

## 面向土壤环境质量等级划分的统计推断与加密采样优化方法研究综述

高秉博, 郝朝展, 李发东, 胡茂桂, 李晓岚, 郜允兵, 潘瑜春

### 引用本文:

高秉博, 郝朝展, 李发东, 等. 面向土壤环境质量等级划分的统计推断与加密采样优化方法研究综述[J]. 农业环境科学学报, 2021, 40(4): 712-722.

在线阅读 View online: <https://doi.org/10.11654/jaes.2020-1202>

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于时空不确定性分析的北京市农田土壤重金属镉含量等级划分

李晓岚, 高秉博, 周艳兵, 潘瑜春, 郜允兵, 李斌, 胡茂桂

农业环境科学学报. 2019, 38(2): 307-316 <https://doi.org/10.11654/jaes.2018-0552>

### 一种农田土壤重金属影响评价的新方法:土壤和农产品综合质量指数法

王玉军, 刘存, 周东美, 陈怀满

农业环境科学学报. 2016, 35(7): 1225-1232 <https://doi.org/10.11654/jaes.2016.07.001>

### 土壤筛选值在镉污染稻米产地环境评价与分类适用性探讨

王祖光, 周其文, 赵玉杰, 刘潇威, 张铁亮, 王夏晖, 李志涛

农业环境科学学报. 2019, 38(10): 2328-2337 <https://doi.org/10.11654/jaes.2019-0167>

### 三峡库区库尾典型农用地土壤重金属污染特征及潜在风险

王金霞, 罗乐, 陈玉成, 何清明, 詹玲玲

农业环境科学学报. 2018, 37(12): 2711-2717 <https://doi.org/10.11654/jaes.2018-0844>

### 土壤重金属污染评价方法研究——以广州市为例

谢志宜, 张雅静, 陈丹青, 杨剑军, 梁耀杰

农业环境科学学报. 2016, 35(7): 1329-1337 <https://doi.org/10.11654/jaes.2016.07.015>



关注微信公众号, 获得更多资讯信息

高秉博, 郝朝展, 李发东, 等. 面向土壤环境质量等级划分的统计推断与加密采样优化方法研究综述[J]. 农业环境科学学报, 2021, 40(4): 712-722.

GAO Bing-bo, HAO Zhao-zhan, LI Fa-dong, et al. Statistical inference and the additional sampling optimization method for soil environmental quality grade classification: A review[J]. *Journal of Agro-Environment Science*, 2021, 40(4): 712-722.



开放科学 OSID

# 面向土壤环境质量等级划分的统计推断与加密采样优化方法研究综述

高秉博<sup>1,2</sup>, 郝朝展<sup>1</sup>, 李发东<sup>3</sup>, 胡茂桂<sup>3</sup>, 李晓岚<sup>4</sup>, 郜允兵<sup>4</sup>, 潘瑜春<sup>4</sup>

(1. 中国农业大学土地科学与技术学院, 北京 100083; 2. 农业农村部农业灾害遥感重点实验室, 北京 100083; 3. 中国科学院地理科学与资源研究所, 北京 100011; 4. 国家农业信息化工程技术研究中心, 北京 100097)

**摘要:** 环境质量等级划分是土壤环境调查和监测的一项重要目标。不同于总体估计和空间插值, 等级划分只需要准确估计污染物浓度与等级划分阈值之间的大小关系, 同时由于土壤环境质量等级具有空间连续性, 加密采样时应该将样点布设在等级过渡地带, 准确获取等级边界。但是初步调查的样点一般较为稀疏, 导致等级估计结果具有较大的不确定性, 难以反映真实的等级过渡情况, 而且具体应用对两类等级划分错误也有不同的控制要求, 构成了土壤环境质量等级划分及加密采样面临的特殊问题。本文分析梳理了国内外土壤环境质量等级划分统计推断与加密采样布局优化方法的研究进展, 归纳总结了以等级划分为目的的统计推断方法与加密采样布局优化方法体系, 并探讨了下一步研究的方向, 为我国目前开展的农用地土壤环境类别划分、工矿企业土壤污染修复区确定等重大需求提供方法选择依据。

**关键词:** 环境质量; 等级划分; 加密采样; 统计推断

中图分类号: X53 文献标志码: A 文章编号: 1672-2043(2021)04-0712-11 doi:10.11654/jaes.2020-1202

## Statistical inference and the additional sampling optimization method for soil environmental quality grade classification: A review

GAO Bing-bo<sup>1,2</sup>, HAO Zhao-zhan<sup>1</sup>, LI Fa-dong<sup>3</sup>, HU Mao-gui<sup>3</sup>, LI Xiao-lan<sup>4</sup>, GAO Yun-bing<sup>4</sup>, PAN Yu-chun<sup>4</sup>

(1. College of Land Science and Technology, China Agricultural University, Beijing 100083, China; 2. Key Laboratory of Agri-informatics, Ministry of Agriculture and Rural Affairs, Beijing 100083, China; 3. Institute of Geographic Sciences and Natural Resources, Chinese Academy of Sciences, Beijing 100011, China; 4. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China)

**Abstract:** Environmental quality grade classification is one crucial goal of soil surveying and monitoring. Unlike global mean estimation and spatial interpolation, environmental quality grade classification only focuses on estimating the relationship between pollutant concentration and grading threshold. Due to spatial autocorrelation, soil environmental quality grades are usually spatially continuous. Thus, to improve the classification precision of environmental quality grades, more sampling points should be laid in the transition area for environmental quality grades. However, because the primary survey sampling points are usually sparse and produce large errors and uncertainty in the estimated result, it is difficult to reflect the transition area of true grades. Additionally, applications have different requirements for the two types of errors in grade classification. The particularity of environmental quality grade classification needs special inference and sampling optimization design methods. This study analyzed the inference and additional sampling optimization design

收稿日期: 2020-10-19 录用日期: 2020-12-14

作者简介: 高秉博(1984—), 男, 宁夏海原人, 博士, 副教授, 主要从事空间统计分析等研究。E-mail: gaobingbo@cau.edu.cn

基金项目: 国家重点研发计划项目(2016YFD0800301); 国家自然科学基金项目(41601425)

Project supported: National Key R&D Program of China(2016YFD0800301); The National Natural Science Foundation of China(41601425)

method for environmental quality grade classification locally and globally. The method framework is also summarized, and future research is discussed. This work's outcome can provide a method selection basis for several agricultural soil environmental management applications, such as soil environmental quality grade classifications of agricultural land and definition of polluted areas for soil environmental remediation.

**Keywords:** environmental quality; grade classification; additional sampling; statistical inference

严格精准的环境管理与治理是实现生态文明与可持续发展战略的必由之路。为了加强生态环境保护,我国陆续发布了《大气污染防治行动计划》《土壤污染防治行动计划》和《水污染防治行动计划》,部署了环境管理治理任务与目标。将环境按照污染程度划分为不同等级以实施差别化管理与治理是我国环境管控的基本策略,同时不同污染等级的面积变化监测也是考核责任主体的主要手段。因此,环境质量等级划分是环境管理与治理的一项基本需求<sup>[1]</sup>。

在土壤环境质量调查和监测中,空间采样是目前最主要的手段。统计推断是采样的目的,决定采样点布局优化方法的选择。土壤环境质量等级划分是不同于均值估计、空间插值和热点探测的一种特殊的统计推断,其目的是准确估计未采样单元处污染物浓度与等级划分阈值之间的关系。由于空间相关性的存在,土壤环境质量等级在空间上的分布具有连续性,因此土壤环境质量等级划分的目的可以转化为确定不同环境质量等级间的边界。由于常规调查监测一般服务于多个目标且样点分布较为稀疏,为了提高土壤环境质量等级边界划分的准确性,通常需要在等级错误划分发生概率较大的地区进一步开展专项加密采样。但是由于其目标的特殊性,土壤环境质量等级划分需要具有针对性的统计推断与加密采样布局优化方法,以提高采样效率与等级划分精度<sup>[2]</sup>。

本文梳理了土壤环境质量等级划分统计推断与加密采样布局优化方法的研究进展,对比了理论方法的优势与不足,归纳了以等级划分为目的的统计推断方法与加密采样布局优化方法体系,并探讨了下一步研究的方向。

## 1 空间采样布局优化与统计推断之间的关系

采样,也称为抽样,通常指从研究对象全体中抽取一部分进行试验和观察,并获取试验数据的工作<sup>[3-4]</sup>。为了获得精度更高的土壤环境质量等级划分图(即获得研究区域中每个空间单元上污染物的环境质量等级),空间采样和统计推断仍然是目前最重要

的方法。统计推断是采样的目的,指基于样本数据对产生样本总体的某些性质以概率的形式进行推断,从而获得对总体的认知。空间采样是土壤环境调查和监测的常用手段<sup>[5-8]</sup>。De Gruijter等<sup>[9]</sup>认为,一个空间采样方案由样本布设和统计推断两部分组成,并将这些方法分为基于设计的方法和基于模型的方法两大类。基于设计的方法由概率采样和基于设计的统计推断方法组成,适用于获得总体参数的估计,如均值、标准差等全局变量;基于模型的方法由目的性采样和基于模型的统计推断组成,适合估计与具体位置相关的污染物浓度,如空间插值、热点探测和等级划分等<sup>[10-11]</sup>。而空间采样与统计推断三一准则(Trinity principle of spatial sampling and statistical inference)认为,调查精度由总体特征、样本布设方式、统计推断模型,以及三者之间的匹配关系所决定<sup>[12-13]</sup>,不同的组合方法有不同的效率。因此,在具体土壤环境等级划分中,需要根据具体的统计推断目标、样本布设方式和研究区域变量特征制定合适的采样设计方法。本文依据空间采样与统计推断三一准则分析归纳面向土壤环境质量等级划分的统计推断与加密采样布局优化方法。

## 2 土壤环境质量等级划分统计推断方法

### 2.1 方法体系

环境质量等级划分方法可以按照等级划分依据分为物理阈值法和概率阈值法两类,如表1所示。物理阈值方法首先对研究区域中每个空间单元的污染物浓度进行估计,然后通过对比估计值与物理阈值的大小确定待估计单元的环境质量等级。概率阈值方法则首先估计每个单元不超过物理阈值的概率,然后将所估计概率值与设定的概率阈值对比确定待估计单元的环境质量等级。物理阈值法和概率阈值法都需要对研究区域的每个空间单元进行估计。物理阈值法主要包括确定性方法、地统计方法、模糊数学方法和机器学习法。地统计方法按照估计的方式可以进一步分为物理值估计方法和不确定性估计方

法<sup>[14-15]</sup>。物理值估计方法主要目标是根据样点估计环境变量在待估计单元的取值及其估计误差,如普通克里金、回归克里金和协同克里金等方法<sup>[16-20]</sup>。关于土壤环境物理值估计方法更全面的综述,请参见Goovaerts<sup>[21]</sup>、Li等<sup>[17-18]</sup>和史文娇等<sup>[22-23]</sup>的文章。不确定性估计的目标主要是构建待估计单元处污染物浓度的概率分布,该类方法可进一步分为参数法和非参数法两类。参数法通过假设待估计单元处污染物浓度的概率分布模型,并使用样本计算概率分布的关键参数如期望方差等,构建待估计单元处随机变量的概率分布。最常用的方法为multi-Gaussian模型,它假设环境变量服从正态分布,并使用简单克里金的估计值和估计误差作为正态分布的期望和方差<sup>[21,24]</sup>。非参数方法并不预先假设待估计单元处环境变量分布的形式,只估计一些关键阈值所对应的累积概率,必要时使用内插和外插方法生成完整的累积概率分布函数<sup>[25]</sup>。模糊数学方法采用模糊隶属函数量化采样点与待估计单元之间环境相似程度,并据此进行土壤属性插值,是一种基于地理环境相似性的非监督学习方法。构建模糊隶属函数需要两种知识<sup>[26]</sup>:第一种类型称为I型知识,定义特定土壤属性的典型环境条件;第二种类型称为II型知识,定义了环境条件偏差带来的土壤属性变化,具体可参见Zhu等<sup>[26]</sup>、Yang等<sup>[27]</sup>、谢军等<sup>[28]</sup>的文章。机器学习法即统计模型的算法化,采用各种不同形式的基函数开展监督学习,并权衡训练样本拟合精度与模型泛化能力,如随机森林、提升树、人工神经网络模型等,通过学习训练建立相关辅助变量与环境变量之间的关系以估算污染物浓度的空

间分布<sup>[29-31]</sup>。

## 2.2 方法对比分析

环境质量等级划分方法比较分析如表2所示。总体而言,物理阈值法由于需要首先估计未采样单元处污染物浓度,而不是直接估计环境质量等级,因此会引入额外的步骤与误差;而概率阈值可以直接判定未采样单元处污染物浓度与等级划分阈值之间的关系。在物理阈值法中,确定性方法虽然易于理解、参数较少且鲁棒性较强,但是用于环境质量等级划分和加密采样空间布局优化时存在明显不足。由于在土壤环境中经常面临研究对象知识掌握不全面且数据不足的情况,难以建立准确的确定性模型进行估计,因此每个估计结果都存在误差。而确定性方法不能给出计算结果的不确定性程度,同时也不能依据区域空间变异进行计算参数的自适应调整,因此不能有效指导土壤环境管理<sup>[32]</sup>。地统计方法能够实现插值参数的自适应调整,给出插值结果及其不确定性,而且其中的协同克里金、回归克里金等方法还能利用辅助变量信息提高插值精度。但是其中的物理值估计方法在用于指导土壤环境质量等级划分时也存在明显不足:一方面,由于其估计值存在平滑效应,在处理较大或较小阈值的等级划分时存在较大误差,甚至会造成划分结果中等级的丢失;另一方面,其估计误差假设为正态分布,且假设误差只与样点位置有关,难以符合实际情况,不能准确反映估计结果与阈值关系的不确定性<sup>[33]</sup>。而其中的不确定性估计方法用于估计空间单元污染物浓度时,需要首先计算获得其概率分布,然后基于概率分布估计污染物浓度,如使用期望

表1 环境质量等级划分方法体系

Table 1 Environmental quality classification method system

等级划分依据 Classification basis	统计推断方法 Statistical inference method		基本原理 Principle
物理阈值	确定性方法		按照空间单元之间的距离远近关系计算出确定性的估计结果
	地统计方法	物理值估计方法	基于空间半变异函数进行无偏最优估计,在得到待估计单元处环境变量估计值的同时,还能得到估计结果的不确定性
		不确定性估计方法	参数估计
		非参数估计	不预先假设概率分布,只估计一些关键阈值所对应的累积概率,据此估计污染物浓度及其不确定性
	模糊数学方法		基于模糊隶属函数量化采样点与待估计单元之间的相似性,基于此估计污染物浓度
机器学习法		多元回归方法的算法化扩展,基于样本数据构建辅助变量与土壤属性之间的关系模型,计算待估计单元处污染物浓度	
概率阈值	地统计方法	不确定性估计方法	基于概率分布假设获得待估计单元的全概率分布,并获得等级划分阈值处的累积概率
		非参数估计	不假设概率分布,只估计等级划分阈值所对应的累积概率

表2 土壤环境质量等级划分方法比较分析

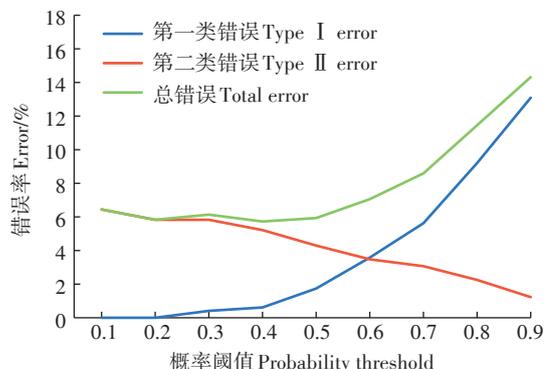
Table 2 Comparison of soil environmental quality classification methods

等级划分依据 Classification basis	统计推断方法 Statistical inference method	优点 Advantage	不足 Shortcoming
物理阈值	确定性方法	易于理解,需要参数较少,鲁棒性强	不能依据区域空间变异进行插值参数的自适应调整,不能获得估计结果的不确定性
	地统计方法	能够实现插值参数的自适应调整,给出插值结果及其不确定性;此外,协同克里金、回归克里金等方法还能利用辅助变量信息提高插值精度	物理值估计方法的平滑效应和估计误差的不准确性会引入误差;不确定性估计方法需要概率分布假设或概率内插和外插,会引入较大误差
	模糊数学方法	能够充分利用辅助数据和先验知识提高土壤属性插值精度,并给出插值结果的不确定程度	没有进行样本的监督训练,过多依赖先验知识,对于空间位置关系考虑不足
	机器学习法	能够有效处理多维多类型的辅助变量数据,充分挖掘利用辅助变量与环境变量之间的相关关系,部分方法能给出插值结果的不确定性	需要较大样本量进行模型训练,未能有效考虑空间相关性与异质性
概率阈值	参数估计	能够给出待估计单元处污染物浓度的全概率分布,一次性获得多个划分阈值对应的累积概率	具有较强的概率分布模型假设,难以符合实际土壤环境变量分布;需要确定概率阈值
	非参数估计	不需要进行概率分布假设,只需要估计等级划分阈值处的累积概率,并直接获得待估计属于环境某一环境质量等级的概率	需要确定概率阈值

估计(E-type estimate)、中位数估计(Median estimates)和百分位数估计(Quantile estimates)等损失函数(Loss function)对污染物浓度进行优化估计<sup>[21,34]</sup>。在优化估计过程中需要完整的累积概率分布模型,而参数法由于其较强的分布假设难以获得较好的结果,非参数法则会因需要概率内插和外插而引入较大的误差。模糊数学方法能够充分利用先验知识和土壤环境相关辅助数据提高土壤属性的插值精度,同时也能给出对插值结果的不确定性估计,但是该方法通常过分依赖先验知识,没有进行样本的监督训练,而且对空间位置关系的考虑不足。机器学习法的优势在于有效处理多维多类型辅助变量,能充分挖掘利用辅助变量与环境变量之间的相关关系提高插值精度,部分方法还能给出插值结果的不确定性,目前已经成为数字土壤制图的主流方法。不足之处在于目前的机器学习方法需要较大样本量进行模型训练,目前尚不能有效融合空间相关性与异质性解决稀疏样点条件下的插值精度问题<sup>[35-36]</sup>。

概率阈值法由于需要获得关于物理阈值的概率,只能与不确定性估计方法结合进行环境质量等级划分。其中非参数方法由于没有较强的分布假设,同时在等级划分时只需要阈值所对应的累积概率,因此更适合用于环境质量等级划分<sup>[32,37]</sup>。Antunes等<sup>[38]</sup>和Chica-Olmo等<sup>[39]</sup>探讨了基于概率阈值方法的空间等级划分,但是这些研究仅止步于将环境质量等级划分问题转化为基于概率阈值的等级划分问题,而对于如何确定概率阈值并没有给出有效的解决方案。高秉

博<sup>[40]</sup>提出了基于指示克里金估计概率及其不确定性的土壤环境等级错误划分方法,建议概率阈值以0.5为基础,依据实际需求进行调整,但未给出定量的确定方法。Gao等<sup>[41]</sup>在此基础上提出了基于交叉检验的自适应概率阈值确定方法,如图1所示,通过设定不同的概率阈值并进行交叉检验,获得研究区域环境质量等级划分的第一类错误、第二类错误和总错误随概率阈值变化的曲线,并依据曲线的特征点确定概率阈值,如总错误曲线最低点、第一类错误曲线与第二类错误曲线交叉点等。该方法能够适应不同区域数据分布的特点,提供了一种定量确定概率预知的方法。李晓岚等<sup>[42]</sup>将该方法拓展到时空维,基于时空克里金统计推断结果与概率阈值自适应确定方法,完成了北京市农田土壤重金属含量的等级划分。

图1 基于交叉检验的自适应概率阈值确定法<sup>[41]</sup>Figure 1 Adaptive probability threshold determination based on cross validation<sup>[41]</sup>

### 3 土壤环境质量等级划分加密采样布局优化方法

#### 3.1 方法体系

当多阶段采样或补充调查需要额外的加密样本时,采样布局优化尤其重要。为了提高土壤环境质量等级划分的精确性,应优化采样布局,使用更少的样点获得更高的等级划分精度。采样方法是否适合取决于采样的目标,当以土壤环境质量等级划分作为统计推断目标时,不需要强调每个空间单元的污染物浓度或全局平均浓度的估计精度,而需要精确估计未采样单元上污染物浓度与等级划分阈值之间的大小关系,以实现土壤环境质量等级的准确划分。由于土壤环境变量的空间自相关性,土壤环境质量等级在地理空间上具有连续性,因此加密采样时应该将更多样点布设在等级过渡地带<sup>[32]</sup>。按照加密采样布局优化目标设定的依据,可以将环境质量等级划分加密采样布局优化方法分为估计值准确度和等级划分精确度两类,具体如下表3所示。

依据估计值准确度加密采样布局优化方法包括随机采样和目的性采样。随机采样中分层随机抽样在实际中应用比较广泛,它基于先验知识、历史调查数据或辅助数据,首先将研究区划分为内部方差较小的若干个较为均质的子区域,然后为每个子区域分配样本量并分别布设样点。目的性采样方法主要通过设定优化目标函数并使用优化求解方法生成采样布局方案<sup>[43]</sup>。常见方法可以分为统计推断误差最小优化方法、地理空间分布优化方法、特征空间分布优化

方法、地理空间与特征空间分布同步优化方法四类<sup>[44]</sup>。统计推断误差最小优化方法一般以最小、最大或者平均统计推断误差作为优化目标函数进行优化布样,如普通克里金误差<sup>[45-47]</sup>、协同克里金误差<sup>[48]</sup>、泛克里金插值误差<sup>[5,49]</sup>、泛协克里金误差<sup>[7]</sup>和非均质表面估计模型误差(Mean of Surface with Nonhomogeneity, MSN)<sup>[2]</sup>等。地理空间分布优化方法通过设计候选样点与待估计点之间的距离函数并将距离函数作为优化目标,常见的目标函数包括平均最短距离最小(Minimization of the Mean of the Shortest Distances, MMSD)<sup>[50]</sup>、加权平均最短距离最小(Weighted Mean of the Shortest Distances, WMMSD)<sup>[51]</sup>、均方距离准则(Mean squared distance to sides, vertices, and boundaries)<sup>[52]</sup>,还包括同时优化插值制图精度与半变异函数精度的组合方法,如Simbahan等<sup>[53]</sup>提出结合MMSD与WM准则(Warrick-Myers criterion)<sup>[54]</sup>的样点空间分布优化方法。特征空间分布优化方法主要优化样点在辅助变量组成的特征空间中的分布,如按照辅助变量分布比例分层的等间距分层设计(Equal Range Stratification design, ER design)<sup>[55]</sup>、特征空间等概率间距分层与优化覆盖的拉丁超立方体方法(Latin Hypercube Sampling, LHS)<sup>[56]</sup>及条件拉丁超立方体方法(Conditioned Latin Hypercube Sampling, cLHS)<sup>[57]</sup>。地理空间与特征空间分布同步优化方法同时优化样点在地理空间与特征空间的分布,解决空间非平稳区域变量插值的采样优化问题,包括基于辅助变量的方差四叉树分步优化方法(Variance Quad-tree)<sup>[58]</sup>和地理空间与特征空间同步优化的空间条件拉超立方体

表3 土壤环境质量等级划分加密采样布局优化方法体系

Table 3 Soil environment quality grade with additional sampling layout optimization method system

优化目标依据 Rationale of optimization objective	采样布局优化方法 Sample layout optimization method		基本原理 Principle
估计值准确度	随机采样	分层随机抽样	基于先验知识、历史调查数据或辅助数据,将研究区划分为内部方差较小的若干个子区域,分别进行样点布设
	目的性采样	统计推断误差最小优化方法	选定统计推断模型,以研究区所有空间单元平均(或最大、或最小)的理论误差方差最小为优化目标布设样点
		地理空间分布优化方法	以样点的地理空间分布模式(如空间均匀分布、点对均衡分布)等作为优化目标布设样点
		特征空间分布优化方法	使用与土壤环境变量相关的辅助变量作为坐标轴,构建特征空间,以样点在特征空间的分布模式为优化目标,优化布设样点
等级划分精确度	地理空间与特征空间分布同步优化方法	组合地理空间和特征空间优化分布的目标函数,同时优化样点在地理空间和特征空间的分布	
	基于等级估计结果的布样方法	基于等级估计概率进行样点布设,将样点布设在容易发生等级错误划分的累积概率段内,如0.2-0.8	
	基于等级估计结果及其不确定性估计的布样方法	综合考虑等级估计概率和估计的不确定性进行加密样点布设	

方法 (Spatial Conditioned Latin Hypercube Sampling method, SCLHS)<sup>[2]</sup>等。

等级划分精确度加密采样布局优化方法直接以等级划分精确度为采样优化目标,主要分为基于等级估计结果的布样方法和基于等级估计结果及其不确定性的布样方法。前者使用前阶段获取的样本数据估计未采样单元属于某一环境质量等级的概率,并基于该概率指导加密样点布设。如 Garcia 等<sup>[59]</sup>认为,在使用非参数估计方法指导布样时,概率值低于 0.2、高于 0.8 的位置不大可能出现等级错划的情况,因此需要重点在概率值介于 0.2~0.8 之间的区域进行采样。该类方法忽略了等级概率估计本身的不确定性,同时划定的采样范围较大,效率不高。后者综合考虑等级估计结果及其不确定性指导样点布设。如 Van Meirvenne 等<sup>[32]</sup>提出的基于条件模拟的加密采样方法,该方法首先基于前期样本数据产生大量的模拟数据,以模拟值与阈值差异的(累积概率的)标准差和期望之比来衡量等级划分的不确定性,用于指导加密采样布局优化。该方法在估计中需要进行大量的条件模拟计算,同时还需要完整的累积概率分布函数,会引入累积概率分布函数内插和外插的误差。Juang 等<sup>[60]</sup>基于阶次地统计提出了包含等级划分中第一类和第二类错误的错划指数,高效地综合了等级估计结果及其不确定性。但是在该方法中物理阈值需被转化为标准阶次并在标准阶次中进行计算分析,在将物理阈值转化为标准阶次时会带来较大误差,如果样本在特征空间的代表性较差,则会从根本上影响该方法的准确性<sup>[61]</sup>。另外,该方法固定以估计值上下 3 倍的标准差作为两类错误的置信区间,不能基于具体需求调整对两类错误的偏好。Gao 等<sup>[62]</sup>基于多高斯模型提出了针对正态分布(或能够转化为正态分布)数据的等级划分错误指数,基于克里插值结果、误差方差和物理阈值量化发生等级划分错误的概率,如公式(1)所示。

$$Index = G(threshold, \hat{z}_0, \delta^2) / G(\hat{z}_0, \hat{z}_0, \delta^2) \quad (1)$$

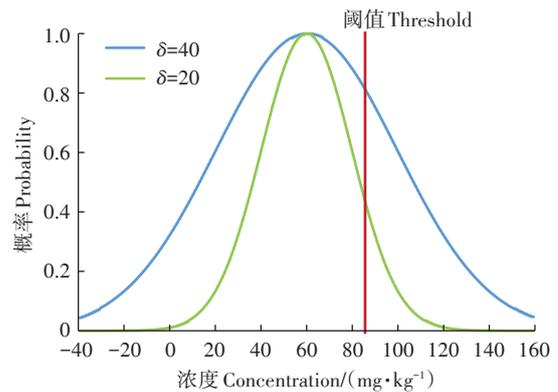
式中: $G$ 为高斯分布函数; $threshold$ 为等级划分的概率阈值; $\hat{z}_0$ 为克里金插值结果; $\delta^2$ 插值的误差方差。

如图 2 所示,其中蓝色与绿色曲线为由待估计点的插值结果和估计方差确定的归一化高斯分布,红色竖线为等级划分物理阈值( $90 \text{ mg} \cdot \text{kg}^{-1}$ ),其中图 2(a)为估计结果相同但估计误差不同的两点的等级划分错误指数,图 2(b)为估计误差相同但估计结果不同的两点的等级划分错误指数,由图可知,估计结果越接近阈值且估计误差越大,越容易发生等级划分错

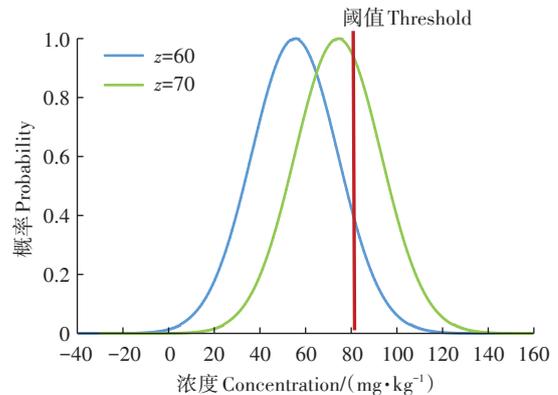
误。Gao 等<sup>[62]</sup>定义了基于该指数的等级划分加密采样优化目标函数,并基于空间模拟退火方法(Spatial Simulated Annealing, SSA)实现了加密样点位置优化布设。

由于土壤环境数据经常不符合正态分布并且难以转化为正态分布,Gao 等<sup>[41]</sup>基于指示克里金估计结果及其不确定性构建了等级划分错误指数,如公式(2)所示,它由阈值错划指数和误差错划指数两部分组成,阈值错划指数反映了累积概率估计值与阈值的接近程度,如图 3(a);误差错划指数反映了累积概率估计值的不确定性,如图 3(b);由于概率分布的分布为均匀分布,因此使用线性相加的方式组合阈值错划指数与误差错划指数形成综合错划指数,如图 3(c)。

$$I(x; Z_c) = I_c(x; Z_c) + I_e(x; Z_c) = F_c - |F_c - \hat{F}(x; Z_c)| + \delta(x; Z_c) \quad (2)$$



(a)相同估计结果不同估计误差  
(a)Error index for two locations with the same mean but different variances



(b)不同估计结果相同估计误差  
(b)Error index for two locations with the same variance but different means

图 2 基于多高斯模型的等级划分错误指数<sup>[62]</sup>

Figure 2 Grade classification Error index based on multi-Gaussian model<sup>[62]</sup>

式中: $x$ 为空间位置; $Z_c$ 为等级划分阈值; $I_c(x; Z_c)$ 为阈值错划指数; $I_e(x; Z_c)$ 为误差错划指数; $F_c$ 为等级划分概率阈值; $\hat{F}(x; Z_c)$ 为 $x$ 位置不超过 $Z_c$ 的累积概率估计值; $\delta(x; Z_c)$ 为累积概率阈值的估计误差标准差。

Gao等<sup>[41]</sup>进一步推荐采用平均错划指数最小、最大错划指数最小等作为优化目标函数,采用图1的自适应方法确定概率阈值,并基于优化算法进行土壤等级划分加密采样布局优化。基于多高斯模型的等级

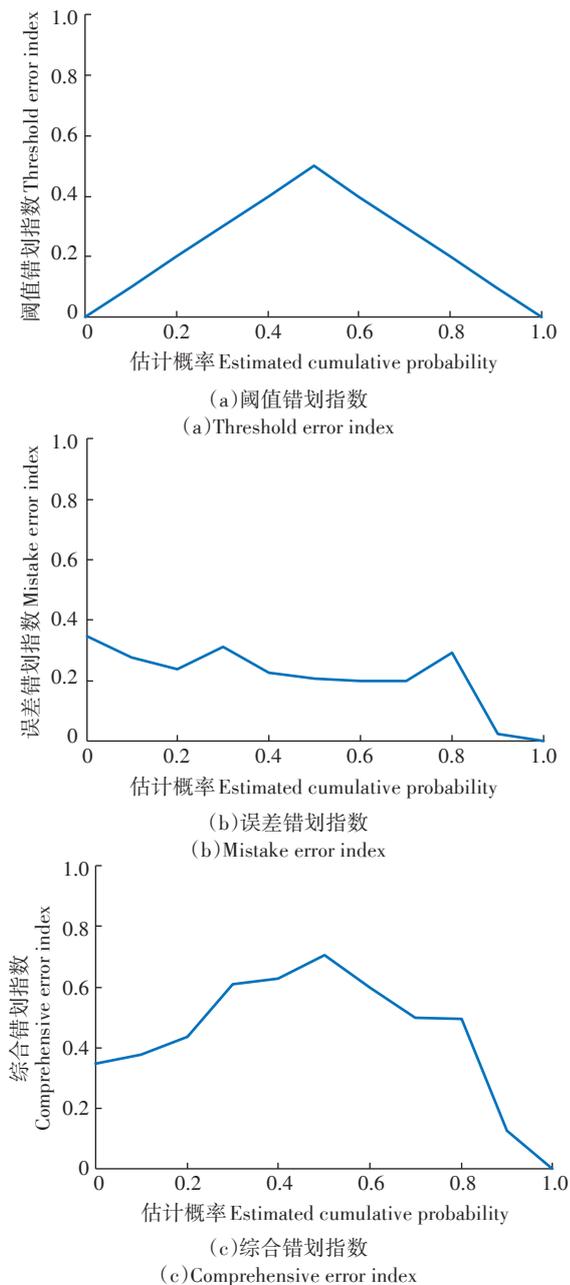


图3 基于指示克里金的等级划分错误指数 Gao等<sup>[41]</sup>

Figure 3 grade classification Error index based on Indicator Kriging<sup>[41]</sup>

划分错误指数和基于指示克里金的等级划分错误指数分别针对正态分布数据和非正态分布数据,通过简单的数学变换综合考虑估计结果与阈值关系的不确定性,能够刻画出土壤环境质量等级可能的边界,可在多阶段采样或补充调查的加密采样布局优化中选用。

### 3.2 方法对比分析

环境质量等级划分方法比较分析如表4所示。分层随机抽样能够考虑环境变量的分层异质性,提高样本代表性,但是关注样点对整个区域的代表性,不能重点关注等级过渡地带,因此适用于均值估计,而不适用于等级划分。在目的性抽样方法中,统计推断误差最小优化方法能够充分利用统计推断模型及参数直接提高样点布设效率;地理空间分布优化方法基于空间自相关性指导样点布设,不需要先验知识和辅助数据;特征空间分布优化方法能够充分利用相关辅助变量指导样点布设,提高样本对环境变量变化特征的代表性;而地理空间与特征空间分布同步优化方法则兼顾了地理空间与特征空间分布,能够同时提高样本对地理空间和变量变化特征的代表性。但是总体而言,依据估计值准确度加密采样布局优化方法没有考虑土壤环境质量等级划分加密采样优化的特殊性,不能重点关注等级过渡地带。而在土壤环境质量等级划分加密采样布局设计中,由于空间自相关性的存在,空间单元的质量等级具有空间连续性,在等级过渡地带(即在值域上接近阈值的区域)更容易发生等级错划,应加密样点;而在远离过渡地带(即在值域上远小于或大于阈值)的区域,增加过多的样点无益于等级划分准确性的提高<sup>[32]</sup>。因此,依据估计值准确度的加密采样方法效率较低,适用于土壤环境质量空间插值,而不适用于土壤环境质量等级划分。

依据等级划分精确度的加密采样布局优化方法直接以等级划分精确度为采样优化目标,将样点布设在容易发生等级错误划分的区域<sup>[2]</sup>。其中基于等级估计结果的加密采样布局优化方法,使用前阶段采样数据估计每个空间单元等级概率并据此进行加密样点布设,侧重于在等级过渡地带进行加密布点。但是由于前阶段采样的样本量不足,导致对等级概率的估计存在不确定性,而该类方法忽略了等级概率估计本身的不确定性,会导致等级过渡带划分不够精细和准确。基于等级估计结果及其不确定性估计的加密采样布局优化方法,在划定不同环境质量等级过渡带时综合考虑了等级概率估计结果及其不确定性,能够获得较合理的等级错误划分概率,进而提高加密采样的

表4 环境质量等级划分加密采样布局优化方法体系

Table 4 Soil environment quality grade with additional sampling layout optimization method system

优化目标依据 Rationale of optimization objective	采样布局优化方法 Sample layout optimization method	优点 Advantage	不足 Shortcoming
估计值准确度	分层随机抽样	考虑环境变量的分层异质性,提高样本代表性	关注样点对整个区域的代表性,不能重点关注等级过渡地带,适用于均值估计
	统计推断误差最小优化方法	充分利用统计推断模型及参数,直接提高样点布设效率	对整个研究区域同等对待,不能重点关注等级过渡地带,使得加密采样效率较低,适用于空间插值
	地理空间分布优化方法	基于空间自相关性指导样点布设,不需要先验知识和辅助数据	
	特征空间分布优化方法	充分利用相关辅助变量指导样点布设,提高样本对环境变量变化特征的代表性	
等级划分精确度	地理空间与特征空间分布同步优化方法	兼顾地理空间与特征空间分布,同时提高样本对地理空间和环境变量变化特征的代表性	
	基于等级估计结果的布样方法	使用等级概率确定等级过渡带,强调在等级过渡地带进行加密布点	忽略了等级概率估计本身的不确定性,等级过渡带划分不够精细
	基于等级估计结果及其不确定性估计的布样方法	综合考虑等级概率估计结果及其不确定性确定等级过渡地带,提高加密采样的布设效率	未充分利用多维辅助变量信息,提高加密样点布设效率

布设效率。目前该类方法不足之处在于未充分利用多维辅助变量信息,提高加密样点布设效率。

## 4 结论与展望

### 4.1 结论

目前,土壤环境质量等级划分统计推断与加密采样布局优化研究已经取得了较大进展。环境质量等级划分统计推断方法可以分为物理阈值法和概率阈值法,两类方法各有优缺点。其中物理阈值法由于需要首先精确估计未采样点的含量,而不是直接估计环境质量等级,因此会引入额外的步骤与误差。而概率阈值法直接以环境质量等级估计为目标,其中非参数估计不需要完整的概率分布,只需要估计与等级划分阈值大小关系的概率,更加适合土壤环境质量等级划分。对应于两类环境质量等级划分统计推断方法,加密采样布局优化也可以分为依据估计值准确度和等级划分精确度的两类方法。前者以研究区域各个空间单元处污染物浓度精确估计为目标,对整个研究区域同等对待,未针对等级边界,采样效率较低。后者以降低等级划分错误为目标,重点将样点布设在等级边界过渡地带,具有更好的采样效率。其中基于等级估计结果的布样方法仅采用候选样点属于某一环境质量等级的估计概率指导加密样点布设,未考虑估计概率的不确定性。而基于等级估计结果及其不确定性估计的布样方法,综合考虑了具有不确定性参数的估计值与等级划分阈值的关系,能够刻画出土壤环境质量等级之间可能的边界带,为多阶段采样或补充调

查的加密样调查提供加密采样布局优化方案。

### 4.2 展望

虽然土壤环境质量等级划分统计推断与加密采样布局优化方法已经初步形成一套理论方法体系,但是土壤环境质量空间格局复杂,影响因素多样,为了能满足进一步满足土壤环境质量等级划分实际工作需求,作者认为还需要进一步研究解决以下三方面的问题:

(1)土壤环境相关辅助数据使用问题:随着土壤环境相关数据的积累,如何在土壤环境质量等级划分的加密采样布局优化和统计推断中充分应用辅助数据,以提高采样效率和等级划分精度。

(2)样本量的确定:如何根据具体的应用需求(如划分修复区边界、划分农用地环境质量类别等),确定环境质量等级划分的等级属性精确度与等级边界空间的精确度参数,进而结合这些参数确定加密调查所需样本量。

(3)如何进行多阶段协同采样优化:目前统计土壤环境质量等级划分加密采样优化大都针对后续加密调查阶段的布样优化,对于没有历史调查数据的区域(如确定废弃工矿区的污染修复边界),如何综合考虑采样成本与等级划分精度,划分调查阶段,在阶段间分配样本量并制定各阶段的样点优化布局方案。

随着土壤污染防治法的实施,我国土壤环境质量等级划分需求将不断增加。目前即将完成全国农用地土壤环境质量类别划分,下一步将开展农用地土壤环境质量类别动态调整和安全利用工作,需要开展大量的土壤环境质量等级划分与加密采样工作,由于我

国已经积累了大量农用地土壤污染采样数据,建议后续工作中采用具有针对性的土壤环境质量等级划分加密采样布局优化和统计推断方法(如概率阈值法),基于等级划分精确度的加密采样布局优化方法或者具有针对性的其他先进方法。随着我国土壤污染修复行业不断发展壮大,在划定修复区时,建议采用先进的土壤环境质量等级划分采样布局优化和统计推断方法,在污染修复区范围划定中同时考虑修复成本与健康风险,实现科学决策。

#### 参考文献:

- [1] 徐笠, 陆安祥, 王纪华. 土壤重金属污染评价方法研究进展及其在禁产区划分工作中的比较应用[J]. 食品安全质量检测学报, 2016, 7(3): 1145-1151. XU Li, LU An-xiang, WANG Ji-hua. Research progress of soil pollution assessment methods and their application in the division of non-agricultural producing region[J]. *Journal of Food Safety and Quality*, 2016, 7(3): 1145-1151.
- [2] Gao B B, Wang J F, Fan H M, et al. A stratified optimization method for a multivariate marine environmental monitoring network in the Yangtze River estuary and its adjacent sea[J]. *International Journal of Geographical Information Science*, 2015, 29(8): 1332-1349.
- [3] 高祖新, 陈华钧, 黄志远. 概率论与数理统计[M]. 南京: 南京大学出版社, 1995: 134-144. GAO Zu-xin, CHEN Hua-jun, HUANG Zhi-yuan. *Probability and statistics*[M]. Nanjing: Nanjing University Press, 1995: 134-144.
- [4] 冯士雍, 倪加勋, 邹国华. 抽样调查理论与方法[M]. 二版. 北京: 中国统计出版社, 2012: 65. FENG Shi-yong, NI Jia-xun, ZOU Guo-hua. *Survey sampling: Theory and methods*[M]. Second Edition. Beijing: China Statistical Publishing House, 2012: 65.
- [5] Brus D J, Heuvelink G B M. Optimization of sample patterns for universal kriging of environmental variables[J]. *Geoderma*, 2007, 138(1/2): 86-95.
- [6] Hu M G, Wang J F. A spatial sampling optimization package using MSN theory[J]. *Environmental Modelling & Software*, 2011, 26(4): 546-548.
- [7] Ge Y, Wang J H, Heuvelink G B M, et al. Sampling design optimization of a wireless sensor network for monitoring ecohydrological processes in the Babao River basin, China[J]. *International Journal of Geographical Information Science*, 2015, 29(1): 92-110.
- [8] Zimmerman D L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction[J]. *Environmetrics: The Official Journal of the International Environmetrics Society*, 2006, 17(6): 635-652.
- [9] De Grujter J, Brus D J, Bierkens M F P, et al. Sampling for natural resource monitoring[M]. Berlin: Springer Science & Business Media, 2006: 11-25.
- [10] Brus D J, De Grujter J J. Random sampling or geostatistical modeling? Choosing between design-based and model-based sampling strategies for soil (with discussion) [J]. *Geoderma*, 1997, 80(1/2): 1-44.
- [11] Wang J, Haining R, Cao Z. Sample surveying to estimate the mean of a heterogeneous surface: Reducing the error variance through zoning [J]. *International Journal of Geographical Information Science*, 2010, 24(4): 523-543.
- [12] Wang J F, Jiang C S, Hu M G, et al. Design-based spatial sampling: Theory and implementation[J]. *Environmental Modelling & Software*, 2013, 40: 280-288.
- [13] Wang J F, Gao B B, Alfred S. The spatial statistic trinity: A generic framework for spatial sampling and inference[J]. *Environmental Modelling & Software*, 2020, 134: 104835.
- [14] 侯景儒, 尹镇南, 李维明, 等. 实用地质统计学[M]. 北京: 地质出版社, 1988: 31-68. HOU Jing-ru, YIN Zhen-nan, LI Wei-ming, et al. *Practical geostatistics*[M]. Beijing: Geological Publishing House, 1998: 31-68.
- [15] Najafian A, Dayani M, Motaghian H R, et al. Geostatistical assessment of the spatial distribution of some chemical properties in calcareous soils[J]. *Journal of Integrative Agriculture*, 2012, 11(10): 1729-1737.
- [16] 李润林, 姚艳敏, 唐鹏钦, 等. 县域耕地土壤锌含量的协同克里格插值及采样数量优化[J]. 土壤通报, 2013, 44(4): 830-838. LI Run-lin, YAO Yan-min, TANG Peng-qin, et al. Optimization of spatial interpolation and sampling size for soil zinc content in cultivated land at county scale using cokriging method[J]. *Chinese Journal of Soil Science*, 2013, 44(4): 830-838.
- [17] Li J, Heap A D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors[J]. *Ecological Informatics*, 2011, 6(3/4): 228-241.
- [18] Li J, Heap A D. Spatial interpolation methods applied in the environmental sciences: A review[J]. *Environmental Modelling & Software*, 2014, 53: 173-189.
- [19] Minasny B, McBratney A B. Methodologies for global soil mapping[M]. Springer, Dordrecht: Digital soil mapping, 2010: 429-436.
- [20] Zimmerman D L, Holland D M. Complementary co-kriging: Spatial prediction using data combined from several environmental monitoring networks[J]. *Environmetrics: The Official Journal of the International Environmetrics Society*, 2005, 16(3): 219-234.
- [21] Goovaerts P. Geostatistics for natural resources evaluation[M]. Oxford: Oxford University Press, 1997: 42.
- [22] 史文娇, 岳天祥, 石晓丽, 等. 土壤连续属性空间插值方法及其精度的研究进展[J]. 自然资源学报, 2012, 27(1): 163-175. SHI Wen-jiao, YUE Tian-xiang, SHI Xiao-li, et al. Research progress in soil property interpolators and their accuracy[J]. *Journal of Natural Resources*, 2012, 27(1): 163-175.
- [23] 史文娇, 岳天祥, 石晓丽, 等. 高风险重金属污染土壤识别研究方法综述[J]. 土壤, 2012, 44(2): 197-202. SHI Wen-jiao, YUE Tian-xiang, SHI Xiao-li, et al. Identification methods of soil pollution for high risk heavy metals: A review[J]. *Soils*, 2012, 44(2): 197-202.
- [24] Journel A G, Huijbregts C J. Mining geostatistics[M]. London: Academic Press, 1978: 94-107.
- [25] 罗明, 裴韬. 空间软数据及其插值方法研究进展[J]. 地理科学进

- 展, 2009, 28(5):663-672. LUO Ming, PEI Tao. Review on soft spatial data and its spatial interpolation methods[J]. *Progress in Geography*, 2009, 28(5):663-672.
- [26] Zhu A X, Yang L, Li B, et al. Construction of membership functions for predictive soil mapping under fuzzy logic[J]. *Geoderma*, 2010, 155(3/4):164-174.
- [27] Yang L, Zhu A X, Zhao Y G, et al. Regional soil mapping using multi-grade representative sampling and a fuzzy membership-based mapping approach[J]. *Pedosphere*, 2017, 27(2):344-357.
- [28] 谢军, 秦承志, 肖桂荣, 等. 模糊聚类方法在南方红壤小流域土壤属性制图中的应用——以长汀朱溪河小流域为例[J]. 中国水土保持科学, 2015, 13(5):132-139. XIE Jun, QIN Cheng-zhi, XIAO Gui-rong, et al. Soil property mapping using fuzzy clustering method in small watershed of the red soil region in southern China[J]. *Science of Soil and Water Conservation*, 2015, 13(5):132-139.
- [29] Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author)[J]. *Statistical Science*, 2001, 16(3):199-231.
- [30] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016:4-15. ZHOU Zhi-hua. *Machine Learning*[M]. Beijing: Tsinghua University Press, 2016:4-15.
- [31] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望[J]. 地理科学进展, 2018, 37(1):66-78. ZHU A-xing, YANG Lin, FAN Nai-qing, et al. The review and outlook of digital soil mapping[J]. *Progress in Geography*, 2018, 37(1):66-78.
- [32] Van Meirvenne M, Goovaerts P. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold[J]. *Geoderma*, 2001, 102(1/2):75-100.
- [33] Isaaks E H, Srivastava R M. An introduction to applied geostatistics [M]. Oxford: Oxford University Press, 1989:52-85.
- [34] 姚荣江, 杨劲松, 杨奇勇, 等. 禹城地区土壤铅含量空间分布的指示克里格估值[J]. 生态环境学报, 2011, 20(12):1912-1918. YAO Rong-jiang, YANG Jin-song, YANG Qi-yong, et al. Indicator Kriging of spatial distribution of soil Pb content in Yucheng City[J]. *Ecology and Environmental Sciences*, 2011, 20(12):1912-1918.
- [35] Sekulić A, Kilibarda M, Heuvelink G B, et al. Random forest spatial interpolation[J]. *Remote Sensing*, 2020, 12(10):1687.
- [36] Hengl T, Nussbaum M, Wright M N, et al. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables[J]. *PeerJ*, 2018, 6:e5518.
- [37] Journé A G. Nonparametric estimation of spatial distributions[J]. *Journal of the International Association for Mathematical Geology*, 1983, 15(3):445-468.
- [38] Antunes I, Albuquerque M T D. Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal)[J]. *Science of the Total Environment*, 2013, 442:545-552.
- [39] Chica-Olmo M, Luque-Espinar J A, Rodríguez-Galiano V, et al. Categorical Indicator Kriging for assessing the risk of groundwater nitrate pollution: The case of Vega de Granada aquifer (SE Spain)[J]. *Science of the Total Environment*, 2014, 470:229-239.
- [40] 高秉博. 时空非平稳区域多目标抽样优化方法[D]. 北京: 中国科学院大学, 2015:79-104. GAO Bing-bo. Multi-targets sampling optimization method for a spatio-temporal non-stationary region[D]. Beijing: The University of Chinese Academy of Sciences, 2015:79-104.
- [41] Gao B B, Liu Y, Pan Y C, et al. Error index for additional sampling to map soil contaminant grades[J]. *Ecological Indicators*, 2017, 77:129-138.
- [42] 李晓岚, 高秉博, 周艳兵, 等. 基于时空不确定性分析的北京市农田土壤重金属镉含量等级划分[J]. 农业环境科学学报, 2019, 38(2):307-316. LI Xiao-lan, GAO Bing-bo, ZHOU Yan-bing, et al. Classification of soil heavy metal cadmium content grade in Beijing farmland based on spatio-temporal uncertainty analysis[J]. *Journal of Agro-Environment Science*, 2019, 38(2):307-316.
- [43] Wang J F, Stein A, Gao B B, et al. A review of spatial sampling[J]. *Spatial Statistics*, 2012, 2:1-14.
- [44] Gao B B, Pan Y C, Chen Z Y, et al. A spatial conditioned Latin hypercube sampling method for mapping using ancillary data[J]. *Transactions in GIS*, 2016, 20(5):735-754.
- [45] 陈天恩, 陈立平, 王彦集, 等. 基于地统计的土壤养分采样布局优化[J]. 农业工程学报, 2009, 25(增刊2):49-55. CHEN Tian-en, CHEN Li-ping, WANG Yan-ji, et al. Optimal arrangement of soil nutrient sampling based on geo-statistics[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2009, 25(Suppl2):49-55.
- [46] 杨建宇, 汤赛, 郟文聚, 等. 基于Kriging估计误差的县域耕地等级监测布样方法[J]. 农业工程学报, 2013, 29(9):223-230, 297. YANG Jian-yu, TANG Sai, YUN Wen-ju, et al. Sampling method for monitoring classification of cultivated land in county area based on Kriging estimation error[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(9):223-230, 297.
- [47] Shen Y, Wu Y. Optimization of marine environmental monitoring sites in the Yangtze River estuary and its adjacent sea, China[J]. *Ocean & Coastal Management*, 2013, 73:92-100.
- [48] Passarella G, Vurro M, D'agostino V, et al. Cokriging optimization of monitoring network configuration based on fuzzy and non-fuzzy variogram evaluation[J]. *Environmental Monitoring and Assessment*, 2003, 82(1):1-21.
- [49] 葛咏, 王江浩, 王劲峰, 等. 基于回归克里格的生态水文无线传感器网络布局优化[J]. 地球科学进展, 2012, 27(9):1006-1013. GE Yong, WANG Jiang-hao, WANG Jin-feng, et al. Regression Kriging model-based sampling optimization design for the eco-hydrology wireless sensor network[J]. *Advance in Earth Science*, 2012, 27(9):1006-1013.
- [50] Van Groenigen J W, Stein A. Constrained optimization of spatial sampling using continuous simulated annealing[J]. *Journal of Environmental Quality*, 1998, 27(5):1078-1086.
- [51] Van Groenigen J W. Spatial simulated annealing for optimizing sampling[M]. Springer, Dordrecht: GeoENV I: Geostatistics for Environmental Applications, 1997:351-361.
- [52] Stevens Jr D L. Spatial properties of design-based versus model-based approaches to environmental sampling[C]//7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2006:119-125.

- [53] Simbahan G C, Dobermann A. Sampling optimization based on secondary information and its utilization in soil carbon mapping[J]. *Geoderma*, 2006, 133(3/4):345-362.
- [54] Warrick A W, Myers D E. Optimization of sampling locations for variogram calculations[J]. *Water Resources Research*, 1987, 23(3):496-500.
- [55] Hengl T, Rossiter D G, Stein A. Soil sampling strategies for spatial prediction by correlation with auxiliary maps[J]. *Soil Research*, 2003, 41(8):1403-1422.
- [56] McKay M D, Beckman R J, Conover W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code[J]. *Technometrics*, 2000, 42(1):55-61.
- [57] Minasny B, McBratney A B. A conditioned Latin hypercube method for sampling in the presence of ancillary information[J]. *Computers & Geosciences*, 2006, 32(9):1378-1388.
- [58] Lin Y P, Chu H J, Huang Y L, et al. Monitoring and identification of spatiotemporal landscape changes in multiple remote sensing images by using a stratified conditional Latin hypercube sampling approach and geostatistical simulation[J]. *Environmental Monitoring and Assessment*, 2011, 177(1/2/3/4):353-373.
- [59] Garcia M, Froidevaux R. Application of geostatistics to 3D modelling of contaminated sites: A case study[M]. Springer, Dordrecht: GeoENV I: Geostatistics for Environmental Applications, 1997:309-325.
- [60] Juang K W, Liao W J, Liu T L, et al. Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site[J]. *Science of the Total Environment*, 2008, 389(1):20-28.
- [61] Juang K W, Lee D Y, Ellsworth T R. Using rank-order geostatistics for spatial interpolation of highly skewed data in a heavy-metal contaminated site[J]. *Journal of Environmental Quality*, 2001, 30(3):894-903.
- [62] Gao B B, Lu A X, Pan Y C, et al. Additional sampling layout optimization method for environmental quality grade classifications of farmland soil[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(12):5350-5358.