烷
	正壬烷
	癸烷
	正辛烷
	环己烷
	1, 2, 4-三甲苯
	邻乙基甲苯
餐饮	乙醇
	己醛
	丙烯醛
	甲醛
	戊醛
	丁烯醛
汽修	乙酸丁酯
	丙烯酸丁酯
	异丙烯酸乙酯
	乙醇
	1, 2, 4-三甲苯
	邻-二甲苯
	二氯甲烷
	间, 对-二甲苯
汽修-低 VOCs 含量	乙酸丁酯
	异丙烯酸乙酯
	乙醇
	丙烯酸丁酯
	二氯甲烷