

DOI: 10.11766/trxb201811210470

# 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\*

巫振富<sup>1</sup> 赵彦锋<sup>2</sup> 程道全<sup>3</sup> 陈杰<sup>2†</sup>

(1 郑州大学公共管理学院, 郑州 450001)

(2 郑州大学农学院, 郑州 450001)

(3 河南省土壤肥料站, 郑州 450002)

**摘要** 明确样点数量和空间分布对土壤属性空间预测的影响, 有助于科学制定土壤采样策略、有效提高土壤空间预测精度。从 5 403 个土壤样点中随机抽取验证数据集以及包含不同样点数量的训练数据子集(每个子集包括五种样点空间分布实例), 在研究区表层土壤有机质含量普通克里格(Ordinary Kriging, OK)和反距离加权(Inverse Distance Weighted, IDW)插值结果的基础上, 分析和探讨样点数量与空间分布对土壤空间预测效果的影响。结果显示, 当样点数量从 5 000 降至 39 个时, OK 和 IDW 插值图的局部变异信息逐渐减少, 基于 20 和 10 个样点的插值图存在失真畸变。当样点数量从 5 000 降至 1 250 个时, OK 插值精度相近( $r$  变幅为 0.55~0.59、RMSE 变幅为 3.03~3.15); 从样点数量减少至 625 个开始, OK 插值精度明显下降, 同一训练子集不同样点空间分布的插值精度分异明显。IDW 插值精度随样点数量与空间分布的变化与 OK 插值相似, 不同的是从 1 875 个样点开始出现插值精度的明显下降和不同空间分布插值精度的明显分异。在插值图发生失真畸变之前, OK 平均插值精度大于 IDW。研究结果表明, 样点数量及空间分布均可在不同程度上影响土壤属性空间预测结果, 当样点数量足够多时, 样点数量和空间分布对预测结果的影响非常有限; 当样点数量减少至一定程度时, 随着样点数量的减少, 空间预测图的局部变异信息逐渐减少, 预测精度逐渐下降, 同时样点空间分布对预测结果的影响开始凸显; 在空间预测结果发生失真畸变之前, 与 OK 相比, IDW 插值精度较低且更早响应样点数量和空间分布的变化。

**关键词** 土壤有机质; 土壤样点; 空间分布; 数字化土壤制图

**中图分类号** S159-3 **文献标识码** A

土壤属性及其空间分布信息是区域耕地地力评价、土壤改良利用、精准农业实践、生态环境保护等一系列科研与生产活动不可或缺的数据基础。随着数据获取、处理和挖掘相关技术的迅猛发展, 基于土壤空间预测的数字土壤制图(Digital soil mapping)已成为当前高效组织和表达土壤信息空间分布的最主要方法<sup>[1-2]</sup>。所谓数字化土壤制图, 就是

基于土壤观测数据与土壤学知识, 应用数学模型推断土壤类型和土壤属性的时间与空间变异, 创建和组织空间土壤信息系统的过程<sup>[2-4]</sup>, 被誉为 20 世纪以来继航空摄影应用于土壤调查制图之后的又一次技术革命<sup>[5]</sup>。

样点是土壤属性观测信息、相关环境因子信息以及两者之间关系信息的载体, 是土壤调查与制图

\* 国家自然科学基金项目(40971128)资助 Supported by the National Natural Science Foundation of China (No. 40971128)

† 通讯作者 Corresponding author, E-mail: jchen@zzu.edu.cn

作者简介: 巫振富(1988—), 男, 广西贺州人, 博士研究生, 主要从事土地资源可持续利用与信息化管理研究, E-mail: wfjt1988@163.com

收稿日期: 2018-11-21; 收到修改稿日期: 2019-01-07; 优先数字出版日期(www.cnki.net): 2019-01-28

的基石。在各种尺度上,土壤采样策略均对土壤空间预测效果与数字化土壤制图输出精度具有决定性影响。样点数量与空间布局是土壤采样策略的核心,需要在全面权衡预测精度、采样成本、作业效率等多方面因素的基础上确定。以往相关研究主要分析和探讨样点数量对土壤空间预测精度的影响,其主要目的是寻求确定最优样点数量的策略与途径。此类研究常用的技术手段是在案例研究中设置若干不同数量的样点系列,通过对比不同样点系列的土壤属性空间预测精度以发现和确定最优采样点数量<sup>[6-10]</sup>。如张晓光等<sup>[11]</sup>以土壤盐分为对象,通过对比分析 12 个不同数量的样点系列的预测结果,发现满足黄河三角洲县域尺度下土壤盐分空间变异表达的土壤样点数量每 1 000 km<sup>2</sup> 不能小于 107 个样点。这类研究的不足之处在于探讨样点数量变化对土壤属性预测结果的影响时,忽略了样点数量变化会导致样点空间分布的变化这一客观事实,因此很难判断上述两种因素对土壤空间预测结果的具体影响。一些学者<sup>[12-13]</sup>则基于土壤属性的统计学特征,利用经验公式直接计算得到一定置信水平条件下的最优采样点数量,但经验公式忽略了土壤属性的空间变异情况,以此得到的最佳样点数量是否可靠有待商榷<sup>[14]</sup>。

另外一类研究则将重点放在揭示和分析样点空间分布在土壤空间预测中的重要影响,其核心目的是优化土壤样点的空间分布格局<sup>[15-20]</sup>。如 van Arkel 和 Kaleita<sup>[21]</sup>在田块尺度上利用 K 均值聚类方法确定了用于估算表层土壤水分的最佳采样点位置。Yang 等<sup>[22]</sup>在土壤样点数相同的条件下,比较了综合等级逐步采样 (integrative hierarchical stepwise sampling, IHS)、分层随机抽样 (stratified random sampling, SRS) 以及条件拉丁超立方采样 (conditioned Latin hypercube sampling, cLHS) 三种采样策略,发现 IHS 在获取代表性样点空间位置方面表现最佳。刘雪琦等<sup>[23]</sup>提出一种基于传统土壤图、依据土壤面积对土壤类型分级,并按照等级之间的比例关系布设训练样点,最后通过案例实践验证了该方法的有效性。样点空间分布格局优化相关研究主要探讨和分析最佳采样点空间位置,或在特定样点数量条件下比较和揭示样点空间分布变化对土壤空间预测精度的影响,而样点数量与空间分布同步变化对土壤预测效果的影响及其机制缺乏深入分析与讨论。

本文以河南省新野县耕地表层土壤有机质含量

空间预测为目标,利用简单随机抽样设置十二个分别包含不同样点数量的样点系列,每个系列设置五个样点空间分布,通过比较基于不同样点系列、不同空间分布的土壤有机质空间预测结果的差异,探讨样点数量和空间分布同步变化对土壤属性空间预测效果与数字化制图精度的影响,籍此为县域尺度土壤采样设计相关研究与实践提供参考。

## 1 材料与方法

### 1.1 研究区概况

河南省新野县位于南阳盆地南部,地理坐标 112°14'44"~112°35'42"E、32°19'30"~32°49'08"N 之间,南北长 52 km,东西宽 22 km,国土面积为 1.06 × 10<sup>3</sup> km<sup>2</sup>,其中耕地面积为 7.13 × 10<sup>4</sup> hm<sup>2</sup>。该县属北亚热带地区,具有明显的大陆性季风气候特征,年平均气温 15.1℃,年平均降水量 721.0 mm,全年平均日照总时数 1 816 h,年平均无霜期 228 d。该县地势自西北向东南倾斜,地势平坦,海拔高度为 77~108 m。如图 1 所示,耕作土壤以发育于古河湖相沉积物的湿润锥形土 (砂姜黑土)、发育于黄土母质和风化残积母质上的湿润淋溶土 (黄褐土) 以及发育于近现代河流冲积物上的冲积新成土 (潮土) 等年轻土壤为主,分别各占耕地面积的 64.7%、19.4% 和 15.9%。新野县农业历史悠久、农业基础雄厚,是国家现代农业示范区、全国无公害蔬菜生产基地示范县、国家商品粮生产基地县以及优质棉基地县。

### 1.2 数据来源及处理

研究区土壤样点信息与表层土壤有机质含量数据 (采用油浴加热重铬酸钾容量法,按照 NY/T 1121.6-2006 测定) 源于 2007—2012 年新野县实施的测土配方施肥补贴项目,由河南省土壤肥料站提供。对照野外工作原始记录与工作底图上标注的样点空间位置,对所有样点进行逐一核对,将底图标注与野外记录信息不吻合且无法更正的样点删除。对样点表层土壤有机质测定值采用  $3\sigma$  准则识别异常值,即采用样本平均值  $\mu$  加减三倍标准差  $\sigma$ ,在区间  $(\mu-3\sigma, \mu+3\sigma)$  以外的数据假定为异常值予以删除<sup>[24]</sup>。经样点空间位置核准和属性异常值检验后,最终得到研究区有效土壤样点 5 403 个 (图 1)。

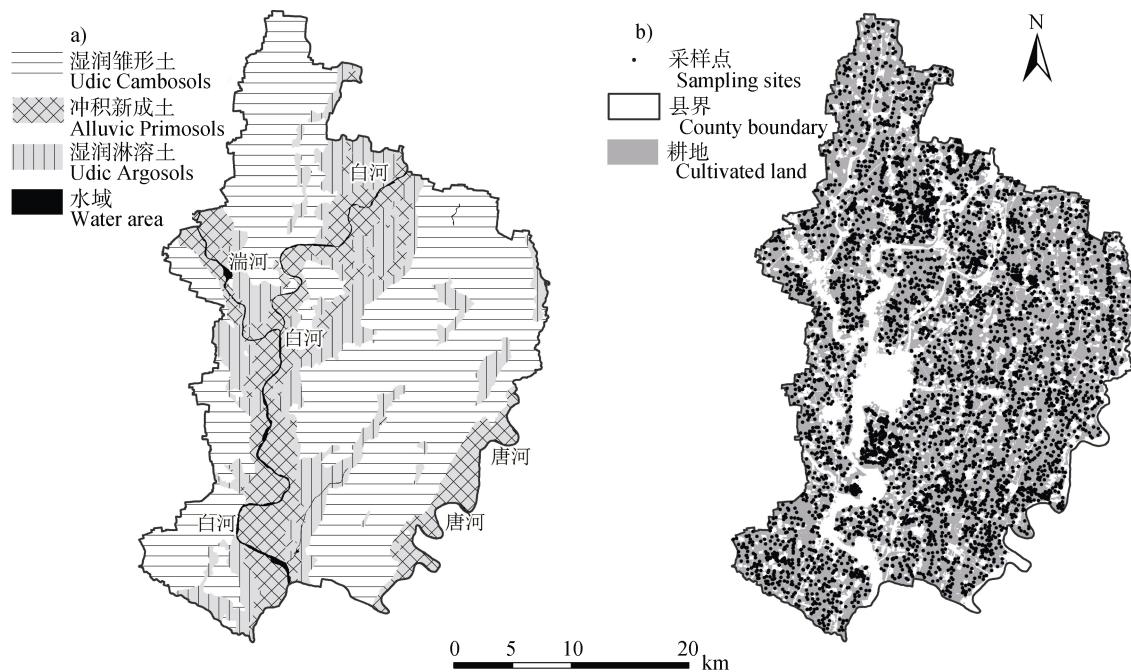


图1 新野县土壤类型 (a)) 和土壤样点空间分布图 (b))

Fig. 1 Soil type (a)) and soil sampling site (b)) distribution maps of Xinye County, Henan Province

### 1.3 研究方法

利用 ArcMap10.1 软件的 Geostatistical Analyst 分析工具, 分别采用普通 Kriging( Ordinary Kriging, OK ) 和反距离加权 ( Inverse Distance Weighted, IDW ) 插值实现研究区表层土壤有机质含量空间预测, 并将预测结果输出为 30 m×30 m 分辨率的栅格图。鉴于 OK 和 IDW 方法均为常用的土壤空间预测方法, 相关计算原理这里不予赘述。插值过程中, 依据样点实际分布情况, 根据每一步长内点对数不小于 20 对、优先拟合近距离经验半方差值等原则, 经过对比筛选确定 OK 插值最优参数<sup>[25-26]</sup>; 根据最小平均误差标准选择 IDW 插值权重计算参数<sup>[27-28]</sup>。

表层土壤有机质含量空间预测基于不同的样点数量与空间分布进行: (1) 首先从研究区 5 403 个有效样点中随机抽取 403 个样点组成验证数据集。 (2) 剩余 5 000 个样点用作训练数据总集, 先从其中随机抽取 2 500 个样点作为一个训练数据子集, 再从剩余的 2 500 个样点中随机抽取 1 250 个样点作为一个训练数据子集, 按上述方法不重复随机抽样形成分别包含 2 500、1 250、625、313、156、78、39、20、10 个样点的 9 个训练数据子集; 然后再从 5 000、2 500 个训练样点中分别随机抽取包含 3 750 和 1 875 个样点的 2 个训练数据子集, 最终形成分

别包含 3 750、2 500、1 875、1 250、625、313、156、78、39、20、10 个样点的 11 个训练数据子集。(3) 重复步骤 (2) 5 次, 获得各训练数据子集的 5 个重复, 代表包含不同样点数量训练子集的 5 种样点空间分布实例, 分别以 I、II、III、IV、V 型分布表示。上述抽样过程利用 ArcMap10.1 软件的 Geostatistical Analyst Tools/Subset Features 工具实现。

用研究区验证样点表层土壤有机质含量预测值与实测值的相关系数 ( $r$ )、预测均方根误差 (Root mean square error, RMSE) 分析和评价基于不同训练样点子集、不同样点空间分布模式获取的研究区土壤有机质含量空间预测效果, 其中:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [Y_i - Y_i^*]^2} \quad (1)$$

式中,  $n$  为研究区验证样点数量 ( $n=403$ ),  $Y_i$  和  $Y_i^*$  分别为第  $i$  个验证样点表层土壤有机质实测值与预测值。

## 2 结果与讨论

### 2.1 研究区表层土壤有机质含量统计学特征

基于全部训练数据 (5 000 个样点) 的表层土壤

有机质含量统计结果表明(表 1),研究区土壤有机质含量变异系数为 22.9%,均值与中值基本一致,偏度和峰度接近 0,K-S 检验呈正态分布( $P>0.05$ )。包含不同数量样点的 11 个训练数据子集统计结果显示(空间分布 I 型样点统计结果见表 1,分布 II、III、IV、V 型统计结果略),当训练子集中样点数量

自 3 750 减少至 20 个时,均值、中值、变异系数与全部训练数据的统计结果相比变化不大,说明较少的土壤样点数量即可满足研究区表层土壤有机质含量描述性整体估算的要求,并且 K-S 正态性检验的  $P$  值均大于 0.05,表明各训练子集表层土壤有机质含量呈正态分布。

表 1 研究区表层土壤有机质含量描述性统计特征

Table 1 Descriptive statistics of organic matter contents in topsoil layer

	样点数 Sizes	均值 Mean/ ( g·kg <sup>-1</sup> )	中值 Median/ ( g·kg <sup>-1</sup> )	变异系数 CV/%	偏度 Skew	峰度 Kurt	K-S 检验 $P$ 值 $P$ values of K-S test
训练数据总集 Training data set	5 000	16.2	16.2	22.9	0.03	-0.22	0.10
	3 750	16.3	16.2	22.8	0.04	-0.23	0.11
	2 500	16.3	16.2	22.4	0.04	-0.25	0.10
	1 875	16.1	16.1	23.1	0.04	-0.26	0.12
训练数据子集 (分布 I 型) Training data subset (distribution pattern I)	1 250	16.1	15.9	23.1	0.02	-0.14	0.12
	625	16.2	16.2	24.0	-0.07	-0.08	0.21
	313	16.3	16.3	22.0	0.17	-0.17	0.92
	156	16.4	16.3	24.3	0.14	-0.71	0.95
	78	16.5	16.2	22.9	0.04	-0.60	0.84
	39	15.9	16.0	21.1	-0.14	-0.85	0.76
	20	16.3	16.5	19.0	-0.06	-1.01	0.08
	10	17.4	18.6	24.6	-0.05	-1.91	0.06

变异函数是地统计学的核心和基本工具,是分析变量空间变异特征的有效手段。在应用 OK 插值实施研究区表层土壤有机质含量预测前,对各训练数据样点表层土壤有机质含量实测值分别进行最优变异函数模型拟合(每个训练数据均以样点空间分布 I 型为例,见表 2),结果表明,对于包含样点数量较少的训练子集,拟合的变异函数模型参数变化较大,当训练子集的样点数量达到 156 个时,表层土壤有机质含量变异函数模型拟合效果较好且更趋稳定,结构性比例为 25%~75%,表现出中等程度的空间相关性<sup>[29]</sup>。

## 2.2 样点数量和空间分布对研究区表层土壤有机质预测结果空间特征的影响

前文关于训练数据的统计学特征分析表明,研究区表层土壤有机质含量满足地统计学插值的数学

条件。以样点空间分布 I 型为例,基于训练数据总集(5 000 个样点)以及不同训练数据子集,应用 OK、IDW 插值获得的研究区表层土壤有机质含量空间预测制图结果见图 2 和图 3。首先,基于训练数据总集的预测结果表明,研究区表层土壤有机质含量较高的区域主要分布于县域西北部、西南部、东部和东北部的湿润锥形土区域,有机质含量较低的区域集中分布在白河、湍河以及唐河两岸的冲积新成土和湿润淋溶土地带。其次,由图 2、图 3 可以看出,表层土壤有机质含量预测结果的空间特征随土壤样点数量变化而发生渐进性变化:当样点数量由 5 000 减至 39 个,表层土壤有机质含量预测值空间分布的局部变异细节信息逐渐减少;进一步减少样点数量至 20、10 个,研究区表层土壤有机质含量空间预测信息出现失真畸变,表现出与样点减少

表 2 表层土壤有机质含量变异函数模型

Table 2 Variogram model of organic matter contents in topsoil layer

样点数 Sizes	模型 Model	步长 Lag/m	变程 Range/m	块金值 Nugget	基台值 Sill	结构性比例 Structural proportion/%		拟合优度 $R^2$ Goodness of fit
5 000	指数 Exponential	1 000	12 000	7.05	14.99	53.0		0.86
3 750	指数 Exponential	1 000	12 000	7.03	14.98	53.1		0.85
2 500	指数 Exponential	1 000	12 000	7.13	14.53	50.9		0.85
1 875	指数 Exponential	1 000	12 000	7.65	15.10	49.3		0.88
1 250	指数 Exponential	1 000	12 000	6.86	15.20	54.9		0.86
625	指数 Exponential	1 000	12 000	7.51	16.38	54.2		0.83
313	球状 Spherical	1 000	10 072	6.72	13.78	51.2		0.79
156	指数 Exponential	1 500	9 981	5.46	14.81	63.1		0.80
78	球状 Spherical	1 500	10 137	7.75	14.95	48.2		0.66
39	球状 Spherical	1 500	14 230	6.34	12.75	50.3		0.29
20	高斯 Gaussian	1 500	1 800	7.33	8.09	9.4		0.20
10	高斯 Gaussian	2 500	28 906	19.24	19.73	2.5		0.00

注: 拟合优度  $R^2$  指变异函数模型与经验半方差值的拟合效果 Note: Fitting goodness  $R^2$  refers to fitting level between variogram model and experimental semi-variance

前不一样的空间分布格局。基于各训练数据子集空间分布Ⅱ、Ⅲ、Ⅳ、Ⅴ型的预测结果与空间分布Ⅰ型呈类似规律, 不再赘述。上述现象表明样点数量对于县域尺度土壤属性空间表征具有重要影响, 这与王志刚等<sup>[30]</sup>、程道全等<sup>[31]</sup>的研究结果一致。

最能凸显研究区表层土壤有机质含量空间分布格局的高、低值区可以视为空间预测与制图表达的“关键区域”, 只有当足够的土壤样点分布于这些“关键区域”, 才可保证预测制图输出结果在最大程度上反映研究区表层土壤有机质含量实际的空间分布格局。根据基于训练数据总集的OK和IDW预测制图结果, 划定研究区表层土壤有机质含量制图表达的“关键区域”(图4)。如图4所示, 当样点数量为313个时, 高值区H1、H2、H3、H4样点数分别为27、3、12、4个, 低值区L1、L2、L3样点数分别为33、15、10个; 当样点数量为20个时, 高值区H2和低值区L3无样点分布, 其余“关键区域”样点数为1或2个; 当样点数量为10个时, 高值区H2、H3和低值区L1、L3无样点分布, 其余“关键区域”样点数为1个。在本研究中, 当样点数量依次减少为

2 500、1 250、625、313、156甚至更少时, 因为能够保证样点在“关键区域”分布, 预测制图输出结果仍然能够表达研究区表层土壤有机质含量空间分布的基本格局, 尽管空间局部变异细节信息的损失随着样点数量的减少而加剧; 当土壤样点数量进一步减少, “关键区域”内样点数量稀少甚至缺失样点时, 预测制图输出结果开始出现失真畸变。

以分别包含2 500、1 250和78个样点的3个训练数据子集为例, 分析和讨论子集内土壤样点空间分布对土壤空间预测结果的影响。图5为上述3个训练子集的不同空间分布类型, 图6、图7则为基于上述3个训练子集的5种样点随机空间分布类型的研究区表层土壤有机质含量OK、IDW插值结果。限于篇幅, 仅列出Ⅰ、Ⅱ、Ⅲ三种空间分布类型及其插值结果。图5表明, 当样点数量为2 500和1 250个时, 同一训练子集的不同空间分布类型在“关键区域”有较多样点; 当样点数量为78个时, 分布Ⅰ、Ⅲ型在“关键区域”存在样点, 但是分布Ⅱ型在西南部的高值区H4无样点分布。对比基于同一训练子集不同样点空间分布的制图结果(图6、图7)可

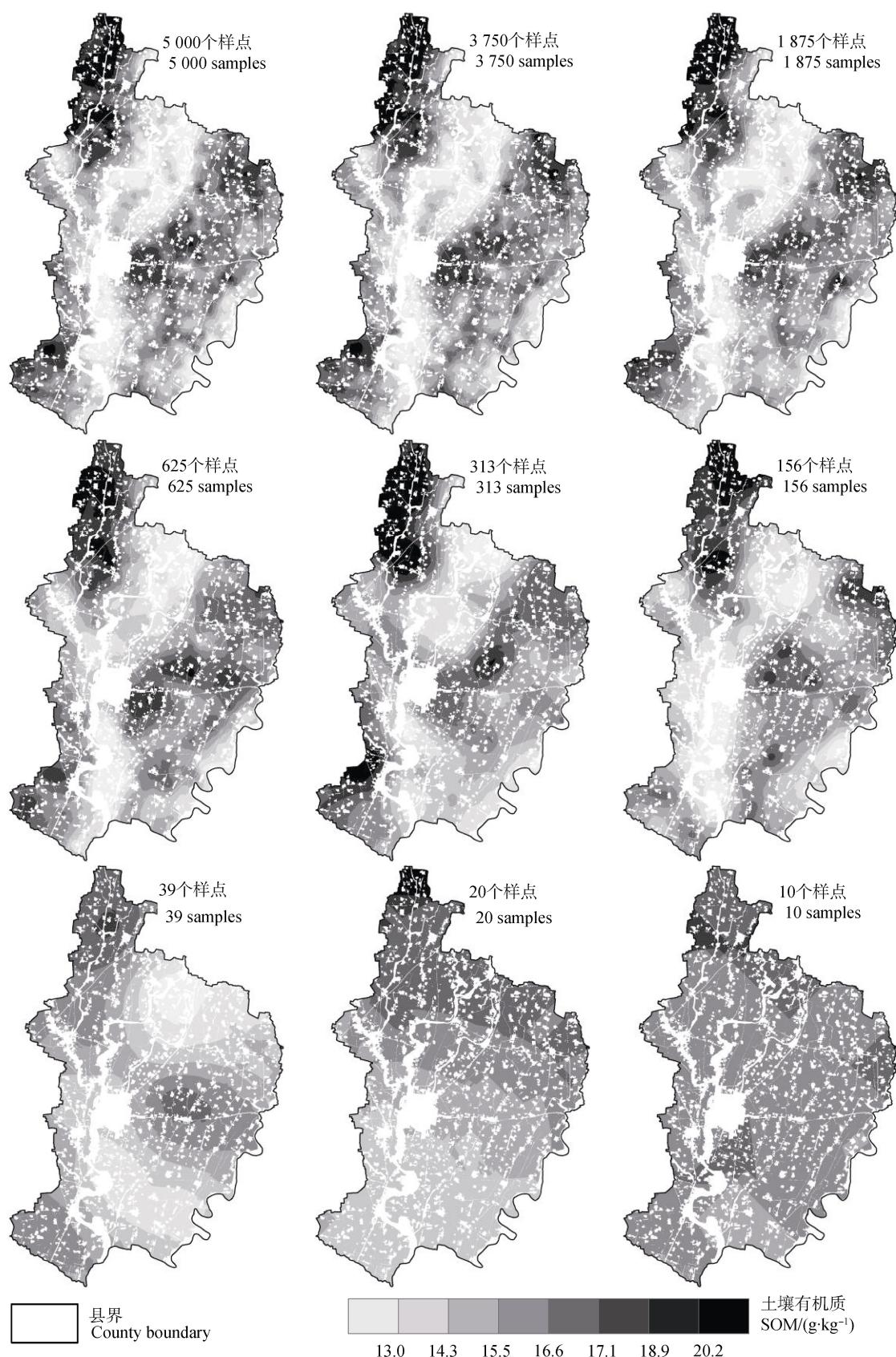


图 2 基于不同训练子集 (样点空间分布 I 型) 的表层土壤有机质含量预测制图 (OK 插值)

Fig. 2 Predictive maps of OM content in topsoil layer based on OK interpolation relative to training subset, Type I in sampling stie distribution pattern

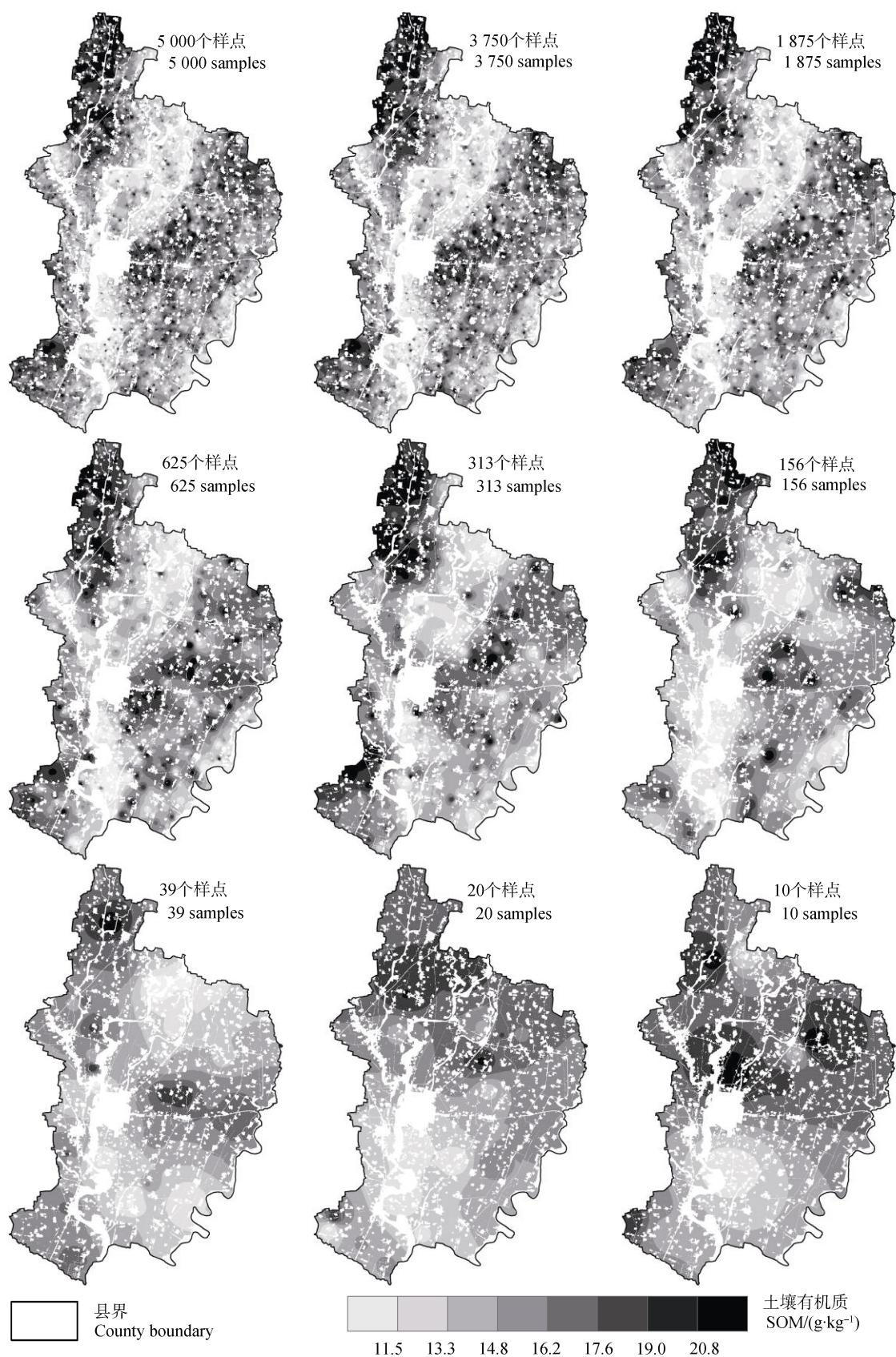
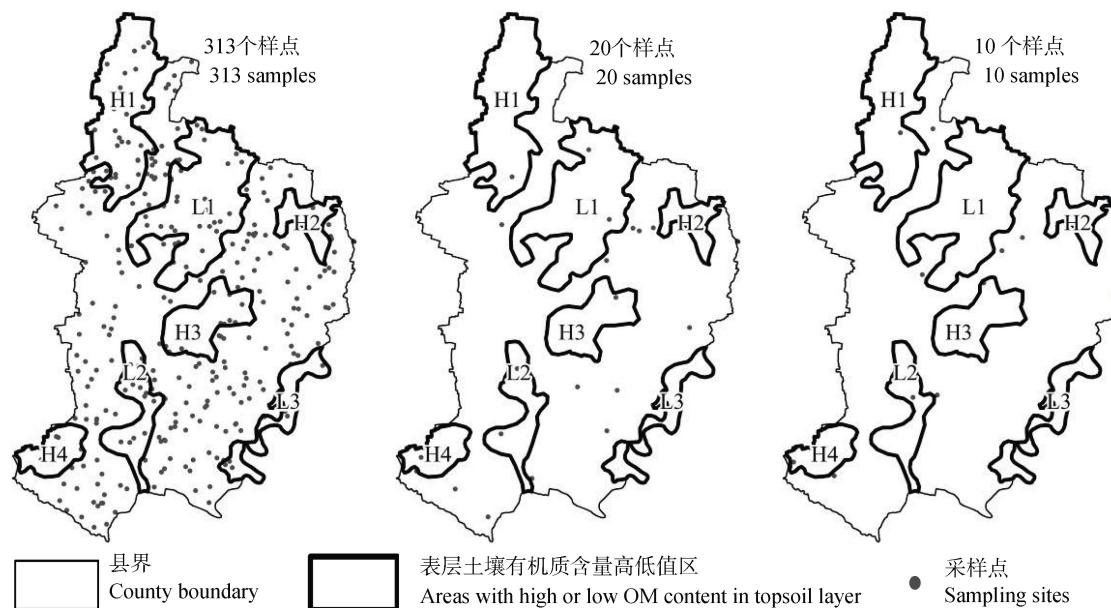


图3 基于不同训练子集(样点空间分布I型)的表层土壤有机质含量预测制图(IDW插值)

Fig. 3 Predictive maps of OM content in topsoil layer based on IDW interpolation relative to training subset, Type I in sampling site distribution pattern



注: H1、H2、H3 和 H4 为表层土壤有机质高值区, L1、L2 和 L3 为表层土壤有机质低值区, 下同 Note: H1, H2, H3 and H4 were areas with high OM content in topsoil layer, and L1, L2 and L3 were areas with low content. The same below

图 4 不同训练子集 (样点空间分布 I 型) 样点在“关键区域”的空间分布

Fig. 4 Spatial distribution of sampling sites in the key area relative to training subset, Type I in sampling site distribution pattern

以发现,当训练子集中土壤样点数量足够多时(如 2 500、1 250 个),样点空间分布的差异不会导致预测制图输出结果的明显差异;而当训练子集中样点数量减少到一定程度时(如 78 个),则样点空间分布的不同可导致制图结果的明显差异。这是因为当训练样点足够多时,随机抽样形成的不同样点空间分布均可确保“关键区域”有样点存在;当训练样点数量较少时,不同样点空间分布能否确保“关键区域”有样点存在就成了一种概率事件,换言之,基于相同样点数量而不同样点空间分布的土壤预测制图结果可能会出现明显差异。Yang 等<sup>[22]</sup>在样点数量相同的条件下通过综合等级逐步采样、分层随机抽样和条件拉丁超立方采样三种方法确定样点空间位置,通过对沙粒含量和土系的预测制图精度发现基于综合等级逐步采样确定的样点分布制图精度优于其他两种方法,证实了在相同样点数量条件下样点空间分布对土壤预测制图精度的影响,这与本研究结果一致。确保在预测变量“关键区域”内布设足够数量的样点,是提高土壤采样代表性和土壤空间预测精度的重要策略和手段之一<sup>[32-33]</sup>。

### 2.3 样点数量和空间分布对研究区表层土壤有机质含量预测精度的影响

图 8 显示了土壤样点数量及其空间分布对表层

土壤有机质含量空间预测精度参数的影响。在应用 OK 插值实施预测制图过程中,当训练子集的样点数量分别为 5 000、3 750、2 500、1 875 和 1 250 个时,研究区验证样点表层土壤有机质含量实测值与预测值之间相关系数  $r$  变幅在 0.55~0.59 之间、预测均方根误差 RMSE 变幅在 3.03~3.15 之间,表现平稳,且各训练子集样点空间分布的变化也未引起预测精度的明显差异。从训练子集样点数量减少至 625 个开始,预测制图的精度随样点数量下降而明显降低,即  $r$  明显减小、RMSE 显著增大;样点分布对预测精度的影响开始凸显,表现为基于同一训练子集不同样点空间分布的预测精度呈现出明显差异;随着样点数量的减少,样点空间分布对预测精度的影响作用可能超过样点数量,如包含 156 个样点的分布 V 型预测精度高于包含 313 个样点的分布 IV 型。李凯等<sup>[34]</sup>研究采样数量对土壤 Cd 污染指数预测精度的影响时得出,基于 2 033、1 830、1 423 和 1 017 个样点的污染指数预测精度变化很小,当样点数量减少至 610 和 203 个时预测精度大幅下降,此结果与本研究得出的样点数量对土壤有机质预测精度的影响规律一致。范曼曼等<sup>[35]</sup>在探讨采样密度对土壤有机质空间预测精度的影响时发现,  $r$  和 RMSE 分别随样点数减少而逐渐减小和增大,但是当样点

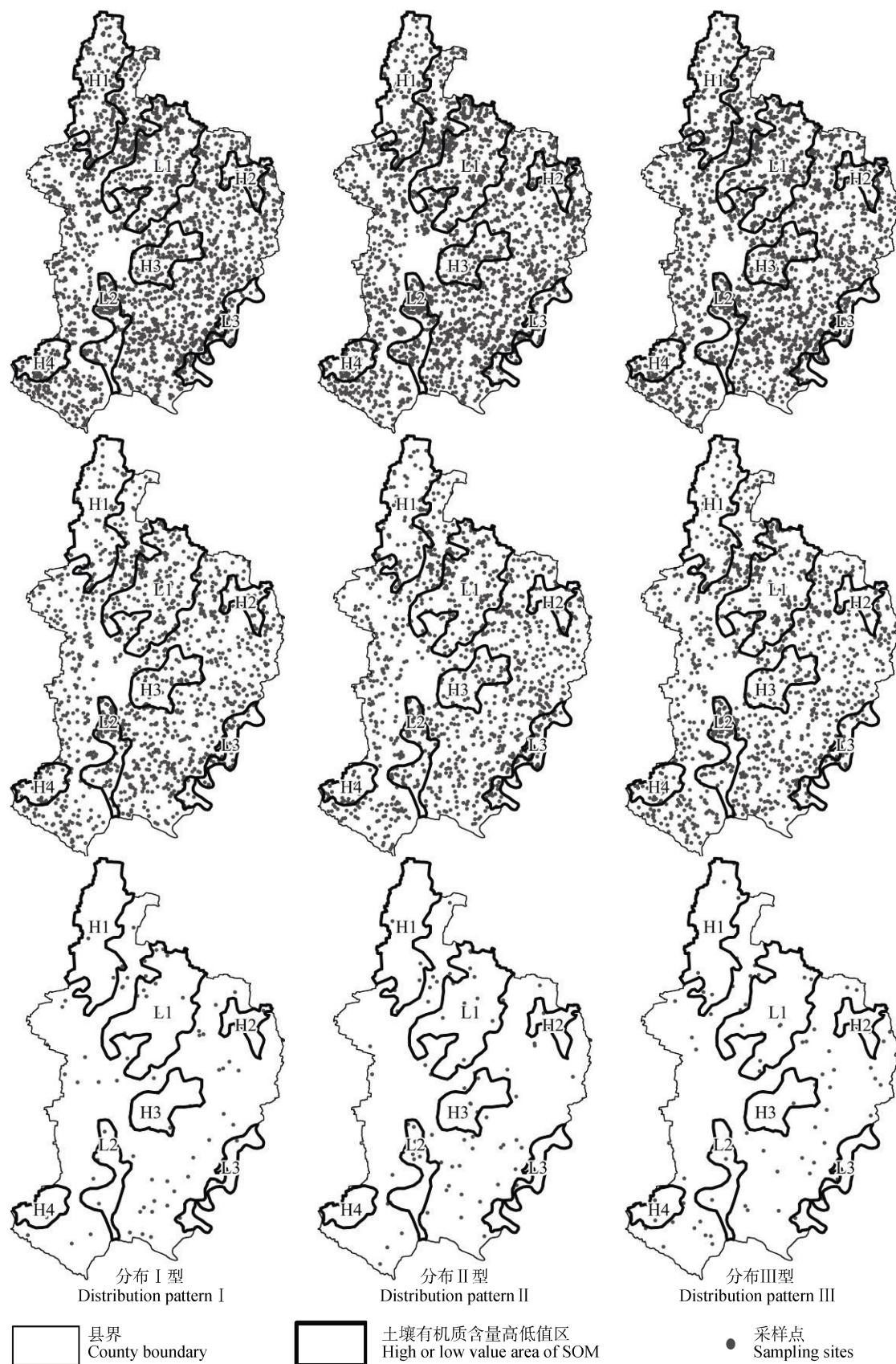


图 5 包含 2 500 (上)、1 250 (中) 和 78 (下) 个样点的训练子集不同空间分布示例

Fig. 5 Examples of sampling site distribution pattern relative to training subset and sample size (2 500 ( top ), 1 250 ( middle ) and 78 ( bottom )

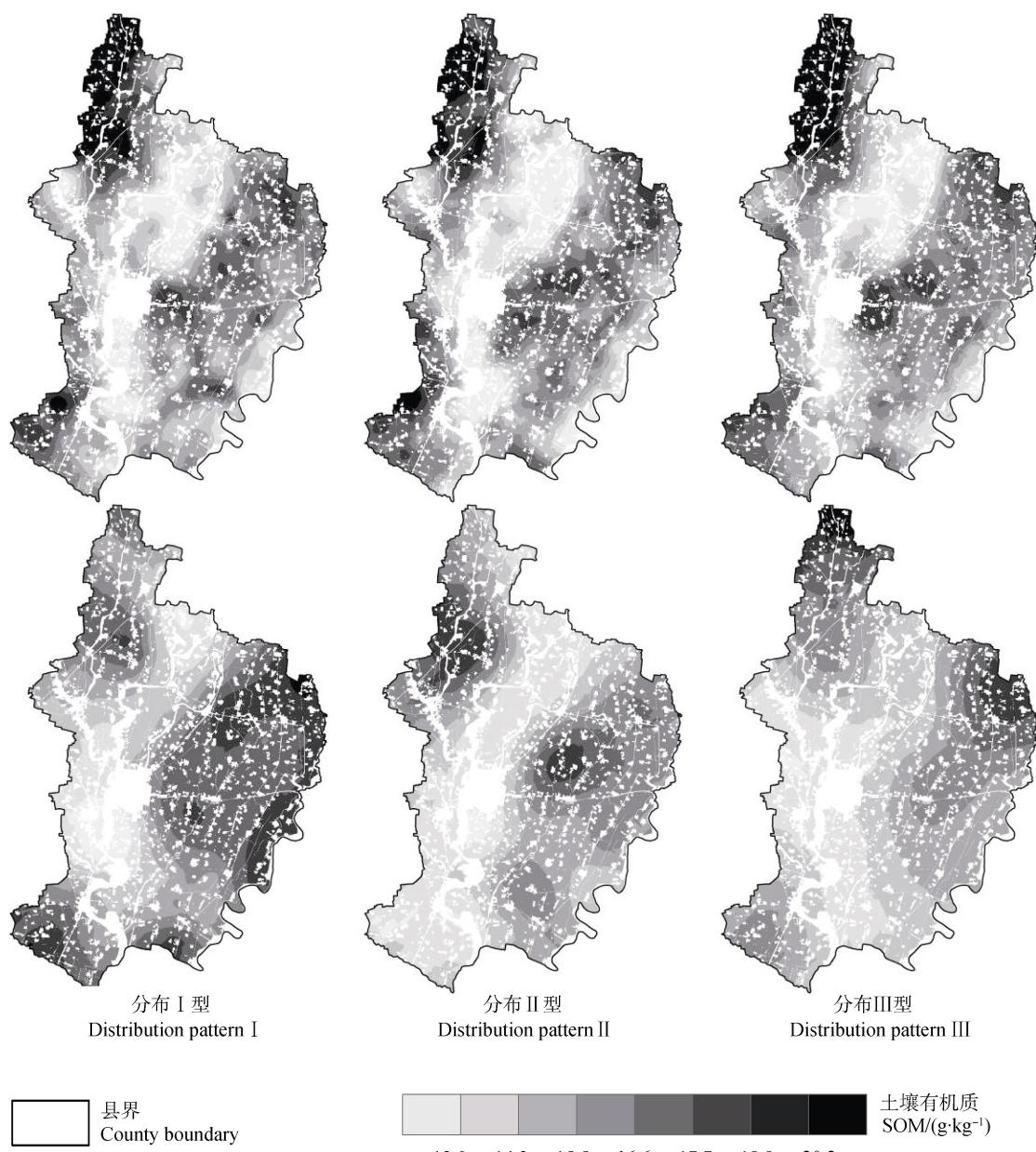


图 6 基于不同样点空间分布的表层土壤有机质含量预测制图 (OK 插值) (上: 1 250 个样点, 下: 78 个样点)

Fig. 6 Predictive maps of OM content in topsoil layers based on OK interpolation relative to sampling site spatial distribution pattern and sample size (top: 1 250 samples and, bottom: 78 samples)

数进一步降至 707 时  $r$  开始增大, RMSE 开始减小; 齐雁冰等<sup>[36]</sup>在研究采样数量与县域土壤有机质 OK 插值精度的关系时也得出类似结果。从本研究结果来看, 出现上述现象的原因可能是当样点数量下降到一定程度之后, 样点空间分布对预测精度的影响作用逐渐凸显, 并超过了样点数量对预测精度的影响。

使用 IDW 插值技术预测研究区表层土壤有机质含量时, 预测结果的精度随样点数量与空间分布的变化表现出与 OK 插值大致相同的趋势。不同的

是, IDW 插值精度在 5 000、3 750、2 500 个样点时表现平稳,  $r$  变幅在 0.49~0.50 之间、RMSE 变幅在 3.31~3.35 之间; 从训练子集样点数量减少至 1 875 个开始, IDW 插值输出结果的精度便随样点数量的减少出现明显下降, 且基于同一训练子集不同样点分布的预测结果之间的精度差异大于 OK 插值, 表明随着样点数量的减少 IDW 插值精度较 OK 插值更早响应样点数量和空间分布的变化, 并且对于空间分布的变化更为敏感; 对比 IDW 和 OK 的平均制图

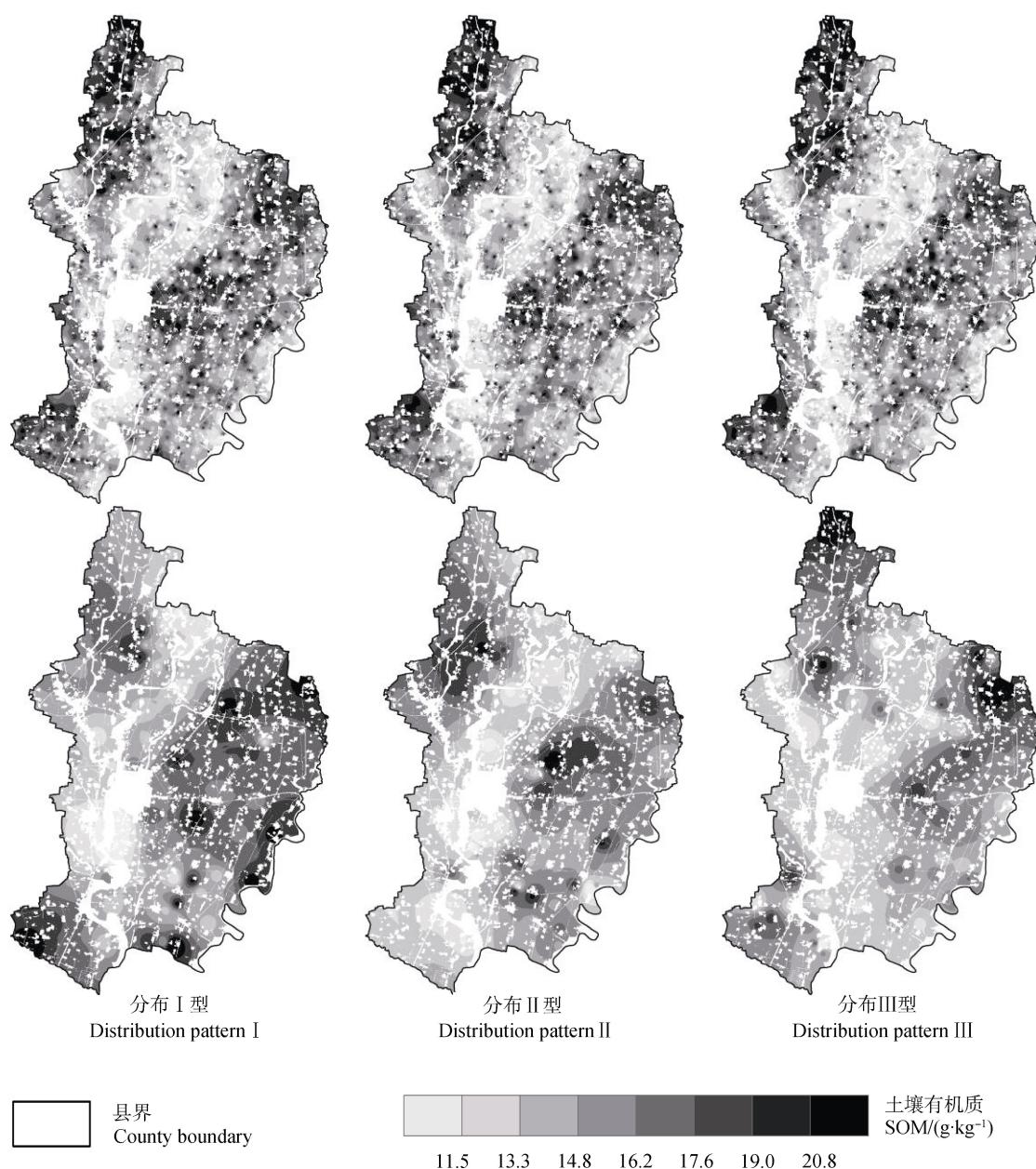
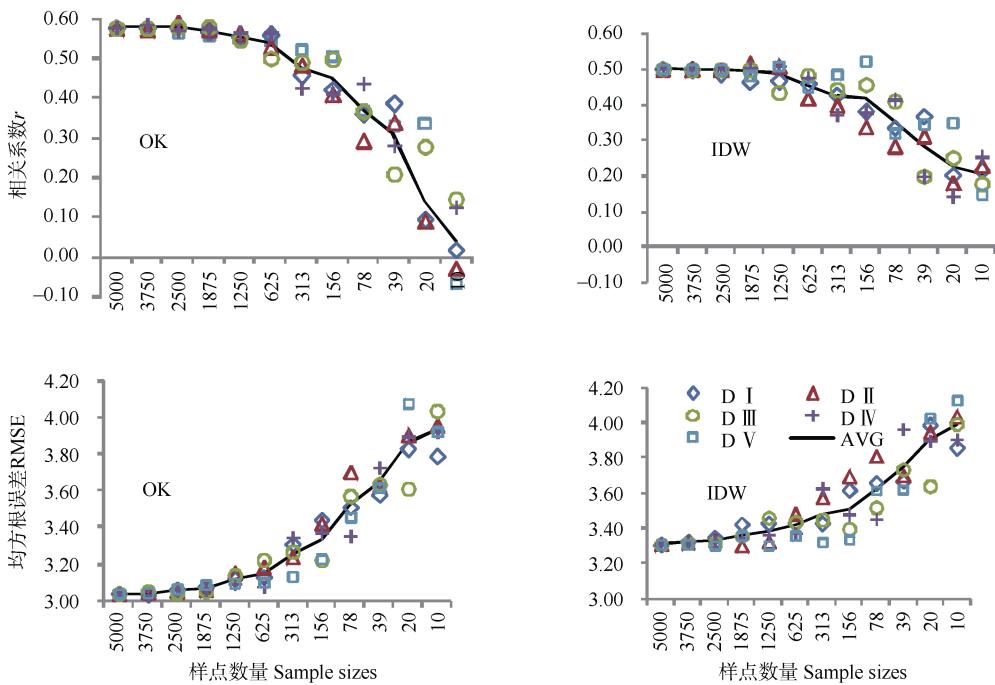


图 7 基于不同样点空间分布的表层土壤有机质含量预测制图 (IDW 插值) (上: 2 500 个样点, 下: 78 个样点)

Fig. 7 Predictive maps of OM content in topsoil layers based on IDW interpolation relative to sampling site spatial distribution pattern and sample size (top: 2 500 samples and, bottom: 78 samples)

精度, 在制图结果发生失真畸变之前 (样点数量大于 20 个), OK 平均插值精度及其整体变幅大于 IDW。陈昌华等<sup>[37]</sup>对不同空间插值方法精度的比较分析表明, 与 OK 相比, IDW 插值精度较低且更快对样本数量的变化做出响应, 这与本研究结果基本一致。导致上述差异的根本原因, 在于 IDW 和 OK 插值技术工作原理的不同: IDW 属于确定性插值方法, 以研究区待测点与训练样点的空间距离确定权

重进行加权平均计算, 训练样点的数量和空间分布对权重的大小具有决定性作用; OK 则属于地统计插值方法, 基于训练样点的土壤属性结构信息 (通过变异函数表征) 进行无偏最优估计, 该方法除了训练样点空间位置信息外还有效利用了训练样点属性值的空间相关性, 与 IDW 方法相比, 土壤属性结构性越强, 训练样点的数量和空间分布对预测结果的影响就越弱<sup>[38]</sup>。



注: D I、D II、D III、D IV、D V 和 AVG 分别表示空间分布 I、II、III、IV、V 型和 5 种分布的平均值 Note: D I, D II, D III, D IV, D V and AVG denotes soil sample distribution pattern I, II, III, IV, V and, the average value of above five distribution patterns, respectively

图 8 土壤样点数量和空间分布对表层土壤有机质含量预测精度的影响

Fig. 8 Influence of sample size and sampling site spatial distribution pattern on accuracy of the prediction of OM contents in topsoil layers

### 3 结 论

在县域尺度上,土壤样点数量和样点空间分布均会对研究区基于 OK 和 IDW 插值技术的土壤空间预测结果产生重要影响,这些影响可通过预测制图结果的空间特征以及预测精度参数的变化进行分析与评估。当样点数量足够多时,研究区样点数量和空间分布对 OK 和 IDW 插值预测制图和预测精度的影响非常有限;当样点数量减少到一定程度时,随着样点数量的减少,预测制图结果的空间局部变异细节信息逐渐减少,预测精度逐渐下降,同时样点空间分布对预测结果的影响不断增强,甚至可能取代样点数量成为预测结果的决定性因素。由于工作原理的不同,在空间预测结果发生失真畸变之前,与 OK 相比, IDW 插值精度较低且更早响应样点数量和空间分布的变化。

### 参 考 文 献

- [ 1 ] Hengl T, Mendes de Jesus J, Heuvelink G B M, et al. Soilgrids250m: Global gridded soil information based on machine learning. *PLoS One*, 2017, 12 (2): 1—40
- [ 2 ] McBratney A B, Mendonça Santos M L, Minasny B. On digital soil mapping. *Geoderma*, 2003, 117: 3—52
- [ 3 ] Lagacherie P, McBratney A B. Chapter 1 spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping // Lagacherie P, McBratney A B. *Developments in soil science*. Elsevier, 2006: 3—22
- [ 4 ] Minasny B, McBratney A B. Digital soil mapping: A brief history and some lessons. *Geoderma*, 2016, 264: 301—311
- [ 5 ] Hartemink A E, McBratney A, Mendonça Santos M L. Digital soil mapping with limited data//Ahrens R J. Foreword of digital soil mapping with limited data. Netherlands: Springer, 2008: V—VII
- [ 6 ] 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. *土壤学报*, 2015, 52 (4): 783—791
- [ 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 (In Chinese). *Acta Pedologica Sinica*, 2015, 52 (4): 783—791
- [ 8 ] 上官魁星, 吴金水, 周脚根, 等. 县域尺度土壤有机碳储量估算的样点密度优化. *土壤学报*, 2014, 51 (1): 41—48
- [ 9 ] ShuangGuan K X, Wu J S, Zhou J G, et al. Optimum density of sampling for estimation of soil organic carbon

- stock at a county scale ( In Chinese ) . *Acta Pedologica Sinica*, 2014, 51 ( 1 ): 41—48
- [ 8 ] 孙志英, 赵彦锋, 张化楠, 等. 基于地图直接对比的土壤空间表达尺度效应研究——以封丘土壤属性制图为例. *土壤学报*, 2014, 51 ( 4 ): 781—794
- Sun Z Y, Zhao Y F, Zhang H N, et al. Study on effect of scaling on soil spatial expression based on direct comparison of maps-A case of soil attribute mapping in Fengqiu County, Henan Province, China ( In Chinese ) . *Acta Pedologica Sinica*, 2014, 51 ( 4 ): 781—794
- [ 9 ] 于雷, 魏东, 王惠霞, 等. 江汉平原县域尺度土壤有机质空间变异特征与合理采样数研究. *自然资源学报*, 2016, 31 ( 5 ): 855—863
- Yu L, Wei D, Wang H X, et al. Spatial variability of soil organic matter and appropriate number of samples on county scale in Jianghan plain ( In Chinese ) . *Journal of Natural Resources*, 2016, 31 ( 5 ): 855—863
- [ 10 ] 张圣民, 许明祥, 张志霞, 等. 黄土高原不同地貌类型区农田土壤有机碳采样布点方法研究. *自然资源学报*, 2018, 33 ( 4 ): 634—643
- Zhang S M, Xu M X, Zhang Z X, et al. Methods of sampling soil organic carbon in farmlands with different landform types on the Loess Plateau ( In Chinese ) . *Journal of Natural Resources*, 2018, 33 ( 4 ): 634—643
- [ 11 ] 张晓光, 王志刚, 宋祥云, 等. 采样点数量对黄河三角洲垦利县土壤盐分空间变异的影响. *资源科学*, 2016, 38 ( 12 ): 2375—2382
- Zhang X G, Wang Z G, Song X Y, et al. Effect of sampling on spatial variability in soil salinity in the Yellow River Delta Area ( In Chinese ) . *Resources Science*, 2016, 38 ( 12 ): 2375—2382
- [ 12 ] 张忠启, 于东升, 胡丹, 等. 县域尺度土壤全氮空间变异及合理采样点数量. *江苏农业学报*, 2016, 32 ( 4 ): 798—802
- Zhang Z Q, Yu D S, Hu D, et al. Spatial variability and reasonable sampling size of soil total nitrogen at county scale ( In Chinese ) . *Jiangsu Journal of Agricultural Sciences*, 2016, 32 ( 4 ): 798—802
- [ 13 ] 张忠启, 于法展, 于东升, 等. 红壤区土壤有机碳时间变异及合理采样点数量研究. *土壤学报*, 2016, 53 ( 4 ): 891—900
- Zhang Z Q, Yu F Z, Yu D S, et al. Temporal variability of soil organic carbon and rational sample size in red soil region ( In Chinese ) . *Acta Pedologica Sinica*, 2016, 53 ( 4 ): 891—900
- [ 14 ] 张贝尔, 黄标, 赵永存, 等. 采样数量与空间插值方法对华北平原典型区土壤质量评价空间预测精度的影响. *土壤*, 2013, 45 ( 3 ): 540—547
- Zhang B E, Huang B, Zhao Y C, et al. Effects of sampling sizes and spatial interpolation method on spatial prediction accuracy of soil fertility quality index in the major grain-producing region of the North China Plain ( In Chinese ) . *Soils*, 2013, 45 ( 3 ): 540—547
- [ 15 ] de Carvalho W, Chagas C D, Muselli A, et al. Conditioned latin hypercube method for soil sampling in the presence of environmental covariates for digital soil mapping. *Revista Brasileira De Ciencia Do Solo*, 2014, 38 ( 2 ): 386—396
- [ 16 ] Li Y, Zhu A X, Shi Z, et al. Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain. *Geoderma*, 2016, 284: 73—84
- [ 17 ] Zhang S J, Zhu A X, Liu J, et al. An heuristic uncertainty directed field sampling design for digital soil mapping. *Geoderma*, 2016, 267: 123—136
- [ 18 ] 韩宗伟, 黄魏, 罗云, 等. 基于路网的土壤采样布局优化——模拟退火神经网络算法. *应用生态学报*, 2015, 26 ( 3 ): 891—900
- Han Z W, Huang W, Luo Y, et al. Application of simulated annealing method and neural network on optimizing soil sampling schemes based on road distribution ( In Chinese ) . *Chinese Journal of Applied Ecology*, 2015, 26 ( 3 ): 891—900
- [ 19 ] 刘京, 朱阿兴, 张淑杰, 等. 基于样点个体代表性的大尺度土壤属性制图方法. *土壤学报*, 2013, 50 ( 1 ): 12—20
- Liu J, Zhu A X, Zhang S J, et al. Large-scaled soil attribute mapping method based on individual representativeness of sample sites ( In Chinese ) . *Acta Pedologica Sinica*, 2013, 50 ( 1 ): 12—20
- [ 20 ] 钟克强, 郭澎涛, 武伟, 等. 基于模拟退火算法的土壤样点布局和土壤景观模型优化. *土壤通报*, 2017, 48 ( 1 ): 14—21
- Zhong K Q, Guo P T, Wu W, et al. Optimization of soil sampling schemes and soil landscape model based on simulate annealing algorithm ( In Chinese ) . *Chinese Journal of Soil Science*, 2017, 48 ( 1 ): 14—21
- [ 21 ] van Arkel Z, Kaleita A L. Identifying sampling locations for field-scale soil moisture estimation using k-means clustering. *Water Resources Research*, 2014, 50 ( 8 ): 7050—7057
- [ 22 ] Yang L, Qi F, Zhu A X, et al. Evaluation of integrative hierarchical stepwise sampling for digital soil mapping. *Soil Science Society of America Journal*, 2016, 80 ( 3 ): 637—651
- [ 23 ] 刘雪琦, 朱阿兴, 杨琳, 等. 土壤图更新中基于土壤类型面积分级的训练样点选择方法. *土壤学报*, 2017, 54 ( 1 ): 36—47
- Liu X Q, Zhu A X, Yang L, et al. Training sample selection method based on grading of soil types by area for updating conventional soil maps ( In Chinese ) . *Acta Pedologica Sinica*, 2017, 54 ( 1 ): 36—47
- [ 24 ] 江叶枫, 郭熙, 叶英聪, 等. 基于辅助变量和神经网络模型的土壤有机质空间分布模拟. *长江流域资源与环*

- 境, 2017, 26 (8): 1150—1158
- Jiang Y F, Guo X, Ye Y C, et al. Simulation of distribution of soil organic matter based on auxiliary variables and neural network model (In Chinese). Resources and Environment in the Yangtze Basin, 2017, 26 (8): 1150—1158
- [25] 巫振富, 赵彦锋, 陈杰. 土壤属性预测中 kriging 参数对插值精度的影响研究. 土壤通报, 2012, 43 (3): 551—558
- Wu Z F, Zhao Y F, Chen J. Effects of parameter condition on prediction precision of soil attributes by kriging interpolation (In Chinese). Chinese Journal of Soil Science, 2012, 43 (3): 551—558
- [26] 吴学文, 晏路明. 普通 kriging 法的参数设置及变异函数模型选择方法——以福建省一月均温空间内插为例. 地球信息科学, 2007, 9 (3): 104—108
- Wu X W, Yan L M. Setting parameters and choosing optimum semivariogram models of ordinary kriging interpolation—A case study of spatial interpolation to January average temperature of Fujian Province (In Chinese). Geo-Information Science, 2007, 9 (3): 104—108
- [27] 陈思萱, 邹滨, 汤景文. 空间插值方法对土壤重金属污染格局识别的影响. 测绘科学, 2015, 40 (1): 63—67
- Chen S X, Zou B, Tang J W. Impact of spatial interpolation methods on identifying structure of heavy metal polluted soil (In Chinese). Science of Surveying and Mapping, 2015, 40 (1): 63—67
- [28] 张铁婵, 常庆瑞, 刘京. 土壤养分元素空间分布不同插值方法研究——以榆林市榆阳区为例. 干旱地区农业研究, 2010, 28 (2): 177—182
- Zhang T C, Chang Q R, Liu J. Comparison of spatial interpolation methods for soil nutrient elements—A case study of Yuyang county, Shanxi Province (In Chinese). Agricultural Research in the Arid Areas, 2010, 28 (2): 177—182
- [29] Wang Y Q, Zhang X C, Zhang J L, et al. Spatial variability of soil organic carbon in a watershed on the Loess Plateau. Pedosphere, 2009, 19 (4): 486—495
- [30] 王志刚, 赵永存, 黄标, 等. 采样点数量对长三角典型地区土壤肥力指标空间变异解析的影响. 土壤, 2010, 42 (3): 421—428
- Wang Z G, Zhao Y C, Huang B, et al. Effects of sample size on spatial characterization of soil fertility properties in an agricultural area of the Yangtze River Delta Region, China (In Chinese). Soils, 2010, 42 (3): 421—428
- [31] 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. 土壤通报, 2013, 44 (4): 844—850
- 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 (In Chinese). Chinese Journal of Soil Science, 2013, 44 (4): 844—850
- [32] 孙孝林, 王会利, 宁源. 样点代表性等级采样法在丘陵山区土壤表层有机质制图中的应用. 土壤, 2014, 46 (3): 439—445
- Sun X L, Wang H L, Ning Y. Application of sampling method based on representativeness grades of sampling points in topsoil organic matter mapping within hilly area (In Chinese). Soils, 2014, 46 (3): 439—445
- [33] 杨琳, 朱阿兴, 张淑杰, 等. 土壤制图中多等级代表性采样与分层随机采样的对比研究. 土壤学报, 2015, 52 (1): 28—37
- Yang L, Zhu A X, Zhang S J, et al. A comparative study of multi-grade representative sampling and stratified random sampling for soil mapping (In Chinese). Acta Pedologica Sinica, 2015, 52 (1): 28—37
- [34] 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. 土壤通报, 2016, 47 (5): 1056—1064
- Li K, Zhao H F, Wu K N, et al. Suitable interpolation method and reasonable sampling quantity of Cd pollution index in soil (In Chinese). Chinese Journal of Soil Science, 2016, 47 (5): 1056—1064
- [35] 范曼曼, 吴鹏豹, 张欢, 等. 采样密度对土壤有机质空间变异解析的影响. 农业现代化研究, 2016, 37 (3): 594—600
- Fan M M, Wu P B, Zhang H, et al. Effect of sampling density on spatial variability analysis of soil organic matter (In Chinese). Research of Agricultural Modernization, 2016, 37 (3): 594—600
- [36] 齐雁冰, 常庆瑞, 刘梦云, 等. 县域农田土壤养分空间变异及合理样点数确定. 土壤通报, 2014, 45 (3): 556—561
- Qi Y B, Chang Q R, Liu M Y, et al. County-scale spatial variability of soil nutrient distribution and determination of reasonable sampling density (In Chinese). Chinese Journal of Soil Science, 2014, 45 (3): 556—561
- [37] 陈昌华, 陈锡云, 徐英. 基于大尺度低密度样点的东北土壤全氮空间插值方法比较. 水土保持通报, 2014, 34 (6): 153—161
- Chen C H, Chen X Y, Xu Y. Comparison of spatial interpolation methods for soil total nitrogen content at large scale using low density soil survey data in Northeast China (In Chinese). Bulletin of Soil and Water Conservation, 2014, 34 (6): 153—161
- [38] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望. 地理科学进展, 2018, 37 (1): 66—78
- Zhu A X, Yang L, Fan N Q, et al. The review and outlook of digital soil mapping (In Chinese). Progress in Geography, 2018, 37 (1): 66—78

## Influences of Sample Size and Spatial Distribution on Accuracy of Predictive Soil Mapping on A County Scale

WU Zhenfu<sup>1</sup> ZHAO Yanfeng<sup>2</sup> CHENG Daoquan<sup>3</sup> CHEN Jie<sup>2†</sup>

(1 School of Public Administration, Zhengzhou University, Zhengzhou 450001, China)

(2 School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China)

(3 Station of Soil and Fertilizer Extension Service, Henan Province, Zhengzhou 450002, China)

**Abstract** 【Objective】This study was conducted to investigate influences of sample size and spatial distribution on prediction of soil mapping, which is contributive to formulating soil sampling strategies scientifically and improving soil prediction accuracy effectively. 【Method】Out of 5 403 soil samples, a validation dataset and training sub-datasets different in number of soil samples were derived randomly, and each subset encompassed five examples different in sampling site spatial distribution pattern. Influences of sample size and spatial distribution on predictive soil mapping, embodied by spatial distribution characteristic and prediction accuracy, were explored on the basis of the prediction of organic matter content (OMC) in topsoil layer with the Ordinary Kriging (OK) or Inverse Distance Weighted (IDW) interpolation method. 【Result】Results show that when the number of soil samples decreased from 5 000 to 39 the OMC predictive maps based on OK or IDW interpolation was gradually losing details of local variation, and when the number dropped down to 20 or 10, the predictive maps became distorted. When the number varied in the range of 5 000~1 250, the predictive maps based on OK interpolation were quite similar in accuracy with  $r$  varying in the range of 0.55~0.59, and RMSE in the range of 3.03~3.15, but when the number dropped down to 625, the predictive maps based on OK interpolation declined significantly in accuracy, and varied sharply between the five groups different in spatial distribution pattern of sampling sites even in the same training subdset. The predictive map based on IDW interpolation varied in accuracy with the number and the distribution pattern of soil sampling sites on a trend similar to that based on OK interpolation, except that the predictive map based on IDW interpolation declined significantly in accuracy with the number of samples starting to drop from 1875 and varied sharply with spatial distribution pattern of the sampling sites. On average, the predictive maps based on OK interpolation were obviously higher than those based on IDW interpolation in accuracy before the map began to turn distorted. 【Conclusion】All the findings in this study indicate that both sample size and spatial distribution pattern have certain impacts on predictive soil map, and the impacts are quite limited when the sample size is large enough. However, when the sample size drops below a certain level, the predictive maps will lose details of some local variations and prediction accuracy as well, while spatial distribution of sampling sites turns up to be the main affecting factor. Compared with OK, IDW is lower in accuracy and responds to the changes in sample size and spatial distribution earlier before the predictive map is distorted.

**Key words** Soil organic matter; Soil sample; Spatial distribution; Digital soil mapping

(责任编辑: 檀满枝)