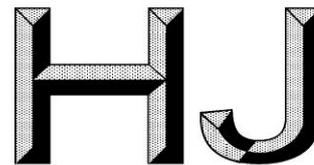


附件 2



# 中华人民共和国国家环境保护标准

HJ □□□—202□

## 区域性土壤环境背景含量统计

### 技术导则

Technical guidelines on the statistics of regional environmental  
background concentration of soil

(征求意见稿)

20□□-□□-□□发布

20□□-□□-□□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 工作目标和程序.....	1
5 区域性土壤环境背景含量数据获取.....	2
6 数据处理分析.....	7
7 区域性土壤环境背景含量统计与表征.....	7
8 区域性土壤环境背景值的确定.....	8
9 实施与监督 .....	8
附录 A（资料性附录） t 分布表.....	9
附录 B（资料性附录） 异常值判别方法 .....	10

## 前 言

为贯彻《中华人民共和国环境保护法》和《中华人民共和国土壤污染防治法》，防止土壤污染，保护生态环境，维护公众健康，制定本标准。

本标准规定了区域性土壤环境背景含量统计工作程序以及区域性土壤环境背景含量数据获取、数据处理分析、统计与表征等技术要求。

本标准的附录 A 和附录 B 为资料性附录。

本标准为首次制定。

本标准由生态环境部土壤生态环境司、法规与标准司组织制订。

本标准主要起草单位：生态环境部南京环境科学研究所、生态环境部环境标准研究所等。

本标准生态环境部 20□□年□□月□□日批准。

本标准自 20□□年□□月□□日起实施。

本标准由生态环境部解释。

# 区域性土壤环境背景含量统计技术导则

## 1 适用范围

本标准规定了区域性土壤环境背景含量统计工作程序以及区域性土壤环境背景含量数据获取、数据处理分析、统计与表征等技术要求。

本标准适用于区域性土壤环境背景含量的统计，不适用于地块尺度土壤环境背景含量的统计。

## 2 规范性引用文件

本标准内容引用了下列文件中的条款。凡是不注明日期的引用文件，其有效版本适用于本标准。

GB/T 4882 数据的统计处理和解释 正态性检验

GB 15618 土壤环境质量 农用地土壤污染风险管控标准（试行）

GB/T 17296 中国土壤分类与代码

GB 36600 土壤环境质量 建设用地土壤污染风险管控标准（试行）

HJ/T 166 土壤环境监测技术规范

## 3 术语和定义

下列术语和定义适用于本标准。

### 3.1

**土壤 soil**

指位于陆地表层能够生长植物的疏松多孔物质层及其相关自然地理要素的综合体。

### 3.2

**土壤环境背景含量 environmental background concentration of soil**

指一定时间条件下，仅受地球化学过程和非点源输入影响的土壤中元素或化合物的含量。

### 3.3

**数据集 data set**

由点位数据组成的集合，每一列代表一个数据信息项（变量），每一行对应一个点位数据。

### 3.4

**土壤剖面 soil profile**

由地表向下直至成土母质的土壤纵切面。由若干层次组成，以其不同的颜色、土壤质地、土壤结构、松紧度以及新生体等而区分。

## 4 工作目标和程序

### 4.1 工作目标

4.1.1 明确调查区域、时间范围和目标元素或化合物。统计与表征调查区域某一时间范围内土壤中目

标元素或化合物的环境背景含量。

4.1.2 调查区域应同时明确平面范围与垂向深度。平面范围可以是省（直辖市、自治区）、市或县（县级市）级行政区域，也可以是行政区域内的一定土壤类型、母质类型、地质地貌类型、流域或土地利用类型等区域，或以上类型组合形成的区域。垂向深度可以是相对固定土壤深度，也可以是一定土壤层次。

4.1.3 时间范围可以是历史上某一时间段，也可以是当前。

4.1.4 土壤中目标元素或化合物可以是某一种、一类或多种等。

## 4.2 工作程序及工作内容

4.2.1 区域性土壤环境背景含量统计的工作程序如图 1 所示。

4.2.2 工作内容包括区域性土壤环境背景含量数据获取、数据处理分析和区域性土壤环境背景含量统计与表征。

### 4.2.2.1 区域性土壤环境背景含量数据获取

应收集整理区域性土壤环境背景含量统计所需相关数据资料，明确是否存在历史环境调查数据：

- a) 当不存在历史数据时，应开展土壤环境背景调查，获取土壤环境背景含量数据。
- b) 当存在历史数据时，应进行点位数据评估。若不存在满足要求的点位数据，应开展土壤环境背景调查，获取土壤环境背景含量数据。
- c) 当存在满足要求的点位数据时，应进行数据集评估。
  - 1) 当数据集满足要求时，将满足要求的点位数据作为土壤环境背景含量数据；
  - 2) 当数据集不满足要求时，应开展土壤环境背景调查，可结合数据集中满足要求的点位数据开展调查，将调查获得的数据和满足要求的点位数据共同作为土壤环境背景含量数据。

### 4.2.2.2 数据处理分析

- a) 根据区域性土壤环境背景含量统计的工作目标，划分统计单元。
- b) 检验区域性土壤环境背景含量数据的分布类型，识别区域性土壤环境背景含量数据异常值，对异常值进行判别和处理。

### 4.2.2.3 区域性土壤环境背景含量统计与表征

- a) 在数据处理分析的基础上，统计不同统计单元的区域性土壤环境背景含量。
- b) 用图和表的形式表征区域性土壤环境背景含量，编制区域性土壤环境背景含量统计技术报告。

## 5 区域性土壤环境背景含量数据获取

### 5.1 数据资料收集与整理

#### 5.1.1 数据资料收集

5.1.1.1 根据工作目标，收集调查区域资料，主要包括：自然和社会经济信息、土地利用及其变化、环境数据资料等。

5.1.1.2 自然和社会经济信息：自然信息包括地理、地形、地貌、植被、土壤类型、母质母岩类型、水文、地质和气象资料等；社会经济信息包括人口密度和分布、敏感目标分布、地方性疾病统计信息、经济现状、产业结构、矿产资源分布等。

5.1.1.3 土地利用及其变化：包括调查区域土地利用现状及其变化图件资料等。

5.1.1.4 环境数据资料：包括由政府机关和权威机构所保存和发布的环境资料（如区域环境保护规划、环境质量公告等）、污染源信息（如工农业生产及排污、污灌、化肥农药施用情况资料等）、历史环境调查数据等。

#### 5.1.2 资料整理

5.1.2.1 根据专业知识和经验识别资料中的错误和不合理的信息。

5.1.2.2 对收集的资料进行归类整理，确定是否存在所需历史环境调查数据，如存在，应建立由单个点位数据组成的环境调查数据集。

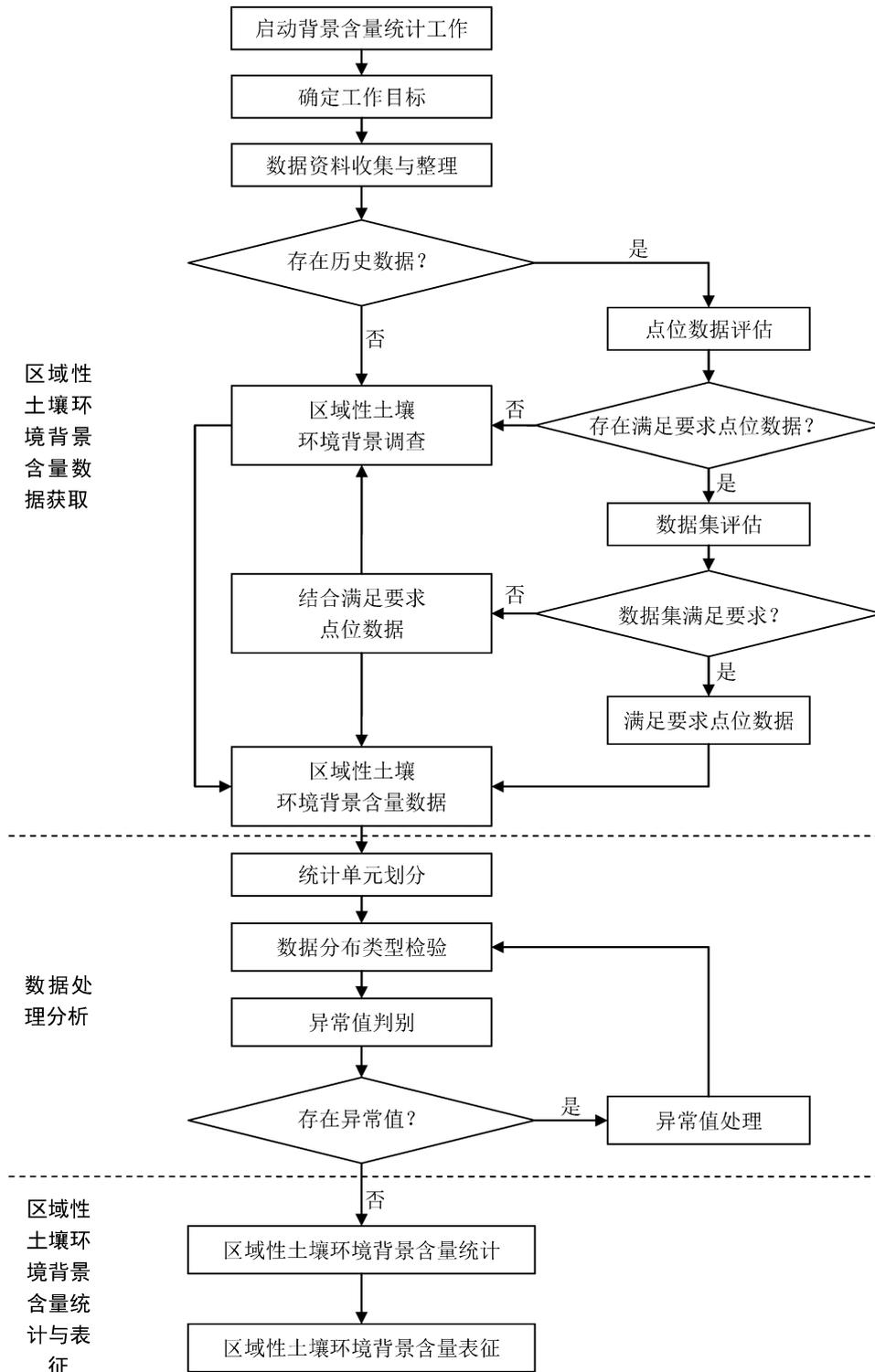


图 1 区域性土壤环境背景含量统计工作程序

## 5.2 区域性土壤环境背景调查

### 5.2.1 调查项目确定

5.2.1.1 调查项目可以是土壤中各种无机元素和有机/无机化合物，需根据工作目标确定。

5.2.1.2 可将部分土壤理化性质参数作为调查项目，包括土壤 pH、土壤质地、容重、有机质含量、阳离子交换量等。

### 5.2.2 调查单元划分

5.2.2.1 在调查区域内，一般按照影响区域性土壤环境背景含量的主导因素土壤类型或母质类型划分调查单元。

a) 按照土壤类型划分调查单元时，省级调查区域以亚类或土类为主划分，市级以土属或亚类进行划分，县级以土种或土属进行划分。

b) 土壤类型采用 GB/T 17296 规定的分类方法。

5.2.2.2 可根据工作目标要求，增加流域或行政区等因素综合划分调查单元。

### 5.2.3 调查点位布设

#### 5.2.3.1 基础样本数量

a) 由均方差和绝对偏差计算样本数，计算公式如 (1)：

$$N = \frac{t^2 s^2}{D^2} \quad (1)$$

式中：N——基础样本数量，个；

t——选定置信水平（土壤环境监测一般选定为 95%）一定自由度下的 t 值（附录 A）；

s<sup>2</sup>——均方差，可从先前的其他研究或者从极差 R 进行估算；

D——可接受的绝对偏差

其中：均方差由公式 (2) 计算：

$$s^2 = \left( \frac{R}{4} \right)^2 \quad (2)$$

式中：s<sup>2</sup>——均方差；

R——极差；

b) 由变异系数和相对偏差计算样本数，计算公式如 (1)，可变为公式 (3)：

$$N = \frac{t^2 C_v^2}{m^2} \quad (3)$$

式中：N——基础样本数量；

t——选定置信水平（土壤环境监测一般选定为 95%）一定自由度下的 t 值（附录 A）；

C<sub>v</sub>——变异系数，可从先前的其他研究资料中估计，%；

m——可接受的相对偏差，土壤环境监测一般限定为 20%~30%，%。

没有历史资料的地区、土壤变异程度不太大的地区，一般 C<sub>v</sub> 可用 10%~30% 粗略估计。

#### 5.2.3.2 布点数量

各调查单元的布点数量应不少于基础样本数量，实际工作中土壤布点数量还要根据调查目的、调查精度和调查区域环境状况等因素确定。一般要求每个调查单元最少布设 30 个采样点。

#### 5.2.3.3 布点方法

a) 针对调查区域内的调查单元，采用系统布点或系统随机布点方法进行布点。

b) 系统布点：网格的数量等于布点数量，将采样单元划分成面积相等的网格，每个网格内布设 1 个采样点。在既有认知较少，自然因素变异较大，或者目标元素或化合物含量变化较大的区域

更适合采用。

- c) 系统随机布点：网格的数量大于布点数量 2 倍，将采样单元划分成面积相等的网格，从中随机抽取满足布点数量的网格，每个网格内布设 1 个采样点。适用于母质类型单一，地形平坦，其他自然因素差异较小的区域。
- d) 根据实际情况可适当调整网格的起始位置，避开过多网格落在道路或河流上，使样品更具代表性。

#### 5.2.3.4 野外点位选择

- a) 点位选择需要具有代表性，避开人为活动的影响。
- b) 采样点的自然景观应符合土壤环境背景研究的要求。采样点选在被采土壤类型特征明显，地形相对平坦、稳定、植被良好的地点；坡脚、洼地等具有从属景观特征的地点不布设采样点。
- c) 不宜在多种土类、多种母质母岩交错分布、面积较小的边缘地区布设采样点。
- d) 采样点以剖面发育完整、层次较清楚、无侵入体为准，不在水土流失严重或表土被破坏处设采样点。
- e) 城镇、住宅、道路、沟渠、粪坑、坟墓附近等人为干扰大的地方，不宜设采样点。
- f) 除特殊要求外，采样点离铁路、公路至少 300 m 以上。
- g) 农用地应在农作物收获后采样，采样点尽量避免在肥料、农药集中使用区，以使样品点尽可能少受人为活动的影响。

#### 5.2.4 样品采集

##### 5.2.4.1 样品采集方法

- a) 一般按照发生层次采集剖面土壤样品。
- b) 可以根据工作目标的特殊需求，采集固定深度土壤样品。

##### 5.2.4.2 剖面土壤样品采集

- a) 对于采集剖面土壤样品的：
  - 1) 每个采样点均采集土壤剖面样品。剖面的规格一般为长 1.5 m，宽 0.8 m，深 1.2 m。挖掘土壤剖面要使观察面向阳，表土和底土分两侧放置。一般每个剖面采集 A 淋溶层、B 淀积层、C 母质层三层土样。过渡层一般不采样。当地下水位较高时，剖面挖至地下水出露时为止；山地丘陵土层较薄时，剖面挖至风化层。在风化壳很厚的地方，除作为特殊研究的剖面外，一般仍挖至 1.2 m 为限。
  - 2) 尽量不利用自然剖面。对于有利用价值的非常典型而又有代表性的自然剖面，需削去其表面土 20 cm，然后按要求取样。
  - 3) 对 B 层发育不完整（不发育）的山地土壤，只采 A、C 两层；
  - 4) 干旱地区剖面发育不完善的土壤，在表层 5~20 cm、心土层 50 cm、底土层 100 cm 左右采样。
  - 5) 水稻土按照 A 耕作层、P 犁底层、C 母质层（或 G 潜育层、潜育层）分层采样，对 P 层太薄的剖面，只采 A、C 两层（或 A、G 层或 A、W 层）。
  - 6) 对 A 层特别深厚，淀积层不甚发育，一米内见不到母质的土类剖面，按 A 层 5~20 cm、A/B 层 60~90 cm、B 层 100~200 cm 采集土壤。
  - 7) 草甸土和潮土一般在 A 层 5~20 cm、C1 层（或 BW 层）50 cm、C2 层 100~120 cm 处采样。
- b) 采样次序自下而上，先采剖面的底层样品，再采中层样品，最后采上层样品。测试重金属的样品尽量用竹片或竹刀去除与金属采样器接触的部分土壤，再用其取样。
- c) 每层剖面样品采集 1 kg 左右，装入样品袋，样品袋一般由棉布缝制而成，如潮湿样品可内衬塑料袋（供无机化合物测定）或将样品置于玻璃瓶内（供有机化合物测定）。采样的同时，由专人

填写样品标签、采样记录；标签一式两份，一份放入袋中，一份系在袋口，标签上标注采样时间、地点、样品编号、分析测试项目、采样深度和经纬度。采样结束，需逐项检查采样记录、样袋标签和土壤样品，如有缺项和错误，及时补齐更正。将底土和表土按原层次回填到采样坑中，方可离开现场，并在采样示意图上标出采样地点，避免下次在相同处采集剖面样，有条件的可以在采样点埋设标志桩。

#### 5.2.4.3 固定深度土壤样品采集

根据工作目标要求，一般按照固定深度 0~20 cm、40~60 cm、80~100 cm 采集土壤样品。

#### 5.2.5 样品制备与保存

样品制备按照 HJ/T 166 中的相关规定执行。

#### 5.2.6 样品分析测试

样品的分析测试按照 HJ/T 166 中的相关规定执行，标准方法参见 GB 15618 和 GB 36600 中的指定方法。

#### 5.2.7 质量控制

为保证所产生的区域性土壤环境背景调查数据具有代表性、准确性、精密性、可比性和完整性。质量控制应涉及调查的全部过程。质量控制参照 HJ/T 166 中的相关规定执行。

### 5.3 历史数据评估

#### 5.3.1 点位数据评估

##### 5.3.1.1 数据的完整性评估

- a) 历史数据应至少包括以下信息项：采样日期、布点方法、选点原则、样点位置（坐标）、样品采集方法、样品采集层次、分析测试项目、样品分析测试方法、野外信息记录（例如土壤类型、母质类型和土地利用类型，是农用地的还要记录作物类型）。
- b) 历史数据达到以上信息项要求的，则其完整性满足要求。

##### 5.3.1.2 数据的符合性评估

- a) 对满足完整性要求的历史数据，进行数据的符合性评估。
- b) 根据工作目标的相关要求，对历史数据进行符合性评估。
- c) 历史数据的采样日期、样品采集层次、分析测试项目等相关信息符合区域性土壤环境背景含量统计工作目标相关要求的，则其符合性满足要求。

##### 5.3.1.3 数据的规范性评估

- a) 对满足符合性要求的历史数据，进行数据的规范性评估。
- b) 对照本标准对区域性土壤环境背景调查点位布设、样品分析测试和质量控制的相关要求，评估数据的规范性。
- c) 历史数据的布点方法、点位选择、样品分析测试和质量控制达到本标准相关要求的，则其规范性满足要求。

#### 5.3.2 数据集评估

5.3.2.1 按照 5.2.2 划分调查单元，统计数据集中各调查单元内满足完整性、符合性和规范性要求的点位数量。

5.3.2.2 当各调查单元内满足要求的点位数量达到本标准 5.2.3.2 布点数量要求时，该数据集满足要求，否则，该数据集不满足要求。

### 5.4 结合历史数据开展区域性土壤环境背景调查

5.4.1 当数据集不满足要求，但存在满足要求的点位数据时，可结合这些点位数据开展调查。

5.4.2 在调查过程中应考虑与满足要求的历史点位数据的衔接：

- a) 在布点数量上，可以减去满足要求的点位数据对应的点位数量。
- b) 在布点位置上，在满足要求的点位数据对应的网格内可不布设采样点位。

## 6 数据处理分析

### 6.1 统计单元划分

6.1.1 按照影响区域性土壤环境背景含量的主导因素土壤类型或母质类型将调查样本数据划分为不同统计单元。可根据工作目标要求，按照流域、行政区等划分统计单元。

6.1.2 基于统计单元内各层次样本数据，进行数据分布类型检验、异常值判别与处理以及区域性土壤环境背景含量统计与表征。

### 6.2 数据分布类型检验

6.2.1 区域性土壤环境背景含量数据的分布类型大致分为正态分布、对数正态分布和其他分布。

6.2.2 数据的正态性检验按照 GB/T 4882 的规定执行。

6.2.3 非正态分布的数据，进行适当的正态转换后再进行正态性检验。

### 6.3 异常值判别与处理

6.3.1 对于明显来源于局部受污染场所的数据，或者因样品采集、分析检测、数据输入错误等原因导致的异常数据，必须进行判别和处理。常用判别样本异常值的方法包括格拉布斯（Grubbs）检验法、狄克逊（Dixon）检验法、T（Thompson）检验法、箱线图法和富集系数法（附录 B）。

6.3.2 应用上述方法进行异常值剔除应注意以下事项：

6.3.2.1 在判断异常值前，应先注意检查原始记录，若查明是过失或错误的的数据，如采样、分析测试操作错误或发生意外污染等，应予舍去。

6.3.2.2 Grubbs 法、Dixon 法、T 法仅适用于来自正态总体的样本。若来自对数正态总体，应先将数据取对数，然后对对数数据样本实施上述判别检验。对于非正态（或对数正态）分布的数据样本可以采用箱线图法进行判别。

6.3.2.3 Dixon 法仅适用于样本容量不大于 100 的样本，Grubbs 法和 T 法两种方法对大、小样本都适用。

6.3.2.4 实际使用中，一般同时用 3 种方法判读异常值，最终结论“以多胜少”法得到。

6.3.2.5 当样本中被怀疑的异常值不止一个，应逐个判断，逐个剔除。先判断最大值是否异常，若是，剔除后再判断次大值是否异常，若是，再剔除，依此进行到所判断的不是异常值为止。

6.3.2.6 异常值可能属于外来污染，也可能来自高背景区。当异常值来自高背景区，则不能轻易地剔除，应考虑取样的实际情况，或结合富集系数等方法进行判断。

## 7 区域性土壤环境背景含量统计与表征

### 7.1 区域性土壤环境背景含量统计

7.1.1 对于异常值处理后的数据，再检验数据分布类型，进行区域性土壤环境背景含量统计。

7.1.2 计算样点数量、最小值、最大值、分位数（2.5%、5%、10%、25%、50%、75%、90%、95%、97.5%、99%）、算术平均值 $\bar{x}$ 、算术标准差 $S$ 、 $\bar{x}+2S$ 、几何平均值 $M$ 、几何标准差 $D$ 、 $M \times D^2$ 等统计量。

## 7.2 区域性土壤环境背景含量表征

7.2.1 用图件和表相结合的形式表征区域性土壤环境背景含量，并给予必要的说明。

7.2.1.1 图件为统计单元图。

7.2.1.2 表格应包括统计单元名称、样点数量、最小值、最大值、分位数（2.5%、5%、10%、25%、50%、75%、90%、95%、97.5%、99%）、 $\bar{x}$ 、 $S$ 、 $\bar{x} + 2S$ 、 $M$ 、 $D$ 、 $M \times D^2$ 和数据分布类型。

7.2.2 编制区域性土壤环境背景含量统计技术报告。

7.2.2.1 区域性土壤环境背景含量统计技术报告应按照本标准规定的工作程序进行编制。

7.2.2.2 区域性土壤环境背景含量统计技术报告内容要完整、详细。内容包括工作目标、工作程序、数据资料收集与整理、区域性土壤环境背景调查、历史数据评估、数据处理分析、区域性土壤环境背景含量统计与表征，记录区域性土壤环境背景含量统计过程。

## 8 区域性土壤环境背景值的确定

依据区域性土壤环境背景含量统计量，选择某个统计量或一组统计量作为背景值：

- a) 数据为正态或接近正态分布时，则一般取 $\bar{x} + 2S$ 作为土壤环境背景值。
- b) 数据为对数正态或近似正态分布时，则一般取 $M \times D^2$ 作为土壤环境背景值。
- c) 数据为其他分布的，则一般取97.5%分位数作为土壤环境背景值。
- d) 在应用过程中，可结合用途和实际情况，对标准差的倍数、分位数进行适当调整，如采用 $\bar{x} + 3S$ 、 $M \times D^3$ 、99%分位数等作为背景值。

## 9 实施与监督

9.1 本标准由地方生态环境主管部门及其他相关主管部门负责监督实施。

附录 A  
(资料性附录)  
t 分布表

单侧	75%	80%	85%	90%	95%	97.50%	99%	99.50%	99.75%	99.90%	99.95%
双侧	50%	60%	70%	80%	90%	95%	98%	99%	99.50%	99.80%	99.90%
1	1	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.92	4.303	6.965	9.925	14.09	22.33	31.6
3	0.765	0.978	1.25	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.19	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5	0.727	0.92	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.86	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.1	1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10	0.7	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.93	4.318
13	0.694	0.87	1.079	1.35	1.771	2.16	2.65	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.14
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.69	0.865	1.071	1.337	1.746	2.12	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.74	2.11	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.33	1.734	2.101	2.552	2.878	3.197	3.61	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.86	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.85
21	0.686	0.859	1.063	1.323	1.721	2.08	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.06	1.319	1.714	2.069	2.5	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.06	2.485	2.787	3.078	3.45	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.69
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.31	1.697	2.042	2.457	2.75	3.03	3.385	3.646
40	0.681	0.851	1.05	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2	2.39	2.66	2.915	3.232	3.46
80	0.678	0.846	1.043	1.292	1.664	1.99	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.29	1.66	1.984	2.364	2.626	2.871	3.174	3.39
120	0.677	0.845	1.041	1.289	1.658	1.98	2.358	2.617	2.86	3.16	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.96	2.326	2.576	2.807	3.09	3.291

附录 B  
(资料性附录)  
异常值判别方法

### 1 格拉布斯 (Grubbs) 检验法

根据公式 (1) 计算 G 值:

$$G = \frac{|x_k - \bar{x}|}{s} \quad (1)$$

其中:

$x_k$  为被怀疑的异常值,  $\bar{x}$ 、 $S$  分别为包括  $x_k$  在内的样本平均值和标准差。

根据样本容量  $n$  的值查 Grubbs 检验临界值表 B.1, 得到临界值  $g_\alpha$ 。

检验结果如下:

若  $G \geq g_\alpha$ , 则  $x_k$  为异常值; 若  $G < g_\alpha$ , 则  $x_k$  不是异常值。

表 B.1 格拉布斯 (Grubbs) 检验的临界值表

$n$	0.9	0.95	0.975	0.99	0.995	$n$	0.9	0.95	0.975	0.99	0.995
3	1.148	1.153	1.155	1.155	1.155	52	2.783	2.971	3.143	3.353	3.500
4	1.425	1.463	1.481	1.492	1.496	53	2.790	2.978	3.151	3.361	3.007
5	1.602	1.672	1.715	1.749	1.764	54	2.798	2.986	3.158	3.368	3.516
6	1.729	1.822	1.887	1.944	1.973	55	2.804	2.992	3.166	3.376	3.524
7	1.828	1.938	2.020	2.097	2.139	56	2.811	3.000	3.172	3.383	3.531
8	1.909	2.032	2.126	2.221	2.274	57	2.818	3.006	3.180	3.391	3.539
9	1.977	2.110	2.215	2.323	2.387	58	2.824	3.013	3.186	3.397	3.546
10	2.036	2.176	2.290	2.410	2.482	59	2.831	3.019	3.193	3.405	3.553
11	2.088	2.234	2.355	2.485	2.564	60	2.837	3.025	3.199	3.411	3.560
12	2.134	2.285	2.412	2.550	2.636	61	2.842	3.032	3.205	3.418	3.566
13	2.175	2.331	2.462	2.607	2.699	62	2.849	3.037	3.212	3.424	3.573
14	2.213	2.371	2.507	2.659	2.755	63	2.854	3.044	3.218	3.430	3.579
15	2.247	2.409	2.549	2.705	2.806	64	2.860	3.049	3.224	3.437	3.586
16	2.279	2.443	2.585	2.747	2.852	65	2.866	3.055	3.230	3.442	3.592
17	2.309	2.475	2.620	2.785	2.894	66	2.871	3.061	3.235	3.449	3.598
18	2.335	2.504	2.651	2.821	2.932	67	2.877	3.066	3.241	3.454	3.605
19	2.361	2.532	2.681	2.854	2.968	68	2.883	3.071	3.246	3.460	3.610
20	2.385	2.557	2.709	2.884	3.001	69	2.888	3.076	3.252	3.466	3.617
21	2.408	2.580	2.733	2.912	3.031	70	2.893	3.082	3.257	3.471	3.622
22	2.429	2.603	2.758	2.939	3.060	71	2.897	3.087	3.262	3.476	3.627
23	2.448	2.624	2.781	2.963	3.087	72	2.903	3.092	3.267	3.482	3.633
24	2.467	2.644	2.802	2.987	3.112	73	2.908	3.098	3.272	3.487	3.638
25	2.486	2.663	2.822	3.009	3.135	74	2.912	3.102	3.278	3.492	3.643
26	2.502	2.681	2.841	3.029	3.157	75	2.917	3.107	3.282	3.496	3.648

$n$	0.9	0.95	0.975	0.99	0.995	$n$	0.9	0.95	0.975	0.99	0.995
27	2.519	2.698	2.859	3.049	3.178	76	2.922	3.111	3.287	3.502	3.654
28	2.534	2.714	2.876	3.068	3.199	77	2.927	3.117	3.291	3.507	3.658
29	2.549	2.730	2.893	3.085	3.218	78	2.931	3.121	3.297	3.511	3.663
30	2.563	2.745	2.908	3.103	3.236	79	2.935	3.125	3.301	3.516	3.669
31	2.577	2.759	2.924	3.119	3.253	80	2.940	3.130	3.305	3.521	3.673
32	2.591	2.773	2.938	3.135	3.270	81	2.945	3.134	3.309	3.525	3.677
33	2.604	2.786	2.952	3.150	3.286	82	2.949	3.139	3.315	3.529	3.682
34	2.616	2.799	2.965	3.164	3.301	83	2.953	3.143	3.319	3.534	3.687
35	2.628	2.811	2.979	3.178	3.316	84	2.957	3.147	3.323	3.539	3.691
36	2.639	2.823	2.991	3.191	3.330	85	2.961	3.151	3.327	3.543	3.695
37	2.650	2.835	3.003	3.204	3.343	86	2.966	3.155	3.331	3.547	3.699
38	2.661	2.846	3.014	3.216	3.356	87	2.970	3.160	3.335	3.551	3.704
39	2.671	2.857	3.025	3.228	3.369	88	2.973	3.163	3.339	3.555	3.708
40	2.682	2.866	3.036	3.240	3.381	89	2.977	3.167	3.343	3.559	3.712
41	2.692	2.877	3.046	3.251	3.393	90	2.981	3.171	3.347	3.563	3.716
42	2.700	2.887	3.057	3.261	3.404	91	2.984	3.174	3.350	3.567	3.720
43	2.710	2.896	3.067	3.271	3.415	92	2.989	3.179	3.355	3.570	3.725
44	2.719	2.905	3.075	3.282	3.425	93	2.993	3.182	3.358	3.575	3.728
45	2.727	2.914	3.085	3.292	3.435	94	2.996	3.186	3.362	3.579	3.732
46	2.736	2.923	3.094	3.302	3.445	95	3.000	3.189	3.365	3.582	3.736
47	2.744	2.931	3.103	3.310	3.455	96	3.003	3.193	3.369	3.586	3.739
48	2.753	2.940	3.111	3.319	3.464	97	3.006	3.196	3.372	3.589	3.744
49	2.760	2.948	3.120	3.329	3.474	98	3.011	3.201	3.377	3.593	3.747
50	2.768	2.956	3.128	3.336	3.483	99	3.014	3.204	3.380	3.597	3.750
51	2.775	2.964	3.136	3.345	3.491	100	3.017	3.207	3.383	3.600	3.754

## 2 狄克逊 (Dixon) 检验法

根据下式计算  $D$  值:

当 $3 \leq n \leq 7$ 时,	$D = \frac{x_n - x_{n-1}}{x_n - x_1}$	怀疑最大值 (2)
	$D = \frac{x_2 - x_1}{x_n - x_1}$	怀疑最小值 (3)
当 $8 \leq n \leq 10$ 时,	$D = \frac{x_n - x_{n-1}}{x_n - x_2}$	怀疑最大值 (4)
	$D = \frac{x_2 - x_1}{x_{n-1} - x_1}$	怀疑最小值 (5)
当 $11 \leq n \leq 13$ 时,	$D = \frac{x_n - x_{n-2}}{x_n - x_2}$	怀疑最大值 (6)
	$D = \frac{x_3 - x_1}{x_{n-1} - x_1}$	怀疑最小值 (7)
当 $14 \leq n \leq 30$ 时,	$D = \frac{x_n - x_{n-2}}{x_n - x_3}$	怀疑最大值 (8)
	$D = \frac{x_3 - x_1}{x_{n-2} - x_1}$	怀疑最小值 (9)
当 $31 \leq n \leq 100$ 时,	$D = \frac{x_n - x_{n-2}}{x_n - x_3}$	怀疑最大值 (10)
	$D = \frac{x_3 - x_1}{x_{n-2} - x_1}$	怀疑最小值 (11)

根据样本容量 $n$ 的值查 Dixon 检验临界值表 B.2, 得到  $D_\alpha$  值。检验结论如下:  
若  $D \geq D_\alpha$ , 则  $x_n$  (或  $x_1$ ) 为异常值; 若  $D \leq D_\alpha$ , 则  $x_n$  (或  $x_1$ ) 不是异常值。

表 B.2 狄克逊 (Dixon) 检验的临界值表

$n$	0.9	0.95	0.99	0.995	$n$	0.9	0.95	0.99	0.995
3	0.885	0.941	0.988	0.994	52	0.269	0.308	0.379	0.405
4	0.679	0.765	0.889	0.920	53	0.267	0.306	0.377	0.402
5	0.557	0.642	0.782	0.823	54	0.265	0.304	0.375	0.400
6	0.484	0.562	0.698	0.744	55	0.264	0.302	0.373	0.398
7	0.434	0.507	0.637	0.680	56	0.262	0.300	0.371	0.396
8	0.479	0.554	0.681	0.723	57	0.261	0.298	0.369	0.394
9	0.441	0.512	0.635	0.676	58	0.259	0.297	0.367	0.392
10	0.410	0.477	0.597	0.638	59	0.258	0.295	0.366	0.391
11	0.517	0.575	0.674	0.707	60	0.256	0.294	0.363	0.388
12	0.490	0.546	0.642	0.675	61	0.255	0.292	0.362	0.387
13	0.467	0.521	0.617	0.649	62	0.253	0.291	0.361	0.385
14	0.491	0.546	0.640	0.672	63	0.252	0.289	0.359	0.383
15	0.470	0.524	0.618	0.649	64	0.251	0.288	0.357	0.382
16	0.453	0.505	0.597	0.629	65	0.250	0.287	0.355	0.380
17	0.437	0.489	0.580	0.611	66	0.249	0.285	0.354	0.379
18	0.424	0.475	0.564	0.595	67	0.247	0.284	0.353	0.377
19	0.412	0.462	0.550	0.580	68	0.246	0.283	0.351	0.376
20	0.401	0.450	0.538	0.568	69	0.245	0.282	0.350	0.374
21	0.391	0.440	0.526	0.556	70	0.244	0.280	0.348	0.372
22	0.382	0.431	0.516	0.545	71	0.243	0.279	0.347	0.371
23	0.374	0.422	0.507	0.536	72	0.242	0.278	0.346	0.370
24	0.367	0.413	0.497	0.526	73	0.241	0.277	0.344	0.368
25	0.360	0.406	0.489	0.519	74	0.240	0.276	0.343	0.368
26	0.353	0.399	0.482	0.510	75	0.239	0.275	0.342	0.366
27	0.347	0.393	0.474	0.503	76	0.238	0.274	0.341	0.365
28	0.341	0.387	0.468	0.496	77	0.237	0.273	0.340	0.364
29	0.337	0.381	0.462	0.489	78	0.236	0.272	0.338	0.363
30	0.332	0.376	0.456	0.484	79	0.235	0.271	0.337	0.361
31	0.327	0.371	0.450	0.478	80	0.234	0.270	0.336	0.360
32	0.323	0.367	0.445	0.473	81	0.233	0.269	0.335	0.359
33	0.319	0.362	0.441	0.468	82	0.232	0.268	0.334	0.358
34	0.315	0.358	0.436	0.463	83	0.232	0.267	0.333	0.356
35	0.311	0.354	0.432	0.458	84	0.231	0.266	0.332	0.356
36	0.308	0.350	0.427	0.454	85	0.230	0.265	0.331	0.355
37	0.305	0.347	0.423	0.450	86	0.229	0.264	0.330	0.353
38	0.301	0.343	0.419	0.446	87	0.228	0.263	0.329	0.352
39	0.298	0.340	0.416	0.442	88	0.228	0.262	0.328	0.352

$n$	0.9	0.95	0.99	0.995	$n$	0.9	0.95	0.99	0.995
40	0.296	0.337	0.413	0.439	89	0.227	0.262	0.327	0.351
41	0.293	0.334	0.409	0.435	90	0.226	0.261	0.326	0.350
42	0.290	0.331	0.406	0.432	91	0.225	0.260	0.325	0.349
43	0.288	0.328	0.403	0.429	92	0.225	0.259	0.324	0.348
44	0.285	0.326	0.400	0.425	93	0.224	0.259	0.323	0.347
45	0.283	0.323	0.397	0.423	94	0.223	0.258	0.323	0.346
46	0.281	0.321	0.394	0.420	95	0.223	0.257	0.322	0.345
47	0.279	0.318	0.391	0.417	96	0.222	0.256	0.321	0.344
48	0.277	0.316	0.389	0.414	97	0.221	0.255	0.320	0.344
49	0.275	0.314	0.386	0.412	98	0.221	0.255	0.320	0.343
50	0.273	0.312	0.384	0.409	99	0.220	0.254	0.319	0.341
51	0.271	0.310	0.382	0.407	100	0.219	0.254	0.318	0.341

### 3 T (Thompson) 检验法

根据下式计算：

$$T = \frac{|x_k - \bar{x}|}{S \sqrt{\frac{n}{n-1}}} \quad (12)$$

式中： $x_k$ 为被怀疑的异常值； $\bar{x}$ 、 $S$ 分别为不包括被怀疑的异常值 $x_k$ 的样本平均值和标准差， $n$ 为样本容量（包括 $x_k$ ）。

根据自由度  $df=n-2$  查  $t$  分布表，得临界值  $t_\alpha$ ，检验结论如下：

若  $T \geq t_\alpha$ ，则  $x_k$  为异常值；若  $T \leq t_\alpha$ ，则  $x_k$  不是异常值。

### 4 箱线图法

为了以简单和可视化的方式检验异常值，通常使用箱线图（Boxplot）（见图 B.1）。箱线图也称箱须图（Box-whisker Plot），是利用数据中的五个统计量：最小值、第一四分位数（ $Q_1$ ）、中位数（ $x_m$ ）、第三四分位数（ $Q_3$ ）与最大值来描述数据的一种方法。

箱形图的绘制步骤：

- 画数轴，度量单位大小和数据批的单位一致，起点比最小值稍小，长度比该数据批的全距稍长。
- 画一个矩形盒，两端边的位置分别对应数据批的上下四分位数（ $Q_3$ 和 $Q_1$ ）。在矩形盒内部中位数（ $x_m$ ）位置画一条线段为中位线。
- 四分位距  $IQR = Q_3 - Q_1$ ，在  $Q_3 + 1.5IQR$  和  $Q_3 - 1.5IQR$  处画两条与中位线一样的线段，这两条线段为异常值截断点，称其为内限；在  $Q_3 + 3IQR$  和  $Q_3 - 3IQR$  处画两条线段，称其为外限。处于内限以外位置的点表示的数据都是异常值，其中在内限与外限之间的异常值为温和的异常值（mild outliers），在外限以外的为极端的异常值（extreme outliers）。
- 从矩形盒两端边向外各画一条线段直到不是异常值的最远点，表示该批数据正常值的分布区间。
- 用“●”标出温和的异常值，用“\*”标出极端的异常值。相同值的数据点并列标出在同一数据线位置上，不同值的数据点标在不同数据线位置上。

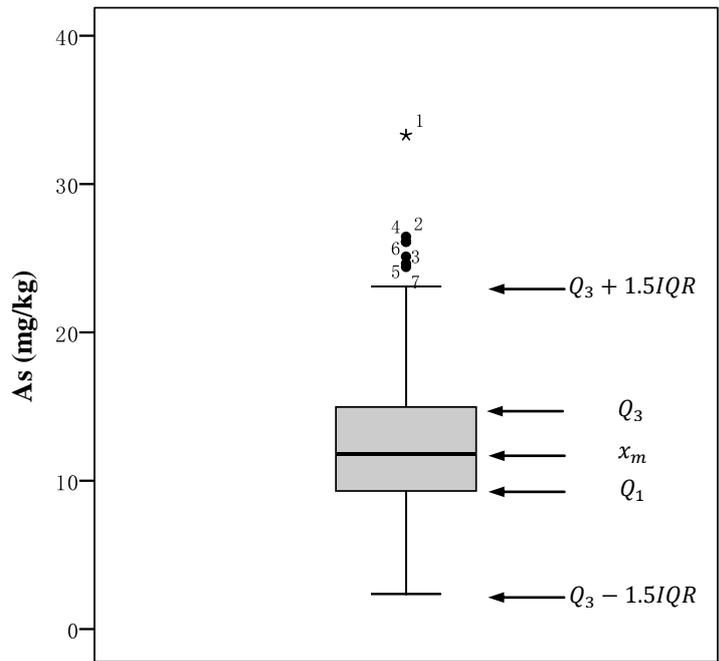


图 B.1 箱线图 中的砷浓度

### 5 富集系数法

由于TiO<sub>2</sub>或Al<sub>2</sub>O<sub>3</sub>的矿物具有高度抗风化性和难于迁移性，在土壤中含量较高，容易测定以及较少受到外来污染影响等特点，因此可根据TiO<sub>2</sub>或Al<sub>2</sub>O<sub>3</sub>含量判断土壤是否被其他元素污染。方法是以TiO<sub>2</sub>或Al<sub>2</sub>O<sub>3</sub>含量作参比，计算其他元素的富集系数，公式为：

$$\text{富集系数} = \frac{\text{土壤中元素含量} / \text{土壤中TiO}_2(\text{或Al}_2\text{O}_3) \text{ 含量}}{\text{母岩中元素含量} / \text{母岩中TiO}_2(\text{或Al}_2\text{O}_3) \text{ 含量}} \quad (13)$$

若富集系数明显大于 1，则该元素有外来污染；若富集系数明显小于 1，则元素富集已受淋溶；若富集系数近似等于 1，则该元素未受污染和淋溶。

富集系数法要求每一个土壤剖面的土层与下层母岩同源。