

中华人民共和国国家生态环境标准

HJ 1185—2021

区域性土壤环境背景含量统计 技术导则（试行）

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

本电子版为正式标准文本，由生态环境部环境标准研究所审校排版。

2021-07-06 发布

2021-08-01 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 工作目标和内容.....2

5 数据获取2

6 数据处理分析.....7

7 统计与表征8

附录 A（资料性附录） t 分布表.....10

附录 B（资料性附录） 异常值判别方法11



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。

本标准为首次发布。

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。

本标准生态环境部 2021 年 7 月 6 日批准。

本标准自 2021 年 8 月 1 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（试行）

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，地块尺度土壤环境背景含量统计的相关标准另行制定。

2 规范性引用文件

本标准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标准。凡是未注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 4882	数据的统计处理和解释 正态性检验
GB/T 8170	数值修约规则与极限数值的表示和判定
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 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

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

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。

3.2

调查单元 survey unit

调查区域按照土壤类型、成土母质（岩）类型、流域、行政区域或土地利用类型等划分的空间单元，同一调查单元，可能在空间上不连续分布。

3.3

统计单元 statistical unit

用于土壤环境背景含量数据统计的单元。

3.4

已有数据 existing data

在开展土壤环境背景含量统计时，调查区域已有的、目标时间范围内的土壤环境调查数据。

4 工作目标和内容

4.1 工作目标

4.1.1 明确调查区域、目标时间范围和目标元素或化合物。统计与表征调查区域目标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。平面范围可以是行政区域，也可以是行政区域内一定土壤类型、成土母质（岩）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 目标时间范围可以是历史上某一段时间段，也可以是当前一定时间段，此时间范围内土壤中目标元素或化合物未发生超预期的富集、损失或含量水平的变化。

4.1.4 土壤中目标元素或化合物可以是某一种、某一类或多种。

4.2 工作程序及内容

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 工作内容包括数据获取、数据处理分析、统计与表征。

4.2.2.1 数据获取

收集整理区域性土壤环境背景含量统计所需的相关数据资料，明确是否存在已有数据：

a) 当存在已有数据时，应进行点位数据评估，若存在满足要求的点位数据，还应进行数据集评估。

b) 当不存在已有数据或已有数据不能完全满足要求时，如目标时间范围为当前一定时间段，应开展或结合已有数据开展土壤环境背景调查以获取数据，如目标时间范围为历史上某一段时间段，则已有数据仅作参考。

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) 检验统计单元土壤环境背景含量数据的分布类型，判别和处理异常值。

4.2.2.3 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 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、土地利用及其变化图件、环境数据资料等，图件的比例尺应与调查区域相匹配。

5.1.1.2 自然和社会经济信息：包括地形、地貌、植被、土壤类型、成土母质（岩）类型、水文、气候和矿产资源分布等。

5.1.1.3 土地利用及其变化图件：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、环境质量公报等）、污染源信息（如工农业生产及排污）、农业投入品使用情况（如农药、肥料、土壤调

理剂、农膜等)、已有土壤环境调查数据等。

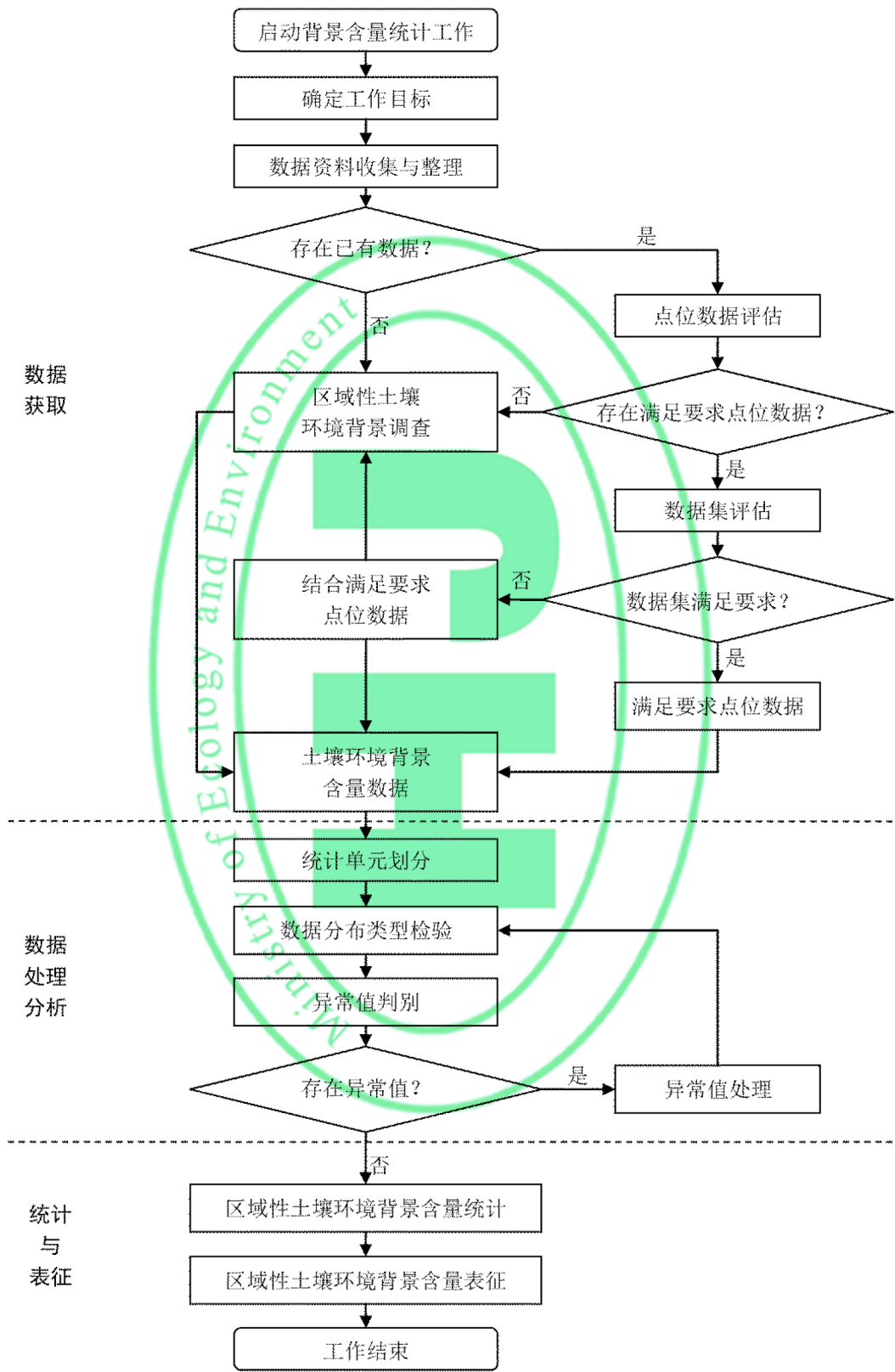


图 1 区域性土壤环境背景含量统计工作程序

5.1.2 资料整理

5.1.2.1 根据专业知识和经验识别处理资料中的错误和不合理信息。

5.1.2.2 对收集的资料进行归类整理，如存在所需已有数据，应建立由点位数据组成的数据集。

5.2 已有数据评估

5.2.1 点位数据评估

5.2.1.1 数据的完整性评估

a) 已有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、成土母质（岩）类型、土地利用类型和作物类型）。

b) 已有数据达到以上信息项要求的，则其完整性满足要求。

5.2.1.2 数据的符合性评估

a) 对满足完整性要求的已有数据，进行数据的符合性评估。

b) 对照工作目标对时间范围、采样层次或深度、目标元素或化合物的相关要求，对已有数据进行符合性评估。

c) 已有数据的采样日期、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。

5.2.1.3 数据的规范性评估

a) 对满足符合性要求的已有数据，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与质量保证的相关要求，评估数据的规范性。

c) 已有数据的布点方法、点位选择、样品分析测试、质量控制与质量保证符合本标准 5.4 相关要求的，则其规范性满足要求。

5.2.2 数据集评估

5.2.2.1 不同来源数据的合并

a) 应评估合并不同来源数据可增加样本代表性的优点与可能引入不准确性的缺点，数据的合并可能需作出数据准确性上的妥协。

b) 应考虑合并后点位数据的空间分布相对均匀性，点位数据存在聚集的，应进行去聚集化处理，综合考虑点位的均匀性和代表性，尽可能随机选择保留点位。

c) 按照发生层次与按照固定深度两种不同土壤样品采集方法获得的数据不宜合并。

d) 对于来自不同分析测试方法的数据，应确保不同分析测试方法之间的等效性。

5.2.2.2 按照本标准 5.4.2 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。

5.2.2.3 当各调查单元内满足要求的点位数量达到本标准 5.4.3.2 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。

5.3 结合已有数据开展区域性土壤环境背景调查

5.3.1 当数据集不满足要求，但存在满足要求的点位数据且目标时间范围为当前一定时间段时，可结合这些点位数据按照本标准 5.4 开展调查。

5.3.2 在调查过程中应考虑与满足要求的已有点位数据的衔接：

- 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。

5.4 区域性土壤环境背景调查

5.4.1 调查项目确定

5.4.1.1 调查项目可以是土壤中各种元素或化合物，需根据工作目标确定。

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

5.4.2 调查单元划分

5.4.2.1 在调查区域内，按照影响该区域土壤环境背景含量的主导因素土壤类型和成土母质（岩）类型等划分调查单元。土壤类型和成土母质（岩）类型图件的比例尺应相对一致。土壤类型采用 GB/T 17296 规定的分类方法，可按照土类、亚类、土属、土种的顺序，结合工作目标进行逐级细化。

5.4.2.2 当工作目标为统计与表征调查区域内不同流域、行政区或土地利用类型的土壤环境背景含量时，在调查区域内，按照流域、行政区、土地利用类型等因素或综合主导因素划分调查单元。

5.4.2.3 当工作目标为统计与表征整个调查区域的土壤环境背景含量时，将整个调查区域作为一个调查单元。

5.4.3 调查点位布设

5.4.3.1 基础样本数量

由变异系数和相对偏差计算样本数量，计算公式如（1）：

$$N = \frac{t^2 C_v^2}{m^2} \quad (1)$$

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）；

C_v ——变异系数，从已有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，%。

已开展过背景含量调查的区域，可以参考当时的 C_v ；没有开展过背景含量调查的区域，可参考上一级区域的 C_v ；没有开展过背景含量调查且土壤变异程度较小的区域，一般 C_v 可用 10%~30% 粗略估计，如土壤变异程度较大，应开展初步调查进行估计。

5.4.3.2 布点数量

a) 实际工作中土壤布点数量要根据调查目的、调查精度和调查区域环境状况等因素确定。

b) 各调查单元的布点数量应同时满足基础样本数量和统计单元最少样本量 30 个的要求。对于以历史上某一时间段为目标时间范围的，若点位数不满足要求，应注明点位数量，所得结果作为参考。

- c) 考虑到土壤变异的不确定性和可能出现异常值等因素，为保证统计数据的有效性，布点数量宜适度增加。
- d) 获取土壤环境背景含量数据后，若实际变异系数大于设定 C_v 且反算的 m 值不可接受时，应考虑补充点位后重新统计。
- e) 当调查项目为多个元素或化合物时，可按照最大变异系数来确定布点数量，协调布设调查点位。

5.4.3.3 布点方法

针对调查区域内的调查单元，一般采用系统布点或系统随机布点方法并结合专业判断进行布点，在保证样点相对均匀分布的情况下，也可以单独采用专业判断布点法。

- a) 系统布点：将调查单元划分成面积相等的网格，网格的数量等于布点数量，每个网格内布设 1 个采样点。适用于既有认知较少，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，也适用于同一土壤类型、流域类型。
- b) 系统随机布点：将调查单元划分成面积相等的网格，网格的数量大于布点数量 2 倍，对每个网格进行编号，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。可以利用掷骰子、抽签、查随机数表的方法随机抽取网格。适用于土壤类型、成土母质（岩）类型单一，地形相对平坦，其他自然因素差异较小的区域。当随机布设的点位分布不均时，需适当增加或调整点位。
- c) 网格间距 L 按公式（2）计算：

$$L = (A/N)^{1/2} \tag{2}$$

式中： L ——网格间距，km 或 m；

A ——调查单元面积， km^2 或 m^2 ；

N ——布点数量（同“5.4.3.2 布点数量”），个。

A 和 L 的量纲要相匹配，如 A 的单位是 km^2 则 L 的单位就为 km。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、河流或裸露岩石上。当网格中所关注类型为非主导类型或面积占比较小时，宜结合点位数量和空间分布情况，考虑取消或调整该点位。

d) 专业判断布点：根据专家判断，选择具有代表性样点进行采样。此方法适用于对调查单元有较充分了解的情形，选择具有代表性样点需要有一定的土壤分类和地质学知识和经验。

5.4.3.4 点位选择

点位选择需要具有代表性，尽量避开人为活动的影响。

- a) 采样点自然景观应符合土壤环境背景研究的要求。采样点选在所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，一般不布设采样点。
- b) 不宜在多种土类、多种成土母质（岩）交错分布、面积较小的边缘地区布设采样点。
- c) 采样点以剖面发育完整、层次较清楚为准，不在水土流失严重或表土被破坏处设采样点。
- d) 现状及历史上的城镇、住宅、工矿企业、交通运输、水利设施、殡葬、粪坑等人为干扰大的区域及其影响范围内不宜设采样点，其对周边的影响范围应根据实际情况进行综合判断，工矿企业、交通运输的影响范围可参考表 1 初步确定。
- e) 农用地一般应在播种和施肥前或农作物成熟后采样，采样点尽量避免在肥料、农药集中使用位置，如坑施和条施位置，以使样点尽可能少受人为活动的影响。

表 1 土壤环境背景含量主要人为影响源影响范围

主要人为影响源	影响范围
工矿企业	工矿企业周边 5000 米

续表

主要人为影响源	影响范围
交通运输	铁路两侧 500 米
	二级以上公路两侧 350 米
	农村道路两侧 50 米

5.4.4 样品采集

- 5.4.4.1 一般按照发生层次采集剖面土壤样品，或根据工作目标特殊需求，采集固定深度土壤样品，同一调查区域应采用相同的样品采集方法。
- 5.4.4.2 剖面土壤样品采集和采样记录参照 HJ/T 166 执行。
- 5.4.4.3 固定深度土壤样品一般按照 0~20 cm、40~60 cm 和 80~100 cm 中的一个或多个深度进行采集，也可以根据工作目标增加采集其他固定深度的土壤样品。以钻孔取样为主，也可以采用槽探的方式进行，采样方法参照 HJ 25.2 执行，样品采集量、采样记录等参照剖面土壤样品采集。

5.4.5 样品保存与流转

样品保存与流转按照 HJ/T 166 中的相关规定执行。

5.4.6 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，其中标准方法优先采用 GB 15618 和 GB 36600 中的分析方法，GB 15618 和 GB 36600 中污染物项目以外的分析方法选用现行有效的国家、行业标准方法。除上述分析方法标准外，本标准实施后发布的其他污染物分析方法标准，如明确适用于土壤样品分析测试，也可采用该分析方法标准。

5.4.7 质量控制与质量保证

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，质量控制与质量保证应涉及调查的全部过程。质量控制与质量保证参照 HJ/T 166 中相关规定执行。

6 数据处理分析

6.1 统计单元划分

- 6.1.1 按照影响土壤环境背景含量的主导因素将土壤环境背景含量数据划分为不同统计单元，使统计单元内土壤环境背景含量的变异性相对较小。
- 6.1.2 可根据工作目标要求，按照划分调查单元的流域、行政区或土地利用类型等划分统计单元。
- 6.1.3 基于统计单元内每层土壤环境背景含量数据，进行数据分布类型检验、异常值判别与处理以及统计与表征，未检出值按检出限的一半参与统计。

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 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判别与处理应注意以下事项:

6.3.2.1 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。对于其他分布的数据样本可以采用箱线图法进行判别。

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

6.3.2.3 对于所判断的异常值,按照以下方式进行处理:

a) 检查原始记录,若是过失或错误的的数据,如样品采集、分析检测、数据输入错误等原因导致的异常数据,应予更正或删除。

b) 考察取样的实际情况,根据目标元素或化合物含量特征和组合特征综合分析,或结合富集系数等方法判断异常原因。若异常值来源于污染,则剔除;若来源于高背景,应予以保留。

c) 若判别出的异常值不止一个,按异常值数字从大到小的顺序逐个判断,逐个处理。

d) 对于呈现多峰的数据,应根据实际情况判断异常值原因,谨慎处理。

6.3.2.4 被剔除或更正的异常值及其理由应予记录,以备查询。处理后样本数量不能满足统计要求的,应补充样本数据。

7 统计与表征

7.1 区域性土壤环境背景含量统计

7.1.1 对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 统计样点数量、最小值、最大值、分位数(2.5%、5%、10%、25%、50%、75%、90%、95%、97.5%)、算术平均值 $\bar{x}$ 、算术标准差 S 、 $\bar{x}+2S$ 、 $\bar{x}-2S$ 、几何平均值 M 、几何标准差 D 、 M/D^2 、 MD^2 等统计量。算术平均值 $\bar{x}$ 的计算公式如(3),算术标准差 S 的计算公式如(4),几何平均值 M 的计算公式如(5),几何标准差 D 的计算公式如(6):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (4)$$

$$M = \log^{-1} \left(\frac{1}{n} \sum_{i=1}^n x'_i \right) \quad (5)$$

$$D = \log^{-1} \left(\sqrt{\frac{\sum_{i=1}^n \left(x'_i - \frac{1}{n} \sum_{i=1}^n x'_i \right)^2}{n-1}} \right) \quad (6)$$

式中: n ——样本量,个;

x_i ——第 i 个样本值;

x'_i —— $\log x_i (i = 1, 2, \dots, n)$;

$\bar{x}$ ——算术平均值;

S ——算术标准差;

M ——几何平均值;

D ——几何标准差。

7.2 区域性土壤环境背景含量表征

7.2.1 用图件和表格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 绘制统计单元分布图，用不同颜色和编号对统计单元进行表达。

7.2.1.2 表格应包括统计单元名称、编号、样点数量、最小值、最大值、分位数（5%、10%、25%、50%、75%、90%、95%）、 $\bar{x}$ 、 S 、 M 、 D 、95%置信范围（正态分布采用 $\bar{x} \pm 2S$ ，对数正态分布采用 $M/D^2 \sim MD^2$ ，其他分布采用 2.5%~97.5%分位数）和数据分布类型。有效数字的修约规则按 GB/T 8170 执行。

7.2.1.3 对调查区域、调查项目、调查单元划分、采样时间、布点数量及方法、样品采集方法、统计单元划分、数据分布类型和异常值判别与处理的概况进行必要的说明。

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 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 技术报告内容要完整、详细。内容包括工作目标、工作程序、调查区域概况（包括自然地理条件、成土母质（岩）类型、土壤类型、土地利用类型等内容）、数据资料收集与整理、已有数据评估、区域性土壤环境背景调查、数据处理分析和统计与表征，记录区域性土壤环境背景含量统计过程。



附 录 A
(资料性附录)
t 分布表

df	置信水平 $(1 - \alpha) \%$ /双侧								
	50	60	70	80	90	95	98	99	99.5
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807

附 录 B
(资料性附录)
异常值判别方法

B.1 格拉布斯 (Grubbs) 检验法

当怀疑最大值为异常时, 根据公式 (1) 计算统计量 G_n :

$$G_n = \frac{x_n - \bar{x}}{S} \quad (1)$$

当怀疑最小值为异常时, 根据公式 (2) 计算统计量 G'_n :

$$G'_n = \frac{\bar{x} - x_1}{S} \quad (2)$$

式中: x_1 为数据经排序后的最小值, x_n 为数据经排序后的最大值, $\bar{x}$ 、 S 分别为样本平均值和标准差。

确定检出水平 α , 根据样本量 n 在表 B.1 中查出临界值; 如果根据实际情况或以往经验, 异常值为单侧情形, 查临界值 $G_{1-\alpha}$; 如果根据实际情况或以往经验异常值可为高端值, 也可以为低端值的双侧情形, 查临界值 $G_{1-\alpha/2}$ 。

检验结果如下:

对于单侧情形, 若 G_n 或 $G'_n > G_{1-\alpha}$, 则 x_n 或 x_1 为异常值。

对于双侧情形, 若 $G_n > G'_n$ 且 $G_n > G_{1-\alpha/2}$, 则 x_n 为异常值; 若 $G'_n > G_n$ 且 $G'_n > G_{1-\alpha/2}$, 则 x_1 为异常值。

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	19	2.361	2.532	2.681	2.854	2.968
4	1.425	1.463	1.481	1.492	1.496	20	2.385	2.557	2.709	2.884	3.001
5	1.602	1.672	1.715	1.749	1.764	21	2.408	2.580	2.733	2.912	3.031
6	1.729	1.822	1.887	1.944	1.973	22	2.429	2.603	2.758	2.939	3.060
7	1.828	1.938	2.020	2.097	2.139	23	2.448	2.624	2.781	2.963	3.087
8	1.909	2.032	2.126	2.221	2.274	24	2.467	2.644	2.802	2.987	3.112
9	1.977	2.110	2.215	2.323	2.387	25	2.486	2.663	2.822	3.009	3.135
10	2.036	2.176	2.290	2.410	2.482	26	2.502	2.681	2.841	3.029	3.157
11	2.088	2.234	2.355	2.485	2.564	27	2.519	2.698	2.859	3.049	3.178
12	2.134	2.285	2.412	2.550	2.636	28	2.534	2.714	2.876	3.068	3.199
13	2.175	2.331	2.462	2.607	2.699	29	2.549	2.730	2.893	3.085	3.218
14	2.213	2.371	2.507	2.659	2.755	30	2.563	2.745	2.908	3.103	3.236
15	2.247	2.409	2.549	2.705	2.806	31	2.577	2.759	2.924	3.119	3.253
16	2.279	2.443	2.585	2.747	2.852	32	2.591	2.773	2.938	3.135	3.270
17	2.309	2.475	2.620	2.785	2.894	33	2.604	2.786	2.952	3.150	3.286
18	2.335	2.504	2.651	2.821	2.932	34	2.616	2.799	2.965	3.164	3.301

续表

n	0.9	0.95	0.975	0.99	0.995	n	0.9	0.95	0.975	0.99	0.995
35	2.628	2.811	2.979	3.178	3.316	68	2.883	3.071	3.246	3.460	3.610
36	2.639	2.823	2.991	3.191	3.330	69	2.888	3.076	3.252	3.466	3.617
37	2.650	2.835	3.003	3.204	3.343	70	2.893	3.082	3.257	3.471	3.622
38	2.661	2.846	3.014	3.216	3.356	71	2.897	3.087	3.262	3.476	3.627
39	2.671	2.857	3.025	3.228	3.369	72	2.903	3.092	3.267	3.482	3.633
40	2.682	2.866	3.036	3.240	3.381	73	2.908	3.098	3.272	3.487	3.638
41	2.692	2.877	3.046	3.251	3.393	74	2.912	3.102	3.278	3.492	3.643
42	2.700	2.887	3.057	3.261	3.404	75	2.917	3.107	3.282	3.496	3.648
43	2.710	2.896	3.067	3.271	3.415	76	2.922	3.111	3.287	3.502	3.654
44	2.719	2.905	3.075	3.282	3.425	77	2.927	3.117	3.291	3.507	3.658
45	2.727	2.914	3.085	3.292	3.435	78	2.931	3.121	3.297	3.511	3.663
46	2.736	2.923	3.094	3.302	3.445	79	2.935	3.125	3.301	3.516	3.669
47	2.744	2.931	3.103	3.310	3.455	80	2.940	3.130	3.305	3.521	3.673
48	2.753	2.940	3.111	3.319	3.464	81	2.945	3.134	3.309	3.525	3.677
49	2.760	2.948	3.120	3.329	3.474	82	2.949	3.139	3.315	3.529	3.682
50	2.768	2.956	3.128	3.336	3.483	83	2.953	3.143	3.319	3.534	3.687
51	2.775	2.964	3.136	3.345	3.491	84	2.957	3.147	3.323	3.539	3.691
52	2.783	2.971	3.143	3.353	3.500	85	2.961	3.151	3.327	3.543	3.695
53	2.790	2.978	3.151	3.361	3.007	86	2.966	3.155	3.331	3.547	3.699
54	2.798	2.986	3.158	3.368	3.516	87	2.970	3.160	3.335	3.551	3.704
55	2.804	2.992	3.166	3.376	3.524	88	2.973	3.163	3.339	3.555	3.708
56	2.811	3.000	3.172	3.383	3.531	89	2.977	3.167	3.343	3.559	3.712
57	2.818	3.006	3.180	3.391	3.539	90	2.981	3.171	3.347	3.563	3.716
58	2.824	3.013	3.186	3.397	3.546	91	2.984	3.174	3.350	3.567	3.720
59	2.831	3.019	3.193	3.405	3.553	92	2.989	3.179	3.355	3.570	3.725
60	2.837	3.025	3.199	3.411	3.560	93	2.993	3.182	3.358	3.575	3.728
61	2.842	3.032	3.205	3.418	3.566	94	2.996	3.186	3.362	3.579	3.732
62	2.849	3.037	3.212	3.424	3.573	95	3.000	3.189	3.365	3.582	3.736
63	2.854	3.044	3.218	3.430	3.579	96	3.003	3.193	3.369	3.586	3.739
64	2.860	3.049	3.224	3.437	3.586	97	3.006	3.196	3.372	3.589	3.744
65	2.866	3.055	3.230	3.442	3.592	98	3.011	3.201	3.377	3.593	3.747
66	2.871	3.061	3.235	3.449	3.598	99	3.014	3.204	3.380	3.597	3.750
67	2.877	3.066	3.241	3.454	3.605	100	3.017	3.207	3.383	3.600	3.754

B.2 狄克逊（Dixon）检验法

根据表 B.2 中公式计算统计量 D_n 值。

表 B.2 D_n 值计算公式

当 $3 \leq n \leq 7$ 时,	怀疑最大值	$D_n = \frac{x_n - x_{n-1}}{x_n - x_1}$ (3)
	怀疑最小值	$D_n = \frac{x_2 - x_1}{x_n - x_1}$ (4)
当 $8 \leq n \leq 10$ 时,	怀疑最大值	$D_n = \frac{x_n - x_{n-1}}{x_n - x_2}$ (5)
	怀疑最小值	$D_n = \frac{x_2 - x_1}{x_{n-1} - x_1}$ (6)

续表

当 $11 \leq n \leq 13$ 时,	怀疑最大值	$D_n = \frac{x_n - x_{n-2}}{x_n - x_2}$ (7)
	怀疑最小值	$D_n = \frac{x_3 - x_1}{x_{n-1} - x_1}$ (8)
当 $14 \leq n \leq 30$ 时,	怀疑最大值	$D_n = \frac{x_n - x_{n-2}}{x_n - x_3}$ (9)
	怀疑最小值	$D_n = \frac{x_3 - x_1}{x_{n-2} - x_1}$ (10)
当 $31 \leq n \leq 100$ 时,	怀疑最大值	$D_n = \frac{x_n - x_{n-2}}{x_n - x_3}$ (11)
	怀疑最小值	$D_n = \frac{x_3 - x_1}{x_{n-2} - x_1}$ (12)

确定检出水平 α ，根据样本量 n 查出临界值 $D_{1-\alpha}$ ；如果为单侧情形，查表 B.3；如果为双侧情形，查表 B.4。检验结论如下：

若 $D_n > D_{1-\alpha}$ ，则 x_n （或 x_1 ）为异常值；若 $D_n \leq D_{1-\alpha}$ ，则 x_n （或 x_1 ）不是异常值。

表 B.3 单侧狄克逊（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	32	0.323	0.367	0.445	0.473
4	0.679	0.765	0.889	0.920	33	0.319	0.362	0.441	0.468
5	0.557	0.642	0.782	0.823	34	0.315	0.358	0.436	0.463
6	0.484	0.562	0.698	0.744	35	0.311	0.354	0.432	0.458
7	0.434	0.507	0.637	0.680	36	0.308	0.350	0.427	0.454
8	0.479	0.554	0.681	0.723	37	0.305	0.347	0.423	0.450
9	0.441	0.512	0.635	0.676	38	0.301	0.343	0.419	0.446
10	0.410	0.477	0.597	0.638	39	0.298	0.340	0.416	0.442
11	0.517	0.575	0.674	0.707	40	0.296	0.337	0.413	0.439
12	0.490	0.546	0.642	0.675	41	0.293	0.334	0.409	0.435
13	0.467	0.521	0.617	0.649	42	0.290	0.331	0.406	0.432
14	0.491	0.546	0.640	0.672	43	0.288	0.328	0.403	0.429
15	0.470	0.524	0.618	0.649	44	0.285	0.326	0.400	0.425
16	0.453	0.505	0.597	0.629	45	0.283	0.323	0.397	0.423
17	0.437	0.489	0.580	0.611	46	0.281	0.321	0.394	0.420
18	0.424	0.475	0.564	0.595	47	0.279	0.318	0.391	0.417
19	0.412	0.462	0.550	0.580	48	0.277	0.316	0.389	0.414
20	0.401	0.450	0.538	0.568	49	0.275	0.314	0.386	0.412
21	0.391	0.440	0.526	0.556	50	0.273	0.312	0.384	0.409
22	0.382	0.431	0.516	0.545	51	0.271	0.310	0.382	0.407
23	0.374	0.422	0.507	0.536	52	0.269	0.308	0.379	0.405
24	0.367	0.413	0.497	0.526	53	0.267	0.306	0.377	0.402
25	0.360	0.406	0.489	0.519	54	0.265	0.304	0.375	0.400
26	0.353	0.399	0.482	0.510	55	0.264	0.302	0.373	0.398
27	0.347	0.393	0.474	0.503	56	0.262	0.300	0.371	0.396
28	0.341	0.387	0.468	0.496	57	0.261	0.298	0.369	0.394
29	0.337	0.381	0.462	0.489	58	0.259	0.297	0.367	0.392
30	0.332	0.376	0.456	0.484	59	0.258	0.295	0.366	0.391
31	0.327	0.371	0.450	0.478	60	0.256	0.294	0.363	0.388

续表

n	0.9	0.95	0.99	0.995	n	0.9	0.95	0.99	0.995
61	0.255	0.292	0.362	0.387	81	0.233	0.269	0.335	0.359
62	0.253	0.291	0.361	0.385	82	0.232	0.268	0.334	0.358
63	0.252	0.289	0.359	0.383	83	0.232	0.267	0.333	0.356
64	0.251	0.288	0.357	0.382	84	0.231	0.266	0.332	0.356
65	0.250	0.287	0.355	0.380	85	0.230	0.265	0.331	0.355
66	0.249	0.285	0.354	0.379	86	0.229	0.264	0.330	0.353
67	0.247	0.284	0.353	0.377	87	0.228	0.263	0.329	0.352
68	0.246	0.283	0.351	0.376	88	0.228	0.262	0.328	0.352
69	0.245	0.282	0.350	0.374	89	0.227	0.262	0.327	0.351
70	0.244	0.280	0.348	0.372	90	0.226	0.261	0.326	0.350
71	0.243	0.279	0.347	0.371	91	0.225	0.260	0.325	0.349
72	0.242	0.278	0.346	0.370	92	0.225	0.259	0.324	0.348
73	0.241	0.277	0.344	0.368	93	0.224	0.259	0.323	0.347
74	0.240	0.276	0.343	0.368	94	0.223	0.258	0.323	0.346
75	0.239	0.275	0.342	0.366	95	0.223	0.257	0.322	0.345
76	0.238	0.274	0.341	0.365	96	0.222	0.256	0.321	0.344
77	0.237	0.273	0.340	0.364	97	0.221	0.255	0.320	0.344
78	0.236	0.272	0.338	0.363	98	0.221	0.255	0.320	0.343
79	0.235	0.271	0.337	0.361	99	0.220	0.254	0.319	0.341
80	0.234	0.270	0.336	0.360	100	0.219	0.254	0.318	0.341

表 B.4 双侧狄克逊 (Dixon) 检验的临界值表

n	0.95	0.99	n	0.95	0.99	n	0.95	0.99
3	0.970	0.994	21	0.479	0.560	39	0.375	0.442
4	0.829	0.926	22	0.469	0.548	40	0.370	0.438
5	0.710	0.821	23	0.460	0.537	41	0.367	0.433
6	0.628	0.740	24	0.449	0.522	42	0.364	0.432
7	0.569	0.680	25	0.441	0.518	43	0.362	0.428
8	0.608	0.717	26	0.436	0.509	44	0.359	0.425
9	0.564	0.672	27	0.427	0.504	45	0.357	0.422
10	0.530	0.635	28	0.420	0.497	46	0.353	0.419
11	0.619	0.709	29	0.415	0.489	47	0.352	0.416
12	0.583	0.660	30	0.409	0.480	48	0.350	0.413
13	0.557	0.638	31	0.403	0.473	49	0.346	0.412
14	0.587	0.669	32	0.399	0.468	50	0.343	0.409
15	0.565	0.646	33	0.395	0.463	51	0.342	0.407
16	0.547	0.629	34	0.390	0.460	52	0.340	0.405
17	0.527	0.614	35	0.388	0.458	53	0.338	0.402
18	0.513	0.602	36	0.438	0.442	54	0.337	0.400
19	0.500	0.582	37	0.380	0.450	55	0.335	0.399
20	0.488	0.570	38	0.377	0.447	56	0.334	0.399

续表

n	0.95	0.99	n	0.95	0.99	n	0.95	0.99
57	0.330	0.396	72	0.309	0.373	87	0.296	0.349
58	0.329	0.393	73	0.308	0.371	88	0.295	0.349
59	0.327	0.390	74	0.306	0.370	89	0.294	0.347
60	0.325	0.389	75	0.305	0.368	90	0.293	0.347
61	0.323	0.387	76	0.304	0.363	91	0.291	0.344
62	0.321	0.385	77	0.304	0.363	92	0.290	0.344
63	0.320	0.383	78	0.303	0.362	93	0.289	0.343
64	0.319	0.382	79	0.303	0.361	94	0.289	0.343
65	0.318	0.379	80	0.302	0.358	95	0.288	0.343
66	0.316	0.377	81	0.301	0.358	96	0.288	0.342
67	0.315	0.375	82	0.301	0.355	97	0.286	0.340
68	0.313	0.376	83	0.301	0.355	98	0.285	0.340
69	0.313	0.375	84	0.298	0.353	99	0.285	0.339
70	0.312	0.375	85	0.297	0.351	100	0.284	0.339
71	0.310	0.373	86	0.297	0.351			

B.3 T（Thompson）检验法

根据下式计算：

$$T = \frac{|x_n - \bar{x}|}{s \sqrt{\frac{n}{n-1}}} \quad (13)$$

式中： x_n ——被怀疑的异常值；

$\bar{x}$ 、 S ——不包括被怀疑的异常值 x_n 的样本平均值和标准差， n 为样本量（包括 x_n ）。

根据自由度 $df=n-2$ 查附录 A 的 t 分布表，得临界值 t_α ，检验结论如下：

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

B.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。

箱线图的绘制步骤：

- 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 $Q_1 - 1.5IQR$ 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； $Q_3 + 3IQR$ 和 $Q_1 - 3IQR$ 处为外限。处于内限以外位置的数据都是异常值，其中在内限与外限之间的异常值为温和异常值（mild outliers），在外限以外的为极端异常值（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“●”标出温和异常值，用“*”标出极端异常值（可以采用不同符号标识）。

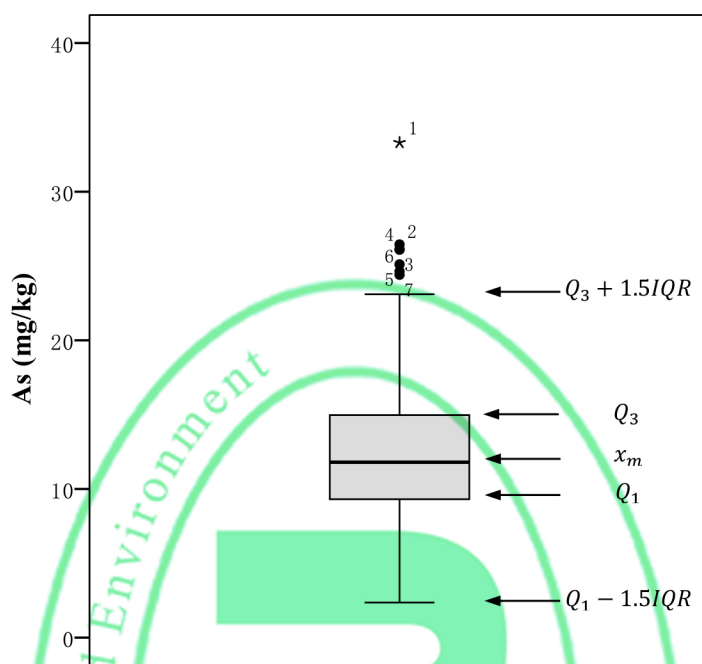


图 B.1 箱线图 中的砷浓度

B.5 富集系数法

由于 TiO_2 或 ZrO_2 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_2 或 ZrO_2 含量判断土壤是否被其他元素污染。方法是以 TiO_2 或 ZrO_2 含量作参比，计算其他元素的富集系数，公式为：

$$\text{富集系数} = \frac{\text{土壤中元素含量} / \text{土壤中} \text{TiO}_2 \text{或} \text{ZrO}_2 \text{含量}}{\text{成土母质(岩)中元素含量} / \text{成土母质(岩)中} \text{TiO}_2 \text{或} \text{ZrO}_2 \text{含量}} \quad (14)$$

若富集系数大于 1.2，一般判断该元素有外来污染。

富集系数法要求每一个土壤剖面的土层与下层成土母质（岩）同源，只适用于无机元素。