



黄亚捷, 贾凤超, 崔艳智, 周岩, 李君超, 刘海东. 偏态分布下已知变异单元内土壤重金属采样数目确定方法研究[J]. 中国农业大学学报, 2024, 29(12): 279-289.
HUANG Yajie, JIA Fenchao, CUI Yanzhi, ZHOU Yan, LI Junchao, LIU Haidong. Study on the method of determining the sampling number of soil heavy metals in known variation units under skew distribution[J]. *Journal of China Agricultural University*, 2024, 29(12): 279-289.
DOI: 10.11841/j.issn.1007-4333.2024.12.24

偏态分布下已知变异单元内土壤重金属采样数目确定方法研究

黄亚捷 贾凤超 崔艳智 周岩 李君超 刘海东*

(生态环境部 环境发展中心, 北京 100029)

摘要 土壤重金属合理采样数目的确定是进行土壤重金属相关研究工作的前提和基础,直接关系到土壤重金属调查、土地可持续利用及修复等工作。以农田土壤镉为研究对象,通过田间采样收集不同尺度下高低变异的17个决策单元内土壤采样数据,基于统计学及经验模型分别建立几何标准差与变异系数的关系,进一步推导偏态分布下土壤重金属合理采样数目确定方法。结果表明:相比于其他方法,基于统计学理论推导出的公式计算的几何标准差更为精确、计算的采样数目与精确采样数目更为接近;基于统计学理论推导出的公式产生的平均误差及相对偏差更小,对高、低变异单元估算的采样数目更为精确,且计算过程更为简单快捷。研究结果为已知变异系数的决策单元提供了一种有效的土壤重金属的采样数目确定方法。

关键词 土壤重金属; 偏态分布; 变异系数; 采样数目; 模型

中图分类号 X53

文章编号 1007-4333(2024)12-0279-11

文献标志码 A

Study on the method of determining the sampling number of soil heavy metals in known variation units under skew distribution

HUANG Yajie, JIA Fenchao, CUI Yanzhi, ZHOU Yan, LI Junchao, LIU Haidong*

(Environmental Development Center, Ministry of Ecology and Environment, Beijing 100029, China)

Abstract Determining the optimum sampling numbers (OSN) of heavy metals in soil is the premise and foundation for research of heavy metals in soil. It is one critical issue to achieve credible results when surveying heavy metals in soil and sustainable land use or remediation decisions. In this study, cadmium in farmland soil was taken as the research object. The data of Cd from 17 high-low variant units at different scales were collected by field sampling. The relationships between geometric standard deviation (GSD) and coefficient of variation (CV) were established by statistical and empirical theory. Based on the relationships, the method of determining the OSN of heavy metals in soil with skewed distribution was further derived. The results showed that: Compared with other methods, geometric standard deviation calculated by the method derived from statistical model was more accurate, and the OSN calculated by the proposed statistical model was closer to the exact sampling number; The mean error and relative deviation generated by the proposed statistical model were smaller, and the OSN calculated by the proposed statistical model in the high or low variation units was more accurate. The total calculation process of the proposed statistical

收稿日期: 2024-04-12

基金项目: 黄河流域生态保护和高质量发展联合研究项目(2022-YRUC-01-050204)

第一作者: 黄亚捷(ORCID:0000-0002-8313-7066), 博士, 高级工程师, 主要从事环境污染监测与治理研究, E-mail: 1060410517@qq.com

通讯作者: 刘海东(ORCID:0009-0006-0951-0041), 正高级工程师, 主要从事环境管理与规划、农业农村环境政策、环境影响评价等研究, E-mail: liuhaidong@edcmep.org.cn

model was simpler and faster. This study provides an effective method to determine the OSN of soil heavy metals in the decision unit with known CV.

Keywords soil heavy metals; skewed distribution; coefficient of variation; sampling number; model

土壤采样是进行土壤相关研究工作的前提和基础。为使有限的样本值能够尽可能精确地代表总体样本的期望值,土壤取样数应当达到一定的数量,即为合理采样数目^[1]。土壤重金属污染对人类健康存在很大的威胁,并在世界范围内得到了广泛的关注^[2-6]。田间土壤重金属的风险评估和污染防治离不开合理的采样数目^[7-8],而受制于采样强度、分析成本和研究结果精度等因素限制,进行大规模的土壤重金属密集采样需要考虑现实性和必要性。因此,如何确定土壤重金属合理采样数目始终是研究者和环境管理者关注的一个重点和难点问题。

以往研究中对土壤重金属采样设计的关注较少。Zhang等^[9]查询在两本重要的环境学英文期刊 *Environmental Science and Technology* 和 *Analytical Chemistry* 上已发表的论文,结果发现关于采样的

论文数目占全部发表论文的数目不足1%。同时,已有研究对土壤重金属采样数目的确定研究多集中于正态分布下的统计学方法^[5-6],在满足对数正态分布下土壤采样数目确定方法的研究更少^[10]。尽管在已知变异的决策单元内提出了对数正态分布下采样数目确定方法,如包括偏态分布下的经典对数正态公式^[11]、Hale’s方法^[12]、Armstrong’s方法^[13]、二次项近似公式^[14]、线性回归估计^[11]、切比雪夫不等式^[15]等,具体公式和方法详见表1。但是,上述修正方法的应用范围有所限制,不足以涵盖土壤重金属在不同变异系数下的采样布点,存在对合理采样数目高估及低估的现象。Huang等^[10]通过对多个上述方法的对比研究,提出了更为精确的已知变异的偏态分布下土壤重金属调查采样数目确定方法。

表1 对数正态分布下合理采样数目确定方法(95%的置信区间)

Table 1 Determining methods of optimum sampling numbers at the 95% confidence level under the log-normal distribution

方法 Method	公式 Equation	参考文献 Reference
经典对数正态公式 Classical lognormal equaion	$N = \frac{Z_{(\alpha,f)}^2 \times (\text{GSD}^{\ln(\text{GSD})} - 1)}{p^2}$	(1) [11]
Hale’ 公式 Hale’s equation	$N = \frac{Z_{(\alpha,f)}^2 S_y^2}{\text{Ln}^2(p + 1)}$	(2) [12]
Armstrong’ 公式 Armstrong’s equation	$p = \exp\left(Z_{(\alpha,f)} \sqrt{\frac{S_y^2}{N_4} + \frac{S_y^4}{2(N_4 - 1)}}\right) - 1$	(3) [13]
二次项近似公式 Quadratic term approximate equation	$N = \frac{\exp(-0.173) \times \text{GSD}^{4.3} \times \text{GSD}^{-0.99\ln(\text{GSD})}}{p^{1.2}}$	(4) [14]
线性回归估计 Linear regression estimation	$N = \beta_0 + \beta_1 \times N_{\text{classic}}$	(5) [11]
切比雪夫不等式 Chebyshev’s inequality	$N = \frac{\left(\frac{1}{\gamma} - 1\right) S_y^2}{p^2}$	(6) [15]

上述方法均为基于几何标准差(Geometric Standard Deviation,GSD)来确定土壤重金属采样数目。然而,土壤重金属变异性并不是通过GSD表达,更多的是通过变异系数(Coefficient of

Variation,CV)表达^[16-17],如Huang等^[10]在土壤重金属采样数目确定方法研究中,通过检索自CNKI、维普及ISI Web of Information citation index等数据库共收集到457个案例,这些案例中土壤重金属变异

情况均以 CV 表达,已提出的方法通过 GSD 计算土壤重金属采样数目的研究并不适用大部分已知 CV 的土壤重金属采样设计。

为进一步探索偏态分布下已知变异系数的土壤重金属采样数目确定方法,并综合考虑我国土壤污染普查及以往的研究调查中主要污染元素为镉,并得到广泛关注^[18]。因此,本研究拟以农田土壤镉为例,通过建立 GSD 与 CV 的关系,对前期工作中建立的偏态分布下土壤重金属采样数目确定公式进一步的推导,以期获得在已知变异系数的决策单元内土壤重金属的采样数目确定公式。

1 数据来源与处理

在确定土壤重金属采样数目时,决策单元的定义至关重要。不同的土壤决策单元定义直接影响土壤采样区域面积及采样数目的确定^[19-20]。联合国粮农组织^[21]、美国洲际技术和法规委员会^[22]、欧洲的一些国家^[23]对土壤污染物的决策单元定义并不相同,基于不同决策单元定义,进而采用不同的方法确定了不同的采样数目。在本研究中,具

有相同的主要污染元素,相似的土壤污染程度及变异程度($CV\leq 100\%$)的农业土壤污染单元定义为本研究的决策单元。在本研究中,通过网格等间距划分共采集了我国 17 个相互独立的土壤污染农田。采集的土壤样品风干后过尼龙筛,土壤粒径为 1 mm,土壤中镉浓度依照 HJ/T 166—2004《中国土壤环境监测技术规范》^[19]的要求测定,同时使用国家标准物质进行质量监控,以确保监测的精确度。

17 个土壤污染单元的采样信息见表 2。可见这 17 个污染单元呈现出不同的空间变异程度及空间尺度。为了获得更多的决策单元,本研究基于田间采样的 17 个污染单元,在不改变原来的采样间距的情况下设置不同的采样数目,进而获取了在 $\leq 1\text{ km}^2$ 及 $1.0\sim 10.0\text{ km}^2$ 及 $10.0\sim 81.5\text{ km}^2$ 尺度下的 855、995 及 877 个决策单元。以决策单元 2 为例,它的原始采样间距及采样数目分别为 200 m 及 138 个。基于决策单元的定义,在采样单元 2 内等间距划分为 8 个不同采样间距的单元,即 300、400、600、800、1 000 及 1 200 m,那么就形成了 8 个不同变异系

表 2 不同污染单元的采样信息
Table 2 Sampling information in different contaminated units

单元 Unit	主要污染物 Main contamination	采样数目 Sampling number	采样面积/ m^2 Sampling area	采样间距/m Sampling distance
1	Cd	121	4.84×10^6	200
2	Cd	138	5.52×10^6	200
3	Cd	142	3.55×10^5	50
4	Cd	143	5.72×10^6	200
5	Cd	65	1.04×10^5	40
6	Cd	17	51	3
7	Cd	29	87	3
8	Cd	15	45	3
9	Cd	18	54	3
10	Cd	15	135	3
11	Cd	20	1.33×10^4	26
12	Cd	104	312	3
13	Cd	500	4.99×10^7	316
14	Cd	457	3.07×10^6	82
15	Cd	816	8.15×10^7	316
16	Cd	211	1.42×10^6	82
17	Cd	173	1.72×10^7	316

数及采样数目的决策单元。以单元15为例,基于决策单元的定义,将单元15(面积为 $8.15 \times 10^7 \text{ m}^2$)划分为5个小面积相互独立的决策单元,那么就形成了5个具有不同变异系数及采样数目的决策单元。

2 研究方法

依据Huang等^[10]研究,提出土壤重金属合理采样数目确定公式,基于GSD确定采样数目,具体计算公式如下:

$$N = 1.20 \times \left[\frac{Z_{(\alpha,f)}^2 \times (\text{GSD}^{\ln(\text{GSD})} - 1)}{p^2} \right] + 8.69 \quad (7)$$

式中: N 为采样数目; $Z_{(\alpha,f)}$ 为某一置信水平下相对应的标准正态偏差, α 为显著水平; f 为自由度; p 为误差,通常指相对误差,本研究取 p 为15%。

已有研究发现土壤重金属变异性并不是通过GSD表达,更多的是通过CV表达^[16-17]。因此,本研究在已知CV的决策单元内建立偏态分布下的土壤重金属采样数目确定公式。

首先,基于统计学模型及经验模型建立GSD与CV的非线性函数。

2.1 基于统计学模型建立GSD与CV的关系

假设 $x_1, x_2, \dots, x_n$ 是独立的随机变量,每个变量都符合未知均值 θ 及方差 η^2 的对数正态分布。为了估算均值 θ 及方差 η^2 的值,通常是假设 $y_i = \ln(x_i)$, $i=1, 2, \dots, n$,将原来的问题简化为正态分布参数的估计问题。假设 $y_i \sim N(\mu, \sigma^2)$ 是符合均值 μ 及方差 σ^2 的正态分布,那么期望值及方差可表达为:

$$E(x) = \theta = e^{\mu + \frac{1}{2}\sigma^2} \quad (8)$$

$$D(x) = \eta^2 = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} = (e^{\sigma^2} - 1)(E(x))^2 \quad (9)$$

式中:对于样品 x_i , $x_a = \frac{\sum x_i}{n}$ 及 $\text{SD}_x^2 = \frac{\sum (x_i - x_a)^2}{n-1}$

分别为变量 x_i 的算术平均值及方差;对于样品 y_i , $y_a = \frac{\sum y_i}{n}$ 及 $\text{SD}_y^2 = \frac{\sum (y_i - y_a)^2}{n-1}$ 分别为变量 y_i 的算术平均值及方差。均值 θ 及方差 η^2 的计算是基于总体样本值,而在实际土壤样本计算时,由于经费的限制,并不能精确的采集到总体的样本。通常通过几个样本的算术平均值 x_a 估算总体样本的均值 θ ,以期接近潜在的真实“总体”均值 θ 。同样地,基于几个样本方差 $D(x)$ 估算总体的方差 η^2 。因此,根据以

往研究^[12],对于满足对数正态分布下的样本 x_i 的算术平均值 x_a 及算术标准差 SD_x 可分别估算为:

$$\hat{x}_a = e^{\mu + \frac{1}{2}\sigma^2} \quad (10)$$

$$\hat{D}_a = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} = (e^{\sigma^2} - 1)(\hat{x}_a)^2 \quad (11)$$

$$\widehat{\text{SD}}_x = e^{\mu + \frac{1}{2}\sigma^2} \sqrt{e^{\sigma^2} - 1} = (\hat{x}_a) \sqrt{e^{\sigma^2} - 1} \quad (12)$$

式中:对数正态分布下的样本 x_i 的几何平均值(x_g)可以估算为 $\hat{x}_g = e^\mu$,几何标准差可以估算为 $\text{GSD} = e^{\text{SD}_y}$,因此, $\hat{\sigma} = e^{\text{SD}_y} = \text{GSD}$ 。基于CV的定义,样本 x_i 的CV可计算为:

$$\text{CV} = \frac{\text{SD}_x}{x_a} \quad (13)$$

将式(10)的 x_a 的估算值($\hat{x}_a$)及式(12)的 SD_x 的估算值($\widehat{\text{SD}}_x$)分别代入式(13),那么CV可表达为:

$$\text{CV} = \sqrt{e^{\sigma^2} - 1} \quad (14)$$

因此,由于 $\text{GSD} = \hat{\sigma}$,那么对数分布下的样本 x_i 的GSD可进一步的估算为:

$$\text{GSD} = e^{\sqrt{\ln(1 + \text{CV}^2)}} \quad (15)$$

由式(15)可知, CV与GSD存在非线性关系,且呈现明显的正相关性,随着CV的增加, GSD也相应的增加。

2.2 基于经验性模型建立GSD与CV的关系

由式(7)可知,在确定可接受误差后,土壤重金属采样数目直接由GSD决定。为更便捷地推导式(7),本研究直接建立 $\text{GSD}^{\ln(\text{GSD})} - 1$ 与CV的关系。由图1可知,在采样面积 $\leq 1 \text{ km}^2$ 、 $1 \sim 10 \text{ km}^2$ 及 $10 \sim 81.5 \text{ km}^2$ 时, GSD与CV关系对应的 R^2 分别为0.98、0.98及0.99,非常接近于1,这说明不同公式呈现出好的拟合效果。因此,基于多个尺度下的不同案例拟合GSD与CV的关系(图1),当决策单元内的采样面积 $\leq 1 \text{ km}^2$ 、 $1 \sim 10 \text{ km}^2$ 及 $10 \sim 81.5 \text{ km}^2$ 时, $\text{GSD}^{\ln(\text{GSD})} - 1$ 与CV的关系可分别表达为式(16)、(17)及(18)。

$$\text{GSD}^{\ln(\text{GSD})} - 1 = 1.30 \times \text{CV}^{2.12} \quad (16)$$

$$\text{GSD}^{\ln(\text{GSD})} - 1 = 1.13 \times \text{CV}^{2.09} \quad (17)$$

$$\text{GSD}^{\ln(\text{GSD})} - 1 = 1.04 \times \text{CV}^{2.03} \quad (18)$$

其次,基于土壤重金属合理采样数目确定式(7),并结合基于统计学模型获取的GSD与CV的关系(即式(15)),式(7)可进一步的表达为:

$$N = 1.20 \times \left[\frac{Z_{(\alpha,f)}^2 \times \text{CV}^2}{p^2} \right] + 8.69 \quad (19)$$

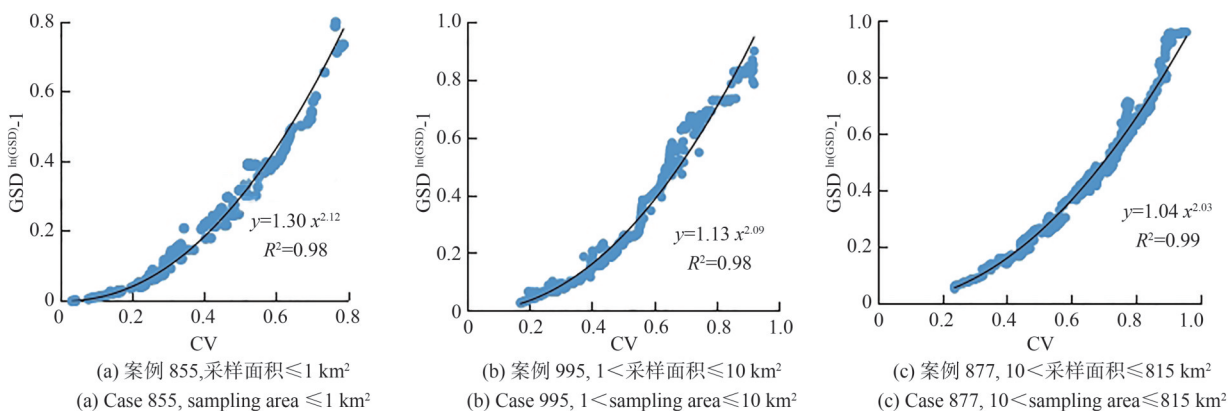


图 1 不同采样面积下 GSD 与 CV 的关系

Fig.1 Relationship between geometric standard deviation (GSD) and coefficient of variation (CV) under different sampling areas

当可接受误差(p)确定时,采样数目与 CV 直接相关。因此,式(19)为通过统计学理论在已知 CV 的决策单元内土壤镉的采样数目确定方法。土壤重金属的变异系数可通过土壤普查、历史区域文献查找或专家经验判断等方式获得。

基于不同的决策单元面积(A , 单位 km^2)下,将式(16)、(17)及(18)分别代入到式(7)中,基于经验模型获得的土壤重金属采样数目确定公式可进一步的推导为:

$$N = \begin{cases} 1.56 \times \left[\frac{Z_{(\alpha, f)}^2 \times CV^{2.12}}{p^2} \right] + 8.69 & A \leq 1 \text{ km}^2 \\ 1.36 \times \left[\frac{Z_{(\alpha, f)}^2 \times CV^{2.09}}{p^2} \right] + 8.69 & 1 \text{ km}^2 < A \leq 10 \text{ km}^2 \\ 1.20 \times \left[\frac{Z_{(\alpha, f)}^2 \times CV^{2.03}}{p^2} \right] + 8.6 & 10 \text{ km}^2 < A \leq 81.5 \text{ km}^2 \end{cases} \quad (20)$$

最后,以不同尺度下高低变异的决策单位为例,分别比较采用不同模型计算的 GSD 以及采用不同方法计算的土壤重金属采样数目,明确在已知 CV 的决策单元内土壤镉的采样数目确定方法。对于不同方法的优劣性主要是通过比较预测值和实际观测值进行评价。通常以平均误差(Mean Error, ME)和相对偏差(Relative Deviation, RD)来评价不同方法的预测精度和准确性。计算公式如下:

$$ME = \frac{1}{n} \sum_{i=1}^n [\hat{z}(x_i) - z(x_i)] \quad (21)$$

$$RD = \text{Abs}[\hat{z}(x_i) - z(x_i)] / z(x_i) \quad (22)$$

式中: n 为验证样点数目; $\hat{z}(x_i)$ 为不同方法的评估值; $z(x_i)$ 为基于采样确定的实测值。

3 结果与分析

3.1 不同方法下 GSD 与 CV 的关系比较

为了更好地验证统计学模型及经验模型计算 GSD 的优劣性,本研究选择了 17 个不同尺度下的

决策单元作为验证数据集,结合式(16)~(18)的特点,采用不同模型计算 $GSD^{\ln(GSD)}$ 结果见表 3。可见:GSD 与 CV 具有明显的正相关性,基于统计学模型推导出的式(15)及经验模型推导出的式(16)~(18)计算的 GSD 与实测值均很接近。

基于不同方法计算的几何标准差的误差比较结果见表 4。可见与经验模型相比,基于统计学模型计算的 GSD 更接近于实测值。其中,基于统计学模型计算的 17 个单元的几何标准差的 ME 为 0.005,低于经验模型计算几何标准差的 ME(0.038);基于统计学模型计算的几何标准差的 RD 为 2.35,低于基于经验模型计算的几何标准差 RD(3.5)。

在小尺度单元内,基于统计学模型计算的 GSD 比经验模型更为精确,如在采样面积为 312 m^2 的决策单元内,统计学模型计算的 RD 为 2.28,经验模型计算的 RD 为 13.311。

3.2 不同模型确定采样数目比较

为了更好地验证基于统计学理论推导出的已知

表3 不同决策单元的土壤几何标准差比较

Table 3 Comparison of geometric standard deviation in different decision-making units

单元 Unit	变异系数实测值/% Measured value of CV	几何标准差实测值/% Measured value of geometric standard deviation	GSD ^{ln(GSD)} Estimated value of geometric standard deviation		
			Estimated value of geometric standard deviation		
			实际值 Measured value	基于统计学模型计算 Calculation based on statistical model	基于经验模型计算 Calculation based on empirical model
1	60.6	1.719	1.354	1.367	1.397
2	47.6	1.634	1.248	1.226	1.239
3	28.9	1.314	1.080	1.083	1.094
4	26.9	1.293	1.070	1.073	1.073
5	32.0	1.328	1.093	1.102	1.116
6	27.2	1.297	1.072	1.074	1.082
7	42.9	1.425	1.157	1.184	1.216
8	66.3	1.739	1.396	1.440	1.544
9	21.3	1.219	1.043	1.045	1.049
10	22.9	1.272	1.056	1.053	1.057
11	11.0	1.123	1.013	1.012	1.012
12	83.1	2.001	1.653	1.691	1.878
13	76.9	1.649	1.406	1.591	1.610
14	84.0	2.362	1.874	1.706	1.785
15	99.1	2.377	2.046	1.982	2.021
16	32.1	1.333	1.094	1.103	1.105
17	43.6	1.471	1.175	1.190	1.193

表4 不同几何标准差计算方法的误差比较

Table 4 Error comparison of calculation methods for different geometric standard deviation

单元 Unit	采样面积/ m ² Sampling area	平均误差 ME		相对偏差 RD	
		基于统计学模型计算 Calculation based on statistical model	基于经验模型计算 Calculation based on empirical model	基于统计学模型计算 Calculation based on statistical model	基于经验模型计算 Calculation based on empirical model
		基于统计学模型计算 Calculation based on statistical model	基于经验模型计算 Calculation based on empirical model	基于统计学模型计算 Calculation based on statistical model	基于经验模型计算 Calculation based on empirical model
1	4.84×10 ⁶	0.013	0.043	0.962	3.134
2	5.52×10 ⁶	−0.022	−0.009	1.768	0.729
3	3.55×10 ⁵	0.003	0.013	0.271	1.214
4	5.72×10 ⁶	0.002	0.002	0.214	0.222
5	1.04×10 ⁵	0.010	0.024	0.875	2.133
6	51	0.002	0.010	0.180	0.962
7	87	0.027	0.059	2.346	5.023
8	45	0.043	0.147	3.106	10.242
9	54	0.003	0.006	0.250	0.612
10	135	−0.003	0.001	0.315	0.117
11	1.33×10 ⁴	−0.001	−0.001	0.068	0.072
12	312	0.038	0.225	2.280	13.311
13	4.99×10 ⁷	0.185	0.204	13.164	12.820
14	3.07×10 ⁶	−0.168	−0.089	8.947	5.210
15	8.15×10 ⁷	−0.064	−0.025	3.126	1.278
16	1.42×10 ⁶	0.009	0.011	0.822	0.992
17	1.72×10 ⁷	0.016	0.018	1.332	1.526
合计 Total		0.005	0.038	2.35	3.51

变异系数下土壤重金属采样数目确定方法式(19)的优势,选取经典统计学方法(具体公式见《土壤环境监测技术规范》^[17])、土壤重金属合理采样数目确定式(7)、基于经验模型推导出的式(20),以及表 1 中的经典对数正态方法、Hale’s 方法、Armstrong’s 方法、二次项近似公式、切比雪夫不等式等方法,以不同尺度下的 17 个单元作为验证单元(表 5),结果发现,相比较于其他方法,基于统计学模型推导出的式(19)及基于经验模型推导出的式(20)计算的采样数目与合理采样数目更为接近,其他方法计算的采样数目一般均比合理的采样数目少。在高、低变异单元内(如决策单元 3、4、5、12),基于统计学模型推导出的式(19)比基于经验模型推导出的式(20)确定的采样数目与合理采样数目更为接近。

表 5 基于不同方法计算的土壤镉的采样数目比较
Table 5 Sampling numbers of soil cadmium by different equations in decision-making units

单元 Unit	变异系数实测值/% Measured value of CV	N	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8
1	60.6	79	63	60	58	66	69	71	84	90
2	47.6	65	39	42	47	53	49	44	55	58
3	28.9	25	14	14	15	13	27	16	26	28
4	26.9	23	12	12	13	29	25	14	24	24
5	32.0	26	17	16	16	16	30	20	30	32
6	27.2	23	13	12	13	14	25	14	24	26
7	42.9	36	31	27	25	23	42	36	46	53
8	66.3	82	75	68	60	69	80	85	99	120
9	21.3	17	8	7	8	8	20	9	18	19
10	22.9	21	9	10	11	11	21	10	19	20
11	11.0	11	2	2	3	3	13	2	11	11
12	83.1	135	118	111	95	114	114	134	150	189
13	76.9	67	101	69	49	119	101	115	130	129
14	84.0	233	120	149	145	156	116	137	153	170
15	99.1	237	168	179	147	188	152	190	210	210
16	32.1	26	18	16	16	31	30	20	30	30
17	43.6	42	32	30	29	43	43	37	48	44

注: N 为式(7)确定的土壤镉合理采样数目。 N_1 为经典统计学公式确定的土壤镉的采样数目,具体公式见《土壤环境监测技术规范》^[19]。
 N_2 、 N_3 、 N_4 、 N_5 、 N_6 、 N_7 、 N_8 分别为式(1)、(2)、(3)、(4)、(6)、(19)、(20)确定的土壤镉的采样数目。
Note: N represents the optimum sampling number of soil cadmium determined by formula (7). N_1 represent the sampling numbers of soil cadmium determined by the classical statistical formula, and the specific formula can be found in *Technical Specification for Soil Environmental Monitoring* ^[19]. N_2 , N_3 , N_4 , N_5 , N_6 , N_7 and N_8 represent the sampling numbers of soil cadmium determined by formulas (1), (2), (3), (4), (6), (19) and (20), respectively. .

基于不同方法计算采样数目的误差进行比较(表 6、表 7),结果发现,比较于不同方法计算的土壤镉采样数目,基于统计学模型推导出的式(19)确定采样数目与合理采样数目更为接近,17 个单元产生的 ME 合计为 9,RD 合计为 2.94,低于其他方法计算的土壤镉采样数目产生的误差。另外,比较于统计学模型推导出的式(19),其他方法,如经验型模型推导出的式(20)过于复杂,不利于快速计算土壤重金属的采样数目。因此。统计学模型推导出的式(19)是偏态分布下更为精确、简单的基于已知变异系数的土壤重金属采样数目确定方法。

表6 基于不同方法计算的采样数目平均误差比较

Table 6 Comparison of sampling mean errors based on different methods

单元 Unit	变异系数实测值/% Measured value of CV	ME ₁	ME ₂	ME ₃	ME ₄	ME ₅	ME ₆	ME ₇	ME ₈
1	60.6	-16	-19	-21	-13	-10	-8	5	11
2	47.6	-26	-23	-18	-12	-16	-21	-10	-7
3	28.9	-11	-11	-10	-12	2	-9	1	3
4	26.9	-11	-11	-10	6	2	-9	1	1
5	32.0	-9	-10	-10	-10	4	-6	4	6
6	27.2	-10	-11	-10	-9	2	-9	1	3
7	42.9	-5	-9	-11	-13	6	0	10	17
8	66.3	-7	-14	-22	-13	-2	3	17	38
9	21.3	-9	-10	-9	-9	3	-8	1	2
10	22.9	-12	-11	-10	-10	0	-11	-2	-1
11	11.0	-9	-9	-8	-8	2	-9	0	0
12	83.1	-17	-24	-40	-21	-21	-1	15	54
13	76.9	34	2	-18	52	34	48	63	62
14	84.0	-113	-84	-88	-77	-117	-96	-80	-63
15	99.1	-69	-58	-90	-49	-85	-47	-27	-27
16	32.1	-8	-10	-10	5	4	-6	4	4
17	43.6	-10	-12	-13	1	1	-5	6	2
合计		-308	-324	-398	-192	-191	-194	9	105

注:ME₁、ME₂、ME₃、ME₄、ME₅、ME₆、ME₇、ME₈分别为采用经典统计学公式、式(1)、(2)、(3)、(4)、(6)、(19)、(20)计算采样数目的平均误差。

Note: ME₁ represents the mean error in calculating the sampling numbers by the classical statistical formula. ME₂, ME₃, ME₄, ME₅, ME₆, ME₇ and ME₈ represent the mean errors in calculating the sampling numbers by formulas (1), (2), (3), (4), (6), (19) and (20), respectively.

4 讨 论

已有研究认为土壤重金属采样数据并不符合正态分布,而是属于偏态分布,且在土壤地球化学背景下多数为对数正态分布^[7-8]。在对数正态分布下,目前最为常用的是经典对数正态方法(表1),然而,这个公式对合理样本数的计算多存在低估的现象^[10]。不同学者提出了经典对数正态方法修正公式,包括表1中的Hale’s方法、Armstrong’s方法、二次项近似公式、线性回归估计等方法,但是不同的方法应用均有所限值,如二次项近似公式更适用于误差<25%,且GSD相对较小(如GSD=1.5)的条件,对于误差>25%且GSD较大时并不推荐^[11]。对于较大的误差>25%的计算,Perez等^[11]通过线性回归估计,给出了不同置信水平下对应不同GSD

(如GSD为1.1,1.5,2.0,2.5,3.0,3.5,4.0)的计算公式。然而,这种方法仍不能计算涵盖不同GSD下土壤重金属的合理采样数目^[10]。值得注意的是,上述不同方法计算采样数目,均是依赖于GSD反映土壤重金属变异情况。

基于此,本研究提出了CV的土壤重金属采样数目确定公式,它是土壤重金属合理采样数目确定公式以及GSD与CV的非线性函数相结合,提出的基于统计学模型推导出的偏态分布下采样数目确定方法。本研究将17个不同尺度下的决策单元作为验证单元,对比分析了经典统计学方法、土壤重金属合理采样数目确定公式(7)、基于经验模型推导出的式(20),以及表1中的经典对数正态方法、Hale’s方法、Armstrong’s方法、二次项近似公式、切比雪夫不等式等方法。结果也发现,相比较于上

表 7 基于不同方法计算的采样数目相对偏差比较

Table 7 Comparison of sampling relative deviation based on different methods

单元 Unit	变异系数实测值/% Measured value of CV	RD ₁	RD ₂	RD ₃	RD ₄	RD ₅	RD ₆	RD ₇	RD ₈
1	60.6	0.20	0.24	0.27	0.16	0.13	0.10	0.06	0.14
2	47.6	0.40	0.35	0.28	0.18	0.25	0.32	0.15	0.11
3	28.9	0.44	0.44	0.40	0.48	0.08	0.36	0.04	0.12
4	26.9	0.48	0.48	0.43	0.26	0.09	0.39	0.04	0.04
5	32.0	0.35	0.38	0.38	0.38	0.15	0.23	0.15	0.23
6	27.2	0.43	0.48	0.43	0.39	0.09	0.39	0.04	0.13
7	42.9	0.14	0.25	0.31	0.36	0.17	0.00	0.28	0.47
8	66.3	0.09	0.17	0.27	0.16	0.02	0.04	0.21	0.46
9	21.3	0.53	0.59	0.53	0.53	0.18	0.47	0.06	0.12
10	22.9	0.57	0.52	0.48	0.48	0.00	0.52	0.10	0.05
11	11.0	0.82	0.82	0.73	0.73	0.18	0.82	0.00	0.00
12	83.1	0.13	0.18	0.30	0.16	0.16	0.01	0.11	0.40
13	76.9	0.51	0.03	0.27	0.78	0.51	0.72	0.94	0.93
14	84.0	0.48	0.36	0.38	0.33	0.50	0.41	0.34	0.27
15	99.1	0.29	0.24	0.38	0.21	0.36	0.20	0.11	0.11
16	32.1	0.31	0.38	0.38	0.19	0.15	0.23	0.15	0.15
17	43.6	0.24	0.29	0.31	0.02	0.02	0.12	0.14	0.05
合计 Total		6.40	6.21	6.52	5.80	3.03	5.33	2.94	3.78

注：RD₁、RD₂、RD₃、RD₄、RD₅、RD₆、RD₇和 RD₈分别为采用经典统计学公式、式(1)、(2)、(3)、(4)、(6)、(19)和(20)计算采样数目的相对偏差。

Note: RD₁ represents the relative deviation in calculating the sampling numbers by the classical statistical formula. RD₂, RD₃, RD₄, RD₅, RD₆, RD₇ and RD₈ represent the relative deviations in calculating the sampling numbers by formulas (1), (2), (3), (4), (6), (19) and (20), respectively.

述方法,基于统计学模型推导出的式(19)计算的采样数目与合理采样数目更为接近,计算中的ME及RD更小;对于高、低变异的决策单元内,基于统计学模型推导出的式(19)可以有效地克服其他方法在确定采样数目时的高估性及低估性,计算土壤重金属的采样数目更加精确,且计算过程更为简单、快捷。因此,基于统计学模型推导出的偏态分布下采样数目确定方法是一种有效、简单的土壤重金属合理采样数目确定方法。

尽管本研究提出了偏态分布下已知变异系数的土壤重金属采样数目确定方法,但是该方法只考虑了在一定置信条件下的土壤重金属采样数目,并未考虑土壤重金属的空间变异性,也不能决定采样点的空间布局。因此,在未来研究中,会进一步考虑土壤重金属的空间结构性和随机性要素,不仅确

定土壤重金属的采样数目,而且也会明确土壤重金属采样点的空间布局。

5 结 论

本研究以不同尺度不同变异的 17 个决策单元内土壤重金属镉为研究对象,以 ME 及 RD 作为评价指标,分别采用统计学模型及经验模型建立 GSD 与 CV 的关系,并结合偏态分布下已知 GSD 的土壤重金属合理采样数目确定公式,探究在已知变异系数的决策单元内土壤重金属采样数目确定方法。主要结论如下:1)相比较于经验模型,基于统计学模型计算的 GSD 更接近于实测值,计算 GSD 过程中的 ME 及 RD 更小。2)相比较于经验模型,以及表 1 中的经典对数正态方法、Hale’s 方法、Armstrong’s 方法、二次项近似公式、切比雪夫不等

式等方法,基于统计学模型推导出的式(19)计算的采样数目与合理采样数目更为接近,计算中的ME及RD更小,对高、低变异单元估算的采样数目更为精确,且计算过程更为简单、快捷。因此,研究认为式(19)是偏态分布下更为精确、简单的基于已知变异系数的土壤重金属采样数目确定方法。

参考文献 References

- [1] 王子龙,陈伟杰,付强,姜秋香,常广义,胡石涛.土壤优化采样策略研究进展[J].水土保持通报,2017,37(5):205-212
Wang Z L, Chen W J, Fu Q, Jiang Q X, Chang G Y, Hu S T. Research progress of soil sampling strategy optimization[J]. *Bulletin of Soil and Water Conservation*, 2017, 37(5): 205-212 (in Chinese)
- [2] 黄亚捷,李菊梅,马义兵.土壤重金属调查采样数目的确定方法研究进展[J].农业工程学报,2019,35(24):235-245
Huang Y J, Li J M, Ma Y B. Research progress of methods for determining sampling numbers of soil heavy metals survey[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(24): 235-245 (in Chinese)
- [3] 黄亚捷,崔艳智,周岩,李君超,卞华锋,刘海东.土壤重金属混合样采样数目和空间变异耦合规律研究进展[J].中国农业大学学报,2024,29(2):47-57
Huang Y J, Cui Y Z, Zhou Y, Li J C, Bian H F, Liu H D. Research progress on coupling regularity between composite sampling numbers and spatial variations of soil heavy metals[J]. *Journal of China Agricultural University*, 2024, 29(2): 47-57 (in Chinese)
- [4] Ni R X, Ma Y B. Current inventory and changes of the input/output balance of trace elements in farmland across China[J]. *PLoS One*, 2018, 13(6): e0199460
- [5] 楚天舒,王柄雄,高思程,成天一,杨增玲.基于农田土壤重金属污染风险筛选值的畜禽粪污农田承载力估算[J].中国农业大学学报,2021,26(2):125-138
Chu T S, Wang B X, Gao S C, Cheng T Y, Yang Z L. Estimation on farmland carrying capacity for livestock and poultry manure based on risk screening values for farmland soil heavy metal contamination[J]. *Journal of China Agricultural University*, 2021, 26(2): 125-138 (in Chinese)
- [6] 高凤杰,王鑫,韩晶,刘媚媚,郭欣欣.东北黑土区小流域耕地土壤重金属污染特征及健康风险评价:以海沟河小流域为例[J].中国农业大学学报,2020,25(8):73-83
Gao F J, Wang X, Han J, Liu M M, Guo X X. Heavy metal pollution characteristics and its health risk assessment in a mollisol watershed of Northeast China: Taking Haigou watershed as study case[J]. *Journal of China Agricultural University*, 2020, 25(8): 73-83 (in Chinese)
- [7] Huang Y J, Li J M, Ma Y B, Li F B, Chen D L. A simple method to determine the sampling numbers in decision-making units with unknown variations of soil cadmium[J]. *Environmental Monitoring and Assessment*, 2021, 193(9): 552
- [8] Liao Y L, Li D Y, Zhang N X. Comparison of interpolation models for estimating heavy metals in soils under various spatial characteristics and sampling methods[J]. *Transactions in GIS*, 2018, 22(2): 409-434
- [9] Zhang J, Zhang C. Sampling and sampling strategies for environmental analysis[J]. *International Journal of Environmental Analytical Chemistry*, 2012, 92(4): 466-478
- [10] Huang Y J, Li J M, Ma Y B. Determining optimum sampling numbers for survey of soil heavy metals in decision-making units: Taking cadmium as an example[J]. *Environmental Science and Pollution Research International*, 2020, 27(19): 24466-24479
- [11] Perez A, Lefante J. Sample size determination and the effect of censoring when estimating the arithmetic mean of a lognormal distribution[J]. *Communications in Statistics: Theory and Methods*, 1997, 26(11): 2779-2801
- [12] Hale W E. Sample size determination for the log-normal distribution[J]. *Atmospheric Environment (1967)*, 1972, 6(6): 419-422
- [13] Armstrong B G. Confidence intervals for arithmetic means of lognormally distributed exposures[J]. *American Industrial Hygiene Association Journal*, 1992, 53(8): 481-485
- [14] Box G E P, Cox D R. An analysis of transformations[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1964, 26(2): 211-243
- [15] Interstate Technology Regulatory Council (ITRC). *Incremental Sampling Methodology*[M]. Washington, DC: Interstate Technology & Regulatory Council, Incremental Sampling Methodology Team, 2012.
- [16] Wang C, Yang Z F, Zhong C, Ji J F. Temporal-spatial variation and source apportionment of soil heavy metals in the representative river-alluviation depositional system[J]. *Environmental Pollution*, 2016, 216: 18-26
- [17] Liang J, Feng C T, Zeng G M, Gao X, Zhong M Z, Li X D, Li X, He X Y, Fang Y L. Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China[J]. *Environmental Pollution*, 2017, 225: 681-690
- [18] Wang L, Cui X F, Cheng H G, Chen F, Wang J T, Zhao X Y, Lin C Y, Pu X. A review of soil cadmium contamination in China including a health risk assessment[J]. *Environmental Science and Pollution Research International*, 2015, 22(21): 16441-16452
- [19] 国家环境保护总局.土壤环境监测技术规范:HJ/T 166—2004[S].北京:中华人民共和国环境保护行业标准,2004
State Environmental Protection Administration. Technical specification for soil environmental monitoring: HJ/T 166—200 [S]. Beijing: Environmental Protection Industry Standards of PRC (in Chinese), 2004
- [20] Theocharopoulos S P, Wagner G, Sprengart J, Mohr M E, Desaulles A, Muntau H, Christou M, Quevauviller P. European soil sampling guidelines for soil pollution studies[J]. *Science of the Total Environment*, 2001, 264(1-2): 51-62
- [21] FAO. A framework for land evaluation. 1976[DB/OL]. [2019-07-15]. <http://www.fao.org/3/x5310e/x5310e00.htm>
- [22] Interstate Technology and Regulatory Council. *Incremental Sampling Methodology*[M]. Washington, DC: Taylor & Francis, 2012
- [23] Wagner G, Desaulles A, Muntau H, Theocharopoulos S, Quevauviller P. Harmonisation and quality assurance in pre-analytical steps of soil contamination studies: Conclusions and recommendations of the CEEM Soil Project[J]. *The Science of the Total Environment*, 2001, 264(1-2): 103-117

责任编辑:杨爱东



第一作者简介：黄亚捷,博士,高级工程师。研究方向为环境污染监测与治理研究、农业农村环境政策等。发表学术论文 20 余篇。其中,以第一作者身份发表 SCI、EI 等国际顶级刊物发表论文 10 余篇。作为编委,编辑著作 3 本。多篇《环境发展专报》获得生态环境部部领导肯定批示。主持和作为技术骨干参与了国家自然科学基金、地方项目(课题)8 项。



通讯作者简介：刘海东,博士,教授级高级工程师,现任生态环境部环境发展中心环境管理与数据应用研究所副所长。多年来主要从事环境管理与规划、农业农村环境政策、环境影响评价等领域的研究工作,是农业部畜禽养殖废弃物资源化利用技术指导委员会委员、中国环境科学学会环境社会治理专委会常务委员。发表论文多篇,撰写多篇研究与政策信息专报,获得国办、生态环境部有关领导批示。