

DOI: 10.11766/trxb201911150515

姚彩燕, 刘绍贵, 乔婷, 龙军, 于东升, 史学正, 邢世和, 陈瀚阅, 张黎明. 基于时空变异的旱地土壤有机碳高效采样策略研究[J]. 土壤学报, 2021, 58 (3): 638–648.

YAO Caiyan, LIU Shaogui, QIAO Ting, LONG Jun, YU Dongsheng, SHI Xuezheng, XING Shihe, CHEN Hanyue, ZHANG Liming. Strategy for Efficient Sampling Strategy of Upland Soil Based on Spatiotemporal Variation of the Soil[J]. Acta Pedologica Sinica, 2021, 58 (3): 638–648.

## 基于时空变异的旱地土壤有机碳高效采样策略研究<sup>\*</sup>

姚彩燕<sup>1, 2, 3</sup>, 刘绍贵<sup>4</sup>, 乔 婷<sup>1, 3</sup>, 龙 军<sup>1, 3</sup>, 于东升<sup>5</sup>, 史学正<sup>5</sup>, 邢世和<sup>1, 3</sup>,  
陈瀚阅<sup>1, 3</sup>, 张黎明<sup>1, 3†</sup>

(1. 福建农林大学资源与环境学院, 福州 350002; 2. 福建农林大学公共管理学院, 福州 350002; 3. 土壤生态系统健康与调控福建省高校重点实验室, 福州 350002; 4. 扬州市农业环境监测站, 江苏 扬州 225603; 5. 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008)

**摘要:** 土壤有机碳(SOC)空间分布具有时序差异性, 明确样点数量对不同时期SOC预测精度影响是制定高效采样策略的基础。选取 $3.93 \times 10^4 \text{ km}^2$ 江苏北部旱地作为案例区, 运用普通克里金插值方法, 分析样点数量对不同时期SOC空间预测精度的影响。结果表明: 不同数量样点数据集下1980年苏北旱地SOC预测值与实测值的相关系数r和均方根误差RMSE变幅分别在0.15~0.56和2.09~2.63 g·kg<sup>-1</sup>之间, 当样点数量大于75%时, 预测精度较高且能达到相对稳定水平, 最佳采样数目在563个左右; 而2008年r和RMSE变幅分别在0.24~0.60和2.11~2.62 g·kg<sup>-1</sup>之间, 预测精度对于样点数量的变化更为敏感, 70%的样点数量即可达到相对稳定水平, 最佳采样数目在526个左右, 这表明不同时期SOC空间预测精度对于样点数量变化的响应不同, 土壤属性的空间自相关性越大, 预测精度对于采样数量的敏感性越强, 空间信息达到饱和状态所需样点数量也相对较少; 此外, 本研究也发现在SOC高、低值等关键区域设置足够的样点数量是提高土壤空间预测效果的重要手段之一。

**关键词:** 土壤有机碳; 样点数量; 预测精度; 空间变异; 采样设计

中图分类号: S15 文献标志码: A

## Strategy for Efficient Sampling Strategy of Upland Soil Based on Spatiotemporal Variation of the Soil

YAO Caiyan<sup>1, 2, 3</sup>, LIU Shaogui<sup>4</sup>, QIAO Ting<sup>1, 3</sup>, LONG Jun<sup>1, 3</sup>, YU Dongsheng<sup>5</sup>, SHI Xuezheng<sup>5</sup>, XING Shihe<sup>1, 3</sup>,  
CHEN Hanyue<sup>1, 3</sup>, ZHANG Liming<sup>1, 3†</sup>

(1. College of Resource and Environment Science, Fujian Agriculture and Forestry University, Fuzhou, 350002, China; 2. School of Public

\* 国家自然科学基金项目(41971050)、福建省自然科学基金项目(2019J01660)、福建省科技计划项目(2017N5006)和福建农林大学国际合作项目(KXGH17017)共同资助 Supported by the National Science Foundation of China (No. 41971050), the National Science Foundation of Fujian Province, China (No. 2019J01660), the Science and Technology Planning Project of Fujian Province, China (No. 2017N5006) and the International Cooperative Research Program at Fujian Agriculture and Forestry University (No. KXGH17017)

† 通讯作者 Corresponding author, E-mail: fjaulmzhang@163.com

作者简介: 姚彩燕(1995—), 女, 广西贵港人, 硕士研究生, 从事土壤碳氮循环与GIS应用。E-mail: 895123404@qq.com

收稿日期: 2019-11-15; 收到修改稿日期: 2020-01-16; 网络首发日期(www.cnki.net): 2020-03-11

Administration, Fujian Agriculture and Forestry University, Fuzhou, 350002, China; 3. University Key Lab of Soil Ecosystem Health and Regulation in Fujian, Fuzhou, 350002, China; 4. Yangzhou Agriculture Environmental Monitoring Station, Yangzhou Jiangsu 225603, China; 5. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

**Abstract:** 【Objective】 Soil organic carbon (SOC) varies sharply with time and space, as it is always subject to influences of various soil-forming factors, natural environmental factors and human activities. So to determine how an appropriate number of sampling sites could affect accuracy of the prediction of SOC in different time periods is the basis for formulating a scientific strategy for high-efficient soil sampling. 【Method】 In this study, a tract of upland ( $3.93 \times 10^4 \text{ km}^2$ ) in North Jiangsu was delineated and selected as a case area, and the Ordinary Kriging interpolation method commonly used in soil science was adopted in analyzing influences of numbers of sampling sites on prediction and mapping of SOC in different time periods. The study was designed to have 20 treatments, which were set in accordance with the principle of 5% decrease in number. 【Result】 Results show that with the number of soil sampling sites decreasing from 100% to 5%, correlation coefficient ( $r$ ) between the predicted value and the measured value of SOC in 1980 and in 2008 varied in the range of 0.15-0.56 and 0.24-0.60 and root mean square error in the range of 2.09-2.63 and  $2.11\text{-}2.62 \text{ g}\cdot\text{kg}^{-1}$ , respectively. As in 1980, the SOC in the studied region varied quite slightly in spatial autocorrelation and quite drastically and locally, its prediction improved slowly and unsteadily in accuracy, and around 563 samples were needed to make the prediction relatively reliable. However, in 2008, the SOC in the region varied quite sharply in spatial autocorrelation, but mildly locally, and hence its prediction was very sensitive in accuracy to variation of the number of sampling sites. So 526 soil sampling sites were enough to ensure stable prediction accuracy. Standard root mean square errors of the 20 treatments in terms of number of sampling sites varied in the range of 0.34-0.43 and  $0.20\text{-}0.25 \text{ g}\cdot\text{kg}^{-1}$ , in 1980 and in 2008, respectively, and spatial prediction was higher in 2008 than in 1980 in accuracy when the numbers of soil sampling sites were the same. 【Conclusion】 Results of this study indicate that the optimal number of soil sampling sites and their prediction accuracy in the same area are not fixed, but determined in the light of spatial variability of soil attributes, distribution and spatial layout of the sampling sites in each time period. As environment, climate and farmland management practices all vary with time period, SOC content does too in spatial structure and layout, which will greatly affect the optimal number of sampling sites relative to time period.

**Key words:** Soil organic carbon; Number of sampling sites; Prediction accuracy; Spatial variation ; Sampling design

土壤有机碳 (Soil Organic Carbon, SOC) 作为全球陆地生态系统最大的有机碳库, 不仅是评价土壤肥力的重要指标, 而且其含量的轻微变化也可导致大气二氧化碳 ( $\text{CO}_2$ ) 浓度显著变化, 进而影响全球碳平衡<sup>[1]</sup>。受气候、地形、土地利用和土壤类型等多方面因素综合影响, SOC 具有很高的时空变异性, 不同时间和空间位置的 SOC 含量特征复杂且不确定, 准确了解 SOC 时空分布变化, 对于制定科学合理的碳汇管理措施具有重要意义<sup>[2]</sup>。样点是土壤特性时空变异的信息载体, 是土壤调查与空间制图的基础。在进行土壤采样时, 任何采样过程都不可能获得土壤的完全信息, 只能对一定数量的土壤样本数据通过插值形式实现土壤属性空间分布的定量表达<sup>[3]</sup>。然而, 样点数量过小, 土壤样品的偶然性增加, 缺乏代表性, 空间预测将会产生不可接受的误差; 但盲目增大采样数量, 不仅预测误差不会持续减小,

反而加大了采样工作量和样品分析周期, 浪费人力财力物力<sup>[4]</sup>。因此, 明确不同样点数量与 SOC 预测精度的响应关系是制定高效采样策略的基础。

近年来, 国内外学者对不同区域的合理采样数目开展了一系列研究<sup>[5-7]</sup>。但当前最佳土壤采样数目的确定主要集中在若干个样点数量的土壤属性空间预测精度对比上, 并没有量化预测精度与采样数量之间的关系。此外, 目前的大多数研究一般局限在单一时间尺度下, 未考虑不同时期土壤属性差异对适宜采样数目影响。众所周知, 土壤是动态的, 经过一定时间, SOC 含量通常会出现一定范围的变化, 在某些情况下, SOC 时间变异将会超过空间变异的影响<sup>[8]</sup>。因此, 为了准确估计 SOC 含量, 基于样点的预测制图必须要考虑空间和时间维度。旱地作为我国主要的农田类型, 占全国耕地面积的 76.9%, 其 SOC 密度显著低于世界平均水平, 固碳

潜力巨大<sup>[9]</sup>。但由于人类活动以及农田管理措施的影响,该土地利用类型下的土壤碳库呈现强烈的时空变异性,明确不同时期旱地SOC预测精度与采样数量之间的关系是合理制定我国农田管理措施的基础。江苏省北部地区85%的耕地为旱地土壤,尽管面积仅占全国旱地土壤总面积的3.7%,却固定了全国10.2%的旱地SOC,具有较大的“碳汇”空间<sup>[10]</sup>。此外,该地区旱地的主要土壤类型为潮土、盐土和砂姜黑土,分别占全区旱地土壤总面积的52.67%、24.31%和8.13%,这与黄淮海平原主要旱地土壤类型面积比例相一致,有利于研究成果向大区域拓展<sup>[11]</sup>。基于此,本研究以江苏省北部 $3.93 \times 10^4 \text{ km}^2$ 旱地(简称“苏北旱地”)为研究区,利用1980年全国第二次土壤普查和2008年农业部测土配方施肥项目的实测样点数据,通过随机抽样的方式设置20组不同数量的样点数据集,分析这两个时期研究区SOC的最佳采样数量,并在此基础上揭示土壤属性时空变异性下预测精度与样点数量之间量化关系,研究结果可为苏北旱地和未来我国第三次土壤普查制定合理的有机碳野外调查方案提供理论依据。

## 1 材料与方法

### 1.1 研究区概况

苏北旱地( $116^{\circ}21' \sim 120^{\circ}54'E$ ,  $32^{\circ}43' \sim 35^{\circ}7'N$ )地处黄淮平原与江淮平原的过渡地带,是中国沿海经济带的重要组成部分(图1)。该地区属于亚热带

向暖温带过渡的季风性气候,年平均降水量800~1200 mm,年平均气温13~16 °C,全年平均日照时数2000~2600 h,无霜期约220 d。境内地势平坦,呈西北高、东南低的趋势,地形主要以平原为主,海拔高度在0~50 m之间。研究区面积最大的旱地土壤类型为潮土和盐土,其次为褐土、砂姜黑土和棕壤;而石质土、紫色土和石灰土较小。成土母质主要包括湖相冲积物、黄泛冲积物、河海相冲积物、其他各河流冲积物以及下蜀黄土等。苏北是我国重要的粮食产区,种植制度主要是夏玉米和冬小麦轮作<sup>[12]</sup>。

### 1.2 数据来源与处理

1980年和2008年表层样点数据(0~20 cm)分别来自全国第二次土壤普查中苏北地区29个县(市、区)在《县级土种志》、《地市级土种志》、《省级土种志》和《中国土种志》中记录的旱地剖面资料及农业部测土配方施肥项目。所有土壤样点信息均包括地理位置描述和理化性质等,其中SOC含量可由有机质含量乘以Bemmelen转换系数(0.58)计算。考虑到样点采集与化学分析过程的不确定性,对两期样点的SOC含量进行正态分布检验后进行异常值剔除处理,所采用的剔除方式是结合数据的常规统计学特征和空间位置,将每个样点的SOC含量与总体及其邻近8个样点的均值和标准差进行比较,如果样点SOC含量在总体均值的五倍标准差之外,且大于或是小于邻近样点均值的三倍标准差,则视为异常值剔除,否则予以保留<sup>[13]</sup>。最后,1980年和2008年可利用的数据集分别为884个和1081个(图2)。

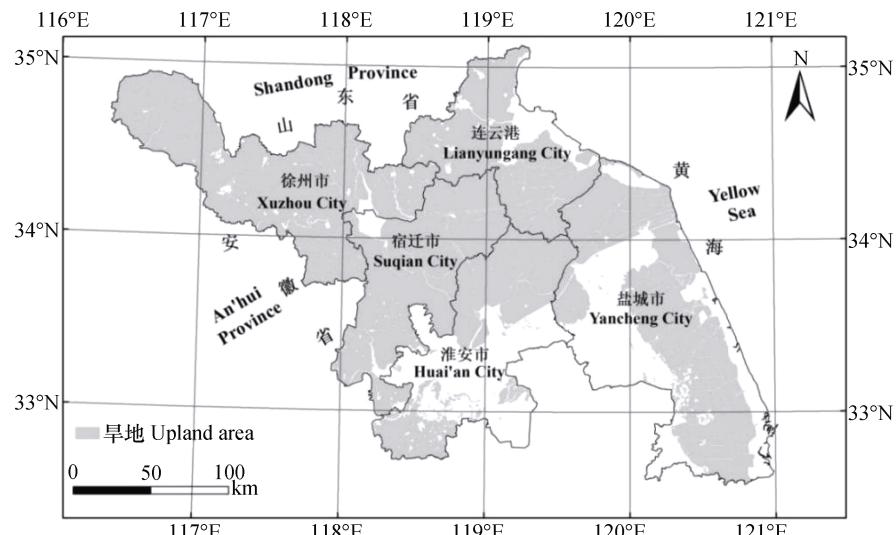


图1 苏北旱地地理位置

Fig. 1 Geographical location map of the upland area studied in Northern Jiangsu

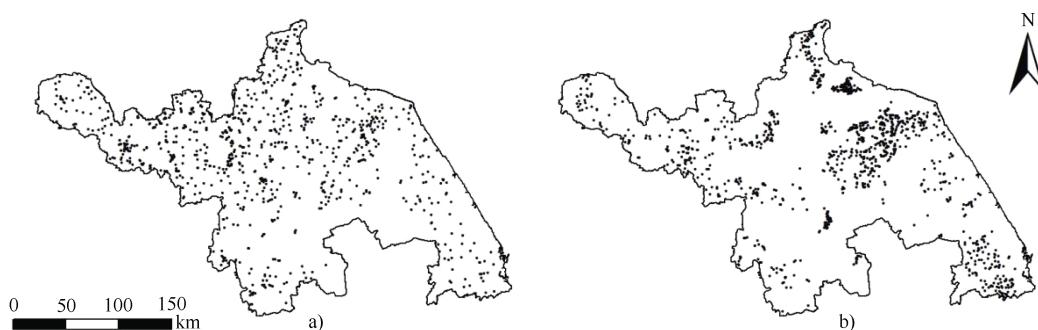


图 2 1980 年 (a) 和 2008 年 (b) 江苏北部地区旱地土壤样点分布  
Fig. 2 Distribution of upland soil sampling sites in Northern Jiangsu in 1980 (a) and 2008 (b)

鉴于矢量数据库可以准确表征 SOC 预测结果的空间分布特征, 本研究使用 1:50 000 土壤矢量数据库实现栅格预测结果的输出。1:50 000 土壤矢量数据库是目前中国大区域尺度最为详尽的土壤数据库, 由研究区第二次土壤普查各个县(市)图件资料 GIS 数字化而成。该数据库有旱地图斑数目为 17 024 个, 基本制图单元为土种, 图例单元命名采用中国土壤发生分类系统。

### 1.3 样点数量设计

不同样点子集的创建采用 ArcGIS 10.2 地统计分析 Create Subsets 模块。首先从 2008 年可利用样点数据集中随机剔除 197 个, 剩余 884 个, 保证 1980 年和 2008 年样点总数一致。然后分别从两期的 884 个样点中随机抽取 85% 的样点 (751 个) 作为训练数据集参与实际的空间插值过程, 剩余的 15% (133 个) 作为验证数据集用于插值预测精度的独立验证和不确定性评价。751 个训练数据集按 5% 的递减方式进行随机抽取。通过 SPSS 软件对各样点子集中的 SOC 进行最大值、最小值、标准差 (SD)、均值、变异系数 (CV)、偏度、峰度等描述性统计和正态分布检验。

### 1.4 空间预测方法及其精度评价

本研究采用当前土壤学使用最多和效果最稳健的普通克里金 (Ordinary Kriging, OK) 方法实现苏北旱地 SOC 含量的空间预测。空间预测结果输出为 30 m×30 m 的栅格, 然后将每个土壤图斑所包含栅格像元 (即 SOC 含量预测值) 平均值提取至矢量土壤图图斑。

本研究选择常用的均方根误差 (Root Mean Square Error, RMSE) 和相关系数 ( $r$ ) 评价不同样点数量下 SOC 空间预测的准确性。由于 1980 年和 2008 年 SOC 数据方差存在一定的波动性, RMSE

和  $r$  不能直接用于不同时期预测精度比较。为了消除该影响, 本研究也采用标准均方根误差 (Normalized Root Mean Square Error, NRMSE) 比较不同时期插值预测精度, 其中:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{mi} - X_{pi})^2} \quad (1)$$

$$NRMSE = RMSE / \frac{1}{n} \sum_{i=1}^n X_{mi} \quad (2)$$

式中,  $n$  为验证样点数量 (本研究  $n=133$ ),  $X_{mi}$  和  $X_{pi}$  分别表示第  $i$  个验证样点中 SOC 含量的实测值和预测值。RMSE 或 NRMSE 越小,  $r$  越大, 预测精度越高。

### 1.5 经典统计学确定合理采样数目的计算公式

本研究所获取的最优采样数量与经典统计学确定的合理采样数目进行对比, 以分析其合理性。在目前大多数研究中, Cochran<sup>[14]</sup>针对与区域范围内纯随机采样而构造的计算公式通常被经典统计学用于最佳采样数量的确定:

$$n = (t \times S_{td})^2 / d^2 \quad (3)$$

式中,  $n$  为所需要的最佳采样数量,  $t$  为与显著水平对应的  $t$  式分布值,  $S_{td}$  为样本标准差,  $d$  为样本平均值与允许误差的乘积。

## 2 结果与讨论

### 2.1 不同样点数量下 1980 年和 2008 年 SOC 含量统计学特征及空间变异结构

基于训练样点总集 (751 个样点) 的 SOC 含量

统计结果表明(表1),1980年研究区SOC含量均值为5.70 g·kg<sup>-1</sup>,而2008年含量均值明显提高,达到10.41 g·kg<sup>-1</sup>。两个时期其余19组样点子集SOC含量均值、变异系数与训练样点总集统计结果相似,表明随机抽取的不同样点数量均具有较好代表性,降低采样数量对研究区SOC基本统计信息影响较小,这与程道全等<sup>[15]</sup>和李凯等<sup>[16]</sup>的研究结果一致。进一步的Kolmogorov-Smirnov检验表明,1980年所有样点子集SOC含量全部符合正态分布,2008年样点子集SOC含量经对数转换后,也均符合正态分布,满足地统计插值的要求。

表征SOC空间变异的指示参数(块金值、基台值、结构性比例和变程)在不同时期或样点数量下

具有较大差异,表明SOC空间变异结构随着样点数量以及时间的推移发生明显变化。1980年研究区SOC的结构性比例较小(3.43%~14.19%),其变异性受结构性因素影响在空间上强烈自相关;而2008年SOC结构性比例在14.29%~50.00%之间,绝大部分样点子集都超过了25%,土壤变异的空间相关性减弱为中等水平,这是因为1980—2008年研究区的耕作、施肥等土壤管理措施经历了较大改变,对SOC空间变异结构造成了一定影响。

## 2.2 不同样点数量下1980年和2008年SOC预测结果的空间分布特征

随着样点数量的减少,苏北旱地SOC的空间分布特征发生明显改变(图3和图4)。训练样点总集

表1 不同样点数量下研究区SOC含量(g·kg<sup>-1</sup>)描述性统计结果及半方差模型拟合参数

Table 1 Descriptive statistical and semivariance models fitting parameters of SOC contents (g·kg<sup>-1</sup>) in the study area relative to number of sampling sites

年份 Year	样点数量 Number of sampling sites	均值 Mean/(g·kg <sup>-1</sup> )	标准差 Standard deviation	变异系数 Coefficient of variation/ (%)	块金值 Nugget	基台值 Sill	结构性比例 Structural proportion/%	变程 Range/km	拟合优度 R2 Goodness of fit
1980年 Year 1980	751	5.70	2.35	41.23	0.73	5.57	13.11	15.60	0.57
	713	5.70	2.34	41.05	0.72	5.54	13.00	15.00	0.56
	676	5.68	2.34	41.20	0.74	5.55	13.33	13.80	0.54
	638	5.68	2.33	41.02	0.72	5.48	13.14	14.40	0.53
	601	5.78	2.36	40.83	0.74	5.62	13.17	15.30	0.52
	563	5.70	2.39	41.93	0.75	5.78	12.98	16.20	0.52
	526	5.75	2.36	41.04	0.72	5.62	12.81	17.40	0.56
	488	5.67	2.33	41.09	0.74	5.52	13.41	13.80	0.55
	451	5.82	2.40	41.24	0.75	5.80	12.93	15.90	0.61
	413	5.65	2.35	41.59	0.75	5.63	13.32	17.10	0.58
	376	5.64	2.35	41.67	0.75	5.63	13.32	14.40	0.56
	338	5.69	2.32	40.77	0.76	5.51	13.79	17.40	0.51
	300	5.74	2.38	41.46	0.68	5.69	11.95	16.50	0.55
	263	5.89	2.39	40.58	0.79	5.83	13.55	19.50	0.52
	225	5.89	2.22	37.69	0.77	5.63	13.68	17.40	0.53
	188	5.76	2.44	42.36	0.76	5.96	12.75	15.90	0.46
	150	5.98	2.42	40.47	0.84	5.92	14.19	8.40	0.04
	113	5.76	2.12	36.81	0.15	4.37	3.43	6.80	0.08
	75	5.89	2.68	45.50	0.41	7.24	5.66	13.50	0.09
	38	5.36	2.06	38.43	0.16	4.15	3.86	17.90	0.08

续表

年份 Year	样点数量 Number of sampling sites	均值 Mean/ ( g·kg <sup>-1</sup> )	标准差 Standard deviation	变异系数 Coefficient of variation/ (%)	块金值 Nugget	基台值 Sill	结构性比例 Structural proportion/%		变程 Range/km	拟合优度 R2 Goodness of fit
							结构性比例 Structural proportion/%	变程 Range/km		
	751	10.41	2.66	25.55	0.03	0.07	42.86	100.20	0.71	
	713	10.4	2.67	25.67	0.03	0.07	42.86	104.70	0.72	
	676	10.41	2.66	25.55	0.03	0.07	42.86	108.30	0.72	
	638	10.38	2.65	25.53	0.03	0.07	42.86	110.10	0.75	
	601	10.34	2.69	26.02	0.03	0.07	42.86	105.30	0.75	
	563	10.34	2.64	25.53	0.03	0.07	42.86	99.30	0.71	
	526	10.41	2.63	25.26	0.03	0.07	42.86	104.70	0.71	
	488	10.38	2.65	25.53	0.02	0.07	28.57	81.30	0.70	
	451	10.32	2.61	25.29	0.03	0.07	42.86	111.60	0.67	
2008 年	413	10.40	2.63	25.29	0.03	0.07	42.86	106.80	0.67	
Year 2008	376	10.44	2.63	25.19	0.02	0.07	28.57	75.60	0.61	
	338	10.47	2.73	26.07	0.02	0.07	28.57	59.40	0.60	
	300	10.46	2.65	25.33	0.02	0.06	33.33	25.81	0.57	
	263	10.50	2.64	25.14	0.01	0.06	16.67	30.00	0.57	
	225	10.30	2.52	24.47	0.03	0.06	50.00	30.31	0.53	
	188	10.04	2.45	24.40	0.03	0.06	50.00	23.21	0.54	
	150	10.35	2.85	27.54	0.01	0.07	14.29	45.30	0.48	
	113	10.61	2.73	25.73	0.03	0.08	37.50	187.80	0.47	
	75	10.63	2.74	25.78	0.03	0.07	42.86	119.40	0.48	
	38	10.50	2.55	24.29	0.01	0.07	14.29	18.30	0.03	

下, SOC 空间分布既能反映整体格局又能体现变异细节, 但随着土壤样点数量的降低, 表征 SOC 局部变异的能力减弱, 并逐渐改变整体空间分布格局, 其中 SOC 含量高值和低值区域面积变化最为剧烈。造成不同数据集极值区域估算结果明显差异的原因可能是因为两端的极值具有较高空间异质性, 样点数量减少导致表征土壤空间异质性的能力减弱。一方面, 随着样点数量的降低, 样点数据集的 SOC 含量最大值呈现减小趋势, 1980 年和 2008 年的变幅分别由 751 个样点的  $13.17 \text{ g}\cdot\text{kg}^{-1}$  减小至 38 个样点的  $11.37 \text{ g}\cdot\text{kg}^{-1}$ ; 而最小值呈增大趋势, 1980 年和 2008 年的变幅分别由 751 个样点的  $0.70 \text{ g}\cdot\text{kg}^{-1}$  减小至 38 个样点的  $1.68 \text{ g}\cdot\text{kg}^{-1}$ , 表明样点数量减少会使得所探测到的 SOC 含量范围降低, 进而影响其表征研究区的空间完整性。另一方面, 巫振富等<sup>[17]</sup>研究

指出, 当最能凸显土壤空间分布格局的高、低值“关键区域”样点数量足够多时, 才可保证预测结果最大程度地反映研究区土壤属性的实际空间分布特征, 但样点数量的减少会导致“关键区域”能否满足样点需求成为一种不确定性事件。

以训样样点总集空间预测结果的高、低值区为基准, 划定研究区 SOC 空间分布的“关键区域”。经统计, 751 个训练样点总集下 1980 年和 2008 年 SOC 高值区的样点数量分别为 147 和 40 个, 低值区分别为 130 和 372 个。当样点数量依次减少为 638 (85%)、563 (75%)、451 (65%) 个时, 位于“关键区域”的土壤样点数量仍然较多, 其中 1980 年 SOC 高值区样点数量分别为 121、116 和 97 个, 低值区样点数量分别为 111、100 和 81 个; 2008 年 SOC 高值区样点数量分别为 33、28 和 29 个, 低值区样

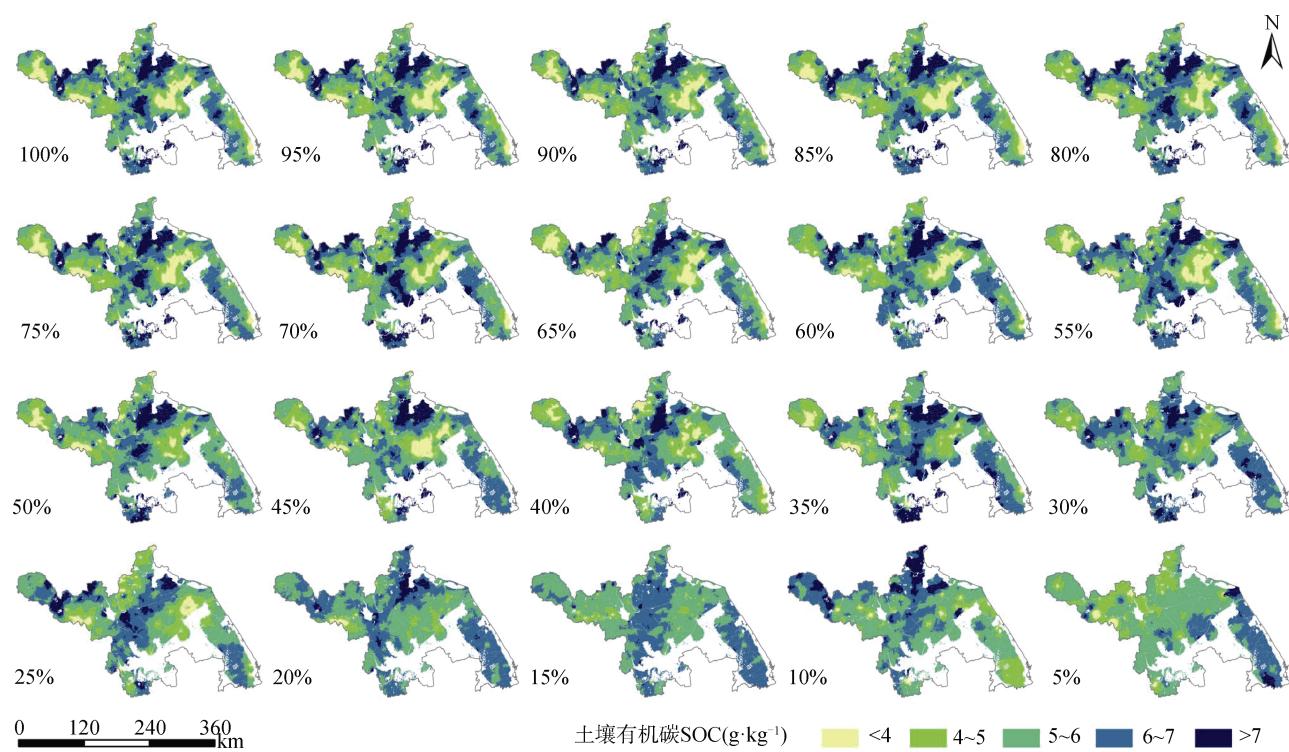


图3 基于不同样点数量 (1980年) 的土壤有机碳含量分布图  
Fig. 3 Prediction maps of SOC contents in 1980 relative to number of sampling sites

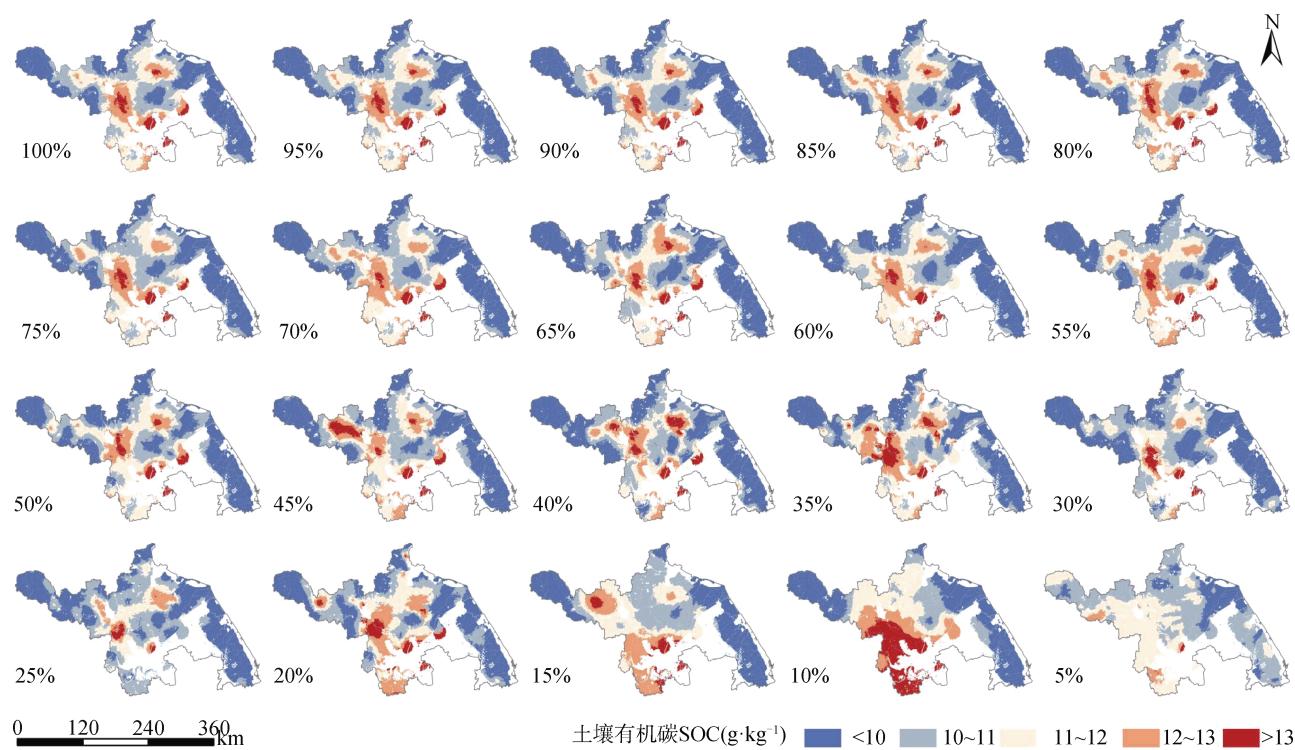


图4 基于不同样点数量 (2008年) 的土壤有机碳含量分布图  
Fig. 4 Prediction maps of SOC contents in 2008 relative to number of sampling sites

点数量分别为320、287和242个, 尽管局部的一些细节信息逐渐被隐藏, 但此时的SOC的整体分布格

局和趋势仍然比较清晰, 样点数量降低并未导致预测制图结果差异明显。然而, 当样点数量进一步减

少至 113 (15%)、38 (5%) 个时, 位于 SOC 高值区的数量 1980 年分别仅有 28 和 3 个, 2008 年分别仅有 8 和 2 个; 位于低值区的数量 1980 年分别仅有 12 和 7 个, 2008 年分别仅有 48 和 17 个, 样点数量无法满足“关键区域”的制图需求, 导致 SOC 在空间分布上的细节信息基本被平滑, 呈现出单一的分布格局, 不能很好地反映苏北旱地 SOC 的真实分布状况, 特别是 5% 样点数量下空间结构出现失真畸变, 表现的空间分布格局与训练样点总集相异。

从图 2 也可以看出, 虽然 2008 年的采样点空间分布较不均匀, 但该时期各个样点集的空间分布差异整体较小, 这是由于其分布于“关键区域”的样点数较多, 能更稳定地反映“关键区域”空间分布的真实情况, 随着训练样点数量的降低, 样点集统计特征的变异系数变化程度也明显小于 1980 年。这表明空间插值精度受采样数量和空间分布的共同影响, 位于“关键区域”的样点数量越多, 其估算结果越稳定。杨琳等<sup>[3]</sup>也认为, 与单纯地增加样点数量相比, 基于样点代表性的土壤采样在土壤空间制图成本和准确性上具有更为明显的优势。因此, 在 SOC 高、低值等“关键区域”设置足够的样点数量, 从而提高采样点所反映的空间异质性, 是提高土壤空间预测效果的重要手段之一。

### 2.3 不同样点数量下 1980 年和 2008 年 SOC 空间预测精度

1980 年和 2008 年研究区 SOC 含量空间预测精度随样点数量变化如图 5 所示。在 1980 年预测制图结果中,  $r$  和 RMSE 随样点数量变化的特征可划分为: 100%~75%、70%~20% 和 15%~5% 三个阶段。在第一阶段,  $r$  和 RMSE 变化幅度较小, 预测精度并未随着采样点数量的减少而产生明显差异, 说明当样点数量  $\geq 75\%$  时, 通过 OK 插值手段获得的 SOC 含量预测结果精度较高且已经达到相对稳定水平, 研究区的最佳采样数目在 75% (563 个) 左右, 这可能是因为样点数量增加至一定程度以后, 空间上的聚集度较高, 导致信息达到一定的饱和水平。在这种情况下, 即使采样点数量增加, 也无法提供进一步的信息。在第二阶段, 随着样点数量的减少, SOC 含量的空间预测精度逐渐降低,  $r$  和 RMSE 变化幅度分别介于 0.39~0.56 和 2.09~2.35 g·kg<sup>-1</sup> 之间, 原因在于预测精度受到样点代表性和密度显著影响。根据极值区域的样点统计, 第二阶段的数量

由 216 个减少至 63 个, SOC 空间数据的代表性显著降低。此外, ArcGIS 10.2 的空间统计分析表明, 第二阶段样点之间的平均最近邻距离为 4~9 km, 明显大于前一阶段。在第三阶段,  $r$  和 RMSE 随着样点数量的减少急剧变化, 预测精度较第二阶段降低得更快。半方差模型拟合结果显示, 15%~5% 样点数量下的拟合优度仅在 0.08 左右, SOC 空间变异结构的推断存在较大误差, 这可能是预测精度急剧降低的主要原因。然而, 与 1980 年相比, 使用 20 组相同样点梯度的训练子集数据对 2008 年研究区 SOC 含量预测时, 预测精度随样点数量变化特征存在一定差异。2008 年预测精度的变化趋势同样表现为先缓慢再剧烈下降, 但阶段划分有所不同, 主要分为: 100%~70%、70%~20% 和 15%~5% 三个。在每个阶段, 预测精度特征与 1980 年也不尽相同。第一阶段, 预测精度并未随着样点数量的降低表现出显著变化, 即 70% 的样点数时预测效果能达到相对稳定水平。可见, 2008 年研究区的最佳采样数目在 526 个左右, 较 1980 年少 5%, 说明随着样点数量的增加, 其预测精度更快趋于稳定。第二阶段, 预测精度随样点数量降低呈现出明显的下降趋势,  $r$  和 RMSE 变化幅度分别处于 0.42~0.63 和 2.11~2.56 g·kg<sup>-1</sup> 之间。在该阶段内, 2008 年各个样点梯度间的预测精度差距明显大于 1980 年, 表明样点数量的增加对其预测效果的提高更为显著。第三阶段, 随着样点数量减少, 样点梯度间预测精度差距进一步增大, 即对样点数量的敏感性增强。但是, 与 1980 年不同的是, 2008 年预测精度在 15% 样点数量下出现了一定回升, 原因可能与样点数量稀少而引起的预测精度不确定性有关。研究表明, 预测精度受到样点数量和空间分布的共同影响, 随着样点数量的减少, 空间分布位置对于预测效果的影响逐渐增大, 甚至可以取代样点数量成为预测效果的主导性因素, 但随机抽样忽略了样点的空间布局, 使得样点是否对研究区 SOC 空间分布具有较好的全局代表性成为一种概率事件<sup>[17]</sup>。

从图 5 也可以看出, 1980 年和 2008 年苏北旱地 SOC 含量空间预测精度随样点数量变化表现出不同的响应特征, 每个时期所需的样点数量有所差异。这主要是由于不同时期的环境、气候以及农田管理措施 (施肥、灌溉和耕作等) 不同, 导致有机碳在土壤中的含量及其空间结构不同而产生响应特

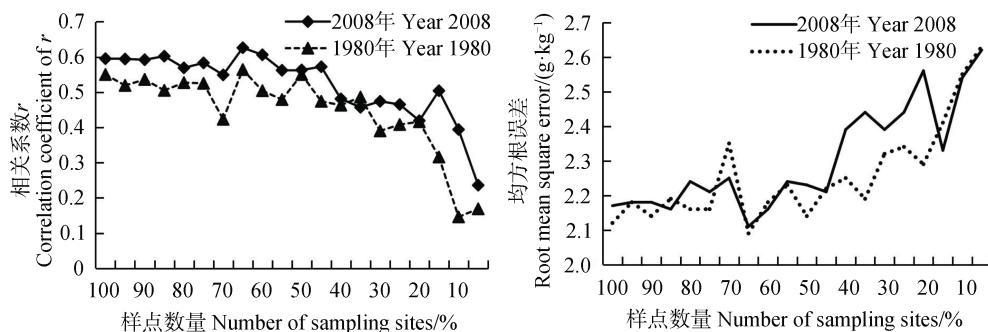


图 5 1980 年和 2008 年不同土壤样点数量的土壤有机碳含量预测精度

Fig. 5 Accuracy of the prediction of soil organic carbon content in 1980 and in 2008 relative to number of sampling sites

征差异性。由于 1980 年研究区 SOC 空间自相关变程较小、局部变异较多, 不仅需要更多的样点来推断空间分布, 并且其预测精度的提高速率也较为缓慢和不稳定; 相比而言, 2008 年研究区 SOC 空间自相关变程较大、局部变异较少, 所以预测精度对于采样数量的敏感性较强, 空间信息达到饱和状态所需样点数量也相对较少。从 NRMSE 来看(图 2), 不同时期 SOC 含量的预测效果差距也十分明显, 20 组样点数量下 1980 年和 2008 年的 NRMSE 分别为  $0.34\sim0.43$  和  $0.20\sim0.25 \text{ g}\cdot\text{kg}^{-1}$ , 在样点数量相同的情况下 2008 年的空间预测精度优于 1980 年, 主要的原因有以下 3 点: 一是空间变异性对 OK 插值精度具有负面影响, 该方法多适用于土壤属性变异较弱的区域环境<sup>[18]</sup>; 二是由于不同时期样点的空间布局差异所致; 三是 2008 年的土壤空间样点尽管分布不均, 但位于关键区域的样点数量较多, 其半方差模型拟合参数明显优于 1980 年, 采样点具有更好的代表性。由此可见, 研究区最优采样点数量及其预测精度不是固定的, 而是由每个阶段土壤属性的空间变异特征以及样点的分布位置、空间布局共同决定。

根据经典统计学确定合理采样数目的计算公式(表 2), 在 95% 的置信区间和 5% 的误差条件下, 1980 年和 2008 年分别只需要 261 和 100 个采样点即可满足空间预测需求, 而在 90% 的置信区间下, 分别仅需 20 和 8 个采样点即可满足 15% 误差条件下的空间预测需求。本研究显示, 1980 年和 2008 年的最优采样数量应分别为 563 和 526 个, 明显多于利用 Cochran 计算公式得到的最佳采样数量。Cochran 在描述土壤元素空间变异时假定研究区土壤元素的含量变化是随机的, 样本之间服从同一已知的概率分布并且相互独立, 认为样品均值是反映研究区土壤特性的最优方式, 决定土壤最佳采样数量的影响因

素主要包括土壤样本的变异系数、允许误差和可接受的置信区间。然而, 地统计研究表明土壤特性的变化并非是完全随机, 其在空间上具有一定的结构性和关联性, Cochran 计算公式虽然能够显著减少采样数量, 但只能概括 SOC 含量变化的宏观趋势而不能反映其局部变异特征, 如果盲目采用 Cochran 公式确定采样数量, 可能会造成超过预期的预测误差和不确定性。本研究采用随机抽样与地统计学的克里金插值相结合的方法确定合理采样数目, 充分考虑了 SOC 的空间变异和结构特征, 较 Cochran 将研究区作为一个均质的整体来确定最优采样数目的方式更具科学性。

表 2 基于 Cochran 采样数量公式获取的苏北旱地土壤有机碳最佳采样数

Table 2 Optimal number of sampling sites for prediction of soil organic carbon in the upland area in North Jiangsu based on the Cochran sampling quantity formula

置信区间 Confidence interval/%	年份 Year	允许误差 Permissible error/%		
		5	10	15
95	1980	261	65	29
	2008	100	25	11
90	1980	184	46	20
	2008	71	18	8

### 3 结 论

本研究通过随机抽样方式设置 20 组不同数量的样点数据集, 结合 1:50 000 高精度土壤矢量数据库明确样点数量对 1980 年和 2008 年 SOC 普通克里金插值预测精度的影响, 结果表明, 1980 年苏北

旱地 SOC 预测值与实测值的相关系数和均方根误差分别在 0.15~0.56 和 2.09~2.63 g·kg<sup>-1</sup> 之间, 样点数量约 563 个时可达到相对可靠的预测水平; 2008 年 SOC 预测值与实测值的相关系数和均方根误差分别在 0.24~0.60 和 2.11~2.62 g·kg<sup>-1</sup> 之间, 空间信息达到稳定预测水平所需样点数量约 526 个。这表明相同区域范围内的最优采样数量及其预测精度不是固定的, 是由每个阶段内土壤属性的空间变异特征以及样点的分布位置、空间布局共同决定。因此, 未来的土壤采样设计需要根据研究区域土壤属性的时空变异特点制定合理的样点数量以及布局方式, 以提高土壤采样的科学性和高效性。此外, 有研究表明系统抽样能够较空间随机抽样更好地反映土壤属性的空间变异, 未来的研究还应进一步考虑系统抽样方法对土壤属性变量空间预测结果的影响<sup>[19-20]</sup>。

## 参考文献 (References)

- [ 1 ] Yao X, Yu K Y, Deng Y B, et al. Spatial distribution of soil organic carbon stocks in Masson pine (*Pinus massoniana*) forests in subtropical China[J]. *Catena*, 2019, 178: 189—198.
- [ 2 ] Huang Y, Sun W J, Zhang W, et al. Changes in soil organic carbon of terrestrial ecosystems in China: A mini-review[J]. *Science China Life Sciences*, 2010, 53 ( 7 ): 766—775.
- [ 3 ] Yang L, Zhu A X, Qin C Z, et al. A soil sampling method based on representativeness grade of sampling points[J]. *Acta Pedologica Sinica*, 2011, 48 ( 5 ): 938—946. [杨琳, 朱阿兴, 秦承志, 等. 一种基于样点代表性等级的土壤采样设计方法[J]. 土壤学报, 2011, 48 ( 5 ): 938—946.]
- [ 4 ] Long J, Liu Y L, Xing S H, et al. Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography[J]. *Ecological Indicators*, 2018, 93: 562—571.
- [ 5 ] Vašátk R, Heuvelink G B M, Borůvka L. Sampling design optimization for multivariate soil mapping[J]. *Geoderma*, 2010, 155 ( 3/4 ): 147—153.
- [ 6 ] Kerry R, Oliver M A. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood[J]. *Geoderma*, 2007, 140 ( 4 ): 383—396.
- [ 7 ] Hai N, Zhao Y C, Tian K, et al. Effect of number of sampling sites on characterization of spatial variability of soil organic matter[J]. *Acta Pedologica Sinica*, 2015, 52 ( 4 ): 783—791. [海南, 赵永存, 田康, 等. 不同样点数量对土壤有机质空间变异表达的影响[J]. 土壤学报, 2015, 52 ( 4 ): 783—791.]
- [ 8 ] Shi Z, Li Y. Application of geostatistics in soil science[J]. Beijing: China Agriculture Press, 2006. [史舟, 李艳. 地统计学在土壤学中的应用[J]. 北京: 中国农业出版社, 2006.]
- [ 9 ] Wang G X, Zhang L M, Li X D, et al. Study of soil organic carbon sequestration rate and potential of upland in northern Jiangsu Province based on high-resolution soil database[J]. *Ecology and Environmental Sciences*, 2016, 25 ( 3 ): 422—431. [王光翔, 张黎明, 李晓迪, 等. 基于高精度土壤数据库的苏北旱地固碳速率和潜力研究[J]. 生态环境学报, 2016, 25 ( 3 ): 422—431.]
- [ 10 ] Zhang L M, Zhuang Q L, Li X D, et al. Carbon sequestration in the uplands of Eastern China: An analysis with high-resolution model simulations[J]. *Soil and Tillage Research*, 2016, 158: 165—176.
- [ 11 ] Zhang L M, Li J J, Yu D S, et al. Map scale effects on soil total phosphorus storage for uplands of China[J]. *Ecology and Environmental Sciences*, 2011, 20( 11 ): 1626—1633. [张黎明, 李加加, 于东升, 等. 不同制图比例尺土壤数据库对旱地磷储量估算的影响[J]. 生态环境学报, 2011, 20 ( 11 ): 1626—1633.]
- [ 12 ] Zhang L M, Liu Y L, Li X D, et al. Effects of soil map scales on simulating soil organic carbon changes of upland soils in Eastern China[J]. *Geoderma*, 2018, 312: 159—169.
- [ 13 ] Yin L X. Characteristics of spatial variability and effect of various spatial interpolation methods for soil chemistry properties of cultivated land in Tong-an County[D]. Fuzhou : College of Resources and Environment, Fujian Agriculture and Forestry University, 2006. [尹兰香. 同安区耕地土壤化学性质空间变异特征及插值模型效果的研究[D]. 福州: 福建农林大学资源与环境学院, 2006.]
- [ 14 ] Cochran W G. *Sampling Techniques*.3rd ed. New York: John Wiley and Sons, 1977.
- [ 15 ] Cheng D Q, Wu Z F, Liu X B, et al. Influences of sample density on spatial prediction of soil organic matter content: A case study from Fengqiu County, Henan Province[J]. *Chinese Journal of Soil Science*, 2013, 44 ( 4 ): 844—850. [程道全, 巫振富, 刘晓冰, 等. 样点密度对土壤有机质空间预测结果的影响——以河南封丘县土壤为例[J]. 土壤通报, 2013, 44 ( 4 ): 844—850.]
- [ 16 ] Li K, Zhao H F, Wu K N, et al. Suitable interpolation method and reasonable sampling quantity of Cd pollution index in soil[J]. *Chinese Journal of Soil Science*, 2016, 47 ( 5 ): 1056—1064. [李凯, 赵华甫, 吴克宁, 等. 土壤重金属 Cd 污染指数的适宜插值方法和合理采样数量研究[J]. 土壤通报, 2016, 47 ( 5 ): 1056—1064.]

- [ 17 ] Wu Z F, Zhao Y F, Cheng D Q, et al. Influences of sample size and spatial distribution on accuracy of predictive soil mapping on a county scale[J]. *Acta Pedologica Sinica*, 2019, 56 ( 6 ): 1321—1335. [巫振富, 赵彦锋, 程道全, 等. 样点数量与空间分布对县域尺度土壤属性空间预测效果的影响[J]. 土壤学报, 2019, 56 ( 6 ): 1321—1335.]
- [ 18 ] Ma L F, Xiong H G, Sun D, et al. Research on a spatial optimal interpolation method of soil organic matter under different degrees of disturbance[J]. *Acta Ecologica Sinica*, 2019, 39 ( 19 ): 7153—7160. [马利芳, 熊黑钢, 等. 不同干扰程度下土壤有机质空间最优插值法研究[J]. 生态学报, 2019, 39 ( 19 ): 7153—7160.]
- [ 19 ] Wang J F, Haining R, Cao Z D. 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.
- [ 20 ] Jiang C S, Wang J F, Cao Z D. A review of geo-spatial sampling theory[J]. *Acta Geographica Sinica*, 2009, 64 ( 3 ): 368—380. [姜成晟, 王劲峰, 曹志冬. 地理空间抽样理论研究综述[J]. 地理学报, 2009, 64 ( 3 ): 368—380.]

(责任编辑: 檀满枝)