

ICS 13.020.40

CCS Z 00

T/SSSC

中 国 土 壤 学 会 团 体 标 准

T/SSSC 014 —2025

基于地下水安全的土壤有机污染物 安全阈值推导技术指南

Technical guidelines for estimating threshold value of soil organic contaminants
based on groundwater safety

2025 - 04 - 15 发布

2025 - 04 - 16 实施

中国土壤学会 发 布

目 次

前 言	3
1 范围	4
2 规范性引用文件	4
3 术语和定义	4
3.1 土-水分配系数 (K_{sw}) Soil-Water Partition Coefficient (K_{sw})	4
3.2 固-液分配系数 (K_d) Solid-Liquid Partition Coefficient (K_d)	5
3.3 包气带衰减因子 (AF) Vadose Zone Attenuation Factor (AF)	5
3.4 污染物地下水稀释因子 (DF) Groundwater Dilution Factor (DF)	5
4 总体要求	5
5 基于地下水安全的土壤有机污染物安全阈值的推导技术框架	5
6 土-水分配过程计算	6
6.1 土-水分配系数计算	6
6.2 固-液分配系数计算	6
7 包气带衰减过程计算	7
8 地下水稀释过程计算	7
9 土壤安全阈值的推导	7
附录 A	9
附录 B	13

前 言

本标准按照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。

本标准由中国科学院生态环境研究中心提出。

本标准由中国土壤学会归口。

本标准起草单位：中国科学院生态环境研究中心、中科院建筑设计研究院有限公司。

本标准主要起草人：王美娥、张雪云、马万凯、谢天、陈卫平。

基于地下水安全的土壤有机污染物安全阈值推导技术指南

1 范围

本标准规定了基于地下水安全的土壤有机污染物安全阈值推导技术的规范性引用文件、术语和定义、总体要求、技术框架、土水分配过程计算、包气带浓度衰减过程计算、地下水稀释过程计算以及土壤安全阈值的推导。

本标准适用于固液分配系数 K_d 小于20的有机污染物的安全阈值推导。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应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。

GB/T 31270.4 化学农药环境安全评价试验准则 第4部分：土壤吸附/解吸试验

GB 14148 地下水质量标准

HJ 164 地下水环境监测技术规范

HJ 166 土壤环境监测技术规范

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与修复监测技术导则

HJ 25.3 建设用地土壤污染风险评估技术导则

HJ 610 环境影响评价技术评价导则 地下水环境

HJ 1019 地块和土壤地下水中挥发性有机物采样技术导则

T/SSSC 013—2025 基于Hydrus-1D的土壤有机污染物包气带衰减因子估算技术指南

3 术语和定义

下列术语和定义适用于本文件。

3.1 土-水分配系数 (K_{sw}) Soil-Water Partition Coefficient (K_{sw})

污染物在土壤中的总浓度和水相中的浓度之比（L/kg）。

3.2 固-液分配系数（ K_d ） Solid-Liquid Partition Coefficient（ K_d ）

污染物在土壤固相和水相中的浓度之比，也称吸附系数（L/kg）。

3.3 包气带衰减因子（ AF ） Vadose Zone Attenuation Factor（ AF ）

土壤表层孔隙水与包气带下层土壤溶液中污染物的浓度之比，无量纲。

3.4 污染物地下水稀释因子（ DF ） Groundwater Dilution Factor（ DF ）

包气带下层土壤渗滤液中的污染物在饱和带中迁移至保护目标（地下水监测井）时的浓度稀释倍数，为包气带下层土壤渗滤液污染物浓度与监测井地下水污染物浓度之比，无量纲。

4 总体要求

- （1）本标准中污染源存在于土壤表层，且在源处不发生衰减；
- （2）本标准假设有机污染物在土-水分配过程中瞬间达到平衡；
- （3）本标准中含水层为非承压含水层；
- （4）本标准不考虑非水相液体的存在。

5 基于地下水安全的土壤有机污染物安全阈值的推导技术框架

基于地下水安全的土壤有机污染物安全阈值的推导技术框架主要包括：

- （1）土-水分配过程计算；
- （2）包气带衰减过程计算；
- （3）地下水稀释过程计算；
- （4）土壤安全阈值的推导。

方法框架如图1所示。

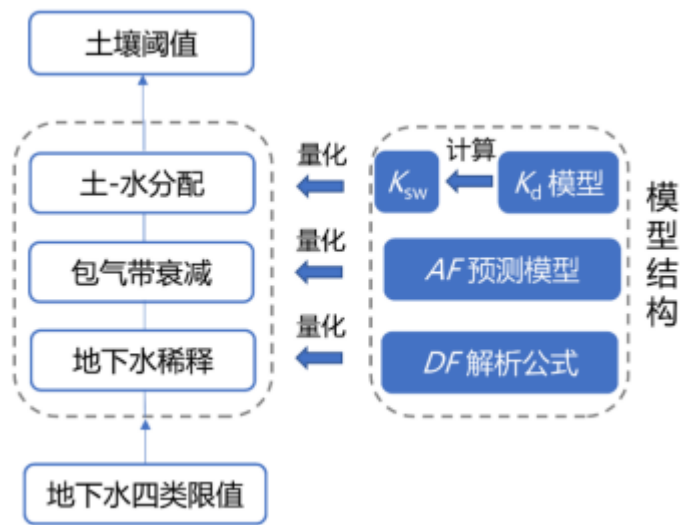


图 1 土壤有机污染物安全阈值的推导技术框架

6 土-水分配过程计算

6.1 土-水分配系数计算

根据土壤有机污染物土-水分配系数 K_{sw} 计算公式如下：

$$K_{sw} = \frac{C_s}{C_e} = \frac{\theta_{ws} + K_d \cdot \rho_b + H' \cdot \theta_{as}}{\rho_b} \quad \text{公式 (1)}$$

其中， C_e 为平衡时液相浓度（mg/L）； C_s 为平衡时土壤中污染物的总浓度（mg/kg）； K_d 为固-液分配系数（L/kg）； θ_{ws} 为包气带中的孔隙水体积比（无量纲）； θ_{as} 为包气带中的孔隙空气体积比（无量纲）； ρ_b 为土壤干容重（g/cm³）； H' 为环境温度下的亨利常数（无量纲）。

6.2 固-液分配系数计算

6.2.1 一般要求

- （1）本标准中的目标化合物的固-液分配系数 K_d 计算方法需适合不同类型土壤；
- （2）本标准中的目标化合物的固-液分配系数 K_d 计算方法包括模型预测和实验室测试。

6.2.2 固-液分配系数模型预测方法

本标准采用机器学习模型获得目标有机污染物的固-液分配系数，包括五个基本步骤：数据收集、数据集筛选、数据预处理、模型构建、模型分析。具体步骤见资料性附录A。

6.2.3 固-液分配系数实验室批量实验测试

本标准采用目标有机污染物在不同类型土壤中的吸附-解吸批量实验获得固-液分配系数，方法可参考资料性附录B所提供的参考文献。

7 包气带衰减过程计算

本标准采用HYDRUS-1D模型估算有机污染物在包气带中衰减因子（ AF ），具体内容参考基于Hydrus-1D的土壤有机污染物包气带衰减因子估算技术指南（T/SSSC 013—2025）。

8 地下水稀释过程计算

本标准中采用包气带衰减因子（ DF ）量化地下水稀释过程， DF 的计算公式如下：

$$DF = \frac{C_d}{C_{gw}} = 1 + \frac{K \cdot i \cdot d}{I \cdot L} \quad \text{公式（2）}$$

其中， C_d 为包气带下层土壤淋滤液中的有机污染物浓度（mg/L）； C_{gw} 为保护目标（地下水监测井）地下水中有有机污染物浓度（mg/L）； DF 与场地特征参数有关，其中，场地特征参数 K 为含水层水力传导度（m/s）； i 为水力梯度（无量纲）； d 为混合区深度（m）； I 为内渗率（m/s）； L 为平行于地下水流的源区长度（m），上述场地特征参数可直接采用HJ 25.3—2019给出的推荐值，并据此推导 DF 的推荐值为20。

9 土壤安全阈值的推导

基于地下水安全的土壤有机污染物安全阈值的计算公式如下：

$$C_s = C_{gw} \cdot K_{sw} \cdot AF \cdot DF \quad \text{公式 (3)}$$

其中， C_s 为保护地下水的土壤有机污染物安全阈值（mg/kg）； C_{gw} 为地下水
中有机污染物的浓度限值（mg/L），部分有机污染物参考地下水质量标准（GB/T
14848-2017）； K_{sw} 为土-水分配系数（L/kg）； AF 为包气带衰减因子； DF 为地下
水稀释因子。

附录 A

(资料性)

基于机器学习方法的有机污染物固-液分配系数 (K_d) 预测模型

一、模型框架

本模型的框架如下图所示，分为四个步骤：（1）数据收集；（2）数据筛选与预处理；（3）模型构建；（4）模型解释。

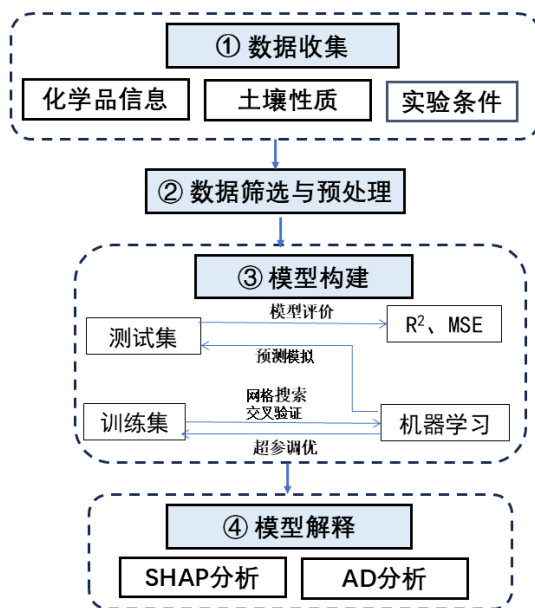


图2 方法框架

二、数据收集

收集目标有机污染物的固-液分配过程相关的变量，包括但不限于化学品信息、土壤性质和实验条件。

（1）化学品信息，包括但不限于Abraham描述符（E、S、A、B和V）、辛醇-水分配系数的对数值 $\log K_{ow}$ 、分子量MW，以及拓扑结构描述符ATSC2dv等。其中，Abraham描述符的可获得途径为UFZ-LSER数据库，对于没有Abraham描述符记录的化学品，可通过UFZ-LSER提供的预测工具或通过相关深度学习模型计算获得； $\log K_{ow}$ 和MW数据可通过Chemicalize计算获得；ATSC2dv可通过Mordred软件包计算获得。

（2）土壤性质，包括目标层土壤的有机碳、黏粒含量、容重、土壤含水率、颗粒密度等理化性质。土壤性质数据可通过实际测量或土壤剖面数据集获得。

(3) 实验条件（包括固-液比、温度等），数据可通过实际测量或使用经验值获得。

二、数据集筛选与预处理

1 数据筛选要求

(1) 研究包含土壤性质（pH、阳离子交换量、土壤有机质和黏粒含量等）、实验条件（包括固-液比、温度和初始溶液浓度等）信息；

(4) 供试土壤样品未受污染且完整，未添加或去除任何成分；

(5) 拟合等温线的决定系数 R^2 大于0.96，以确保吸附实验数据的质量；

(6) 土壤有机污染物吸附平衡实验的操作遵守OECD（欧洲经济合作与发展组织）或USEPA（美国环境保护局）指南，以确保实验结果的标准化和结果可比性。

2 数据预处理要求

(1) 统一数据格式：分别统一不同数据集中各变量的量纲，确保变量单位和符号一致；

(2) 异常数据处理：去除异常值，以提高数据质量；

(3) 特征参数筛选：使用相关分析，计算包括土壤性质（pH、CEC、黏粒含量和SOC）、分子描述符（Abraham描述符、 $\log K_{ow}$ 、MW和ATSC2dv）、实验条件（温度和固-液比）和平衡浓度 C_e 土壤性质（pH、CEC、黏粒含量和SOC等）等变量之间的相关系数。基于相关系数，筛选出对 K_d 预测重要的特征参数，具体过程参考资料性附录B所提供的参考文献；

(4) 归一化处理：输入特征值进行归一化处理公式如下：

$$X_{ij}^s = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad \text{公式 (1)}$$

其中， x_{ij} 指第*i*个样本的第*j*个预测变量的原始值； X_{ij}^s 指 x_{ij} 的归一化值； $\bar{x}_j$ 和 σ_j 分别指第*j*个预测变量的平均值和标准差。

(5) 数据集拆分：按照随机分组原则拆分数数据集，来自同一等温线的 C_e-Q_e 被标记为一个特定组，这个组被随机分配到同一个子数据集中，具体过程参考资料性附录B所提供的参考文献。

三、固-液分配系数模型构建

本方法采用极端梯度提升（XGBoost）预测土壤中有机污染物Kd，具体过程参考资料性附录B所提供的参考文献。如图1所示，训练集用来进行估计模型的参数，采用留一组法进行交叉验证。不断调整模型的超参数以及输入特征，直到验证集的评价指标达到最优。最优超参数和输入特征固定后，采用测试集测试两个模型的外推能力，并根据评价指标选择最佳算法，包括决定系数 R^2 、均方根误差RMSE、平均绝对误差MAE和方差解释度，计算公式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pre} - y_{act})^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{pre} - y_{act}| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{act} - y_{pre})^2}{\sum_{i=1}^n (y_{act} - y_{mean})^2} \quad (4)$$

$$\text{Explained variance} = 1 - \frac{\text{var}(y_{act} - y_{pre})}{\text{var}(y_{act})} \quad (5)$$

式中， y_{act} 指实际值； y_{pre} 指预测值。

由于测试结果与数据集拆分的随机状态有关，本方法将随机种子设置为0（即数据集被随机分割），并将测试过程重复100次，以充分展示模型的外推能力。建模的编程语言推荐Python 3.9.7，使用scikit-learn库构建XGBoost模型（Pedregosa et al., 2011）。

四、模型解释

(1) SHAP分析方法

每个特征的Shapley值为预测值和平均预测值的差（公式6）。

$$f(x_i) = \overline{f(x)} + \sum_{j=1}^k \text{shap}_{ij} \quad (6)$$

$$C_j = \frac{\sum_{i=1}^n |\text{shap}_{ij}|}{n} \quad (7)$$

式中， shap_{ij} 是第j个特征对应的第i行的计算的Shapley值； $\overline{f(x)}$ 是预测值的平均值； $f(x_i)$ 是第i行的预测值； C_j 用于量化第j个特征的贡献； n 是训练样本数； k 是特征数。

具体过程参考资料性附录B所提供的参考文献。

(2) 应用域分析 (AD)

首先，预测变量归一化，归一化的特征乘以它们的贡献 (C_j)，(公式 1)。其次，用公式 8 计算第 i 个样本与其 $n-1$ 个相邻样本的平均欧氏距离，再对训练集中的所有样本计算出平均欧式距离的向量 $\mathbf{d}$ 。再次，最大离群值 (向量 $\mathbf{d}$ 第 75% 分位数加 1.5 倍的的 IQR) 被定义为阈值。最后，计算新的预测样本的 $\overline{d}_i$ 并将其与阈值比较 (公式 9)。如果它大于阈值，该模型就不能应用于新的预测样本，否则就可以应用。

$$X_{ij}^{sw} = C_j \cdot X_{ij}^s \quad (8)$$

$$\overline{d}_i = \frac{\sum_{r=1}^{n-1} D_{ir}}{n-1} \quad (9)$$

式中， X_{ij}^{sw} 指的是特征贡献加权的 X_{ij}^s ； X_{ij}^s 指的是标准化的 X_{ij} ； C_j 是第 j 个特征的贡献； D_{ir} 指的是第 i 和第 r 个样本间的欧氏距离； $\overline{d}_i$ 指的是第 i 个样本和训练集中其他样本的平均距离； n 指的是训练集中的样本数。

具体过程参考资料性附录B所提供的文献。

附录 B

(资料性)

本标准方法参考文献建议

序号	方法	文献
1	基于机器学习的固-液分配系数 K_d 预测模型构建	Ma 等, Environmental Pollution, 2023,323: 121297
2	基于批量实验的固-液分配系数 K_d 测试方法	罗晓丽等, 环境科学学报, 2008, 28 (7) : 1375-1380; Chiou, Partition and adsorption of organic contaminants in environmental systems. 2002, John Wiley & Sons, Inc. ISBN: 0-471-23325
3	数据筛选方法	Drobnic和Pustisek, Electronics, 2020, 9: 761; Graham等, Ecology, 2003, 84: 2809-2815
4	数据集拆分方法	Yang等, Environmental Science & Technology, 2021, 55: 14316-14328; Zhang等, Environmental Science & Technology, 2020, 54: 7008-7018
5	机器学习方法	Pedregosa等, Artificial Intelligence, 2021, 298: 103502
6	SHAP分析法	Aas等, Artificial Intelligence, 2021, 298: 103502
7	应用域分析方法	Meyer等, Methods in Ecology and Evolution, 2021, 12: 1620-1633; Sahigara等, Journal of Cheminformatics, 2013, 5: 27