

基于土壤—地形关系的合理样点对土壤有机质空间预测的影响

李广月, 毕如田*, 朱洪芬, 陈卓

(山西农业大学资源环境学院, 山西 太谷 030801)

摘要: 针对土壤精细化管理体系中合理样点数及空间预测优化问题, 本文将基于土壤—地形关系, 探讨了不同采样方式下以局部样点数量为代表提供最优数据的可能性。以地统计学、土壤—地形关系、地理加权回归克里格 (GWRK) 模型为基础, 经系统格网、地形单元分区和地形起伏度最佳统计单元等三种采样方式分析合理样点集的空间分布对土壤有机质空间预测精度的影响。结果表明: (1) 确定地形起伏度最佳统计单元大小为 10×10 像元, 且平原区样点分布最为密集, 合理样点数为 1656; (2) 高程、坡向、地形位置指数、相对位置指数、地形起伏度是影响土壤有机质空间变异的主要因素, 能够解释研究区内有机质含量空间变异的 69.2%; (3) GWRK 模型精度均比普通克里金插值 (OK) 精度高, 且山脊、背坡、陡坡、坡脚等坡位内合理样点数分别为 39、481、9、28。在样点数最多时 ($n = 2806$), GWRK 精度提高幅度及样点数量对预测结果影响有限。当样点数量减少时, 有机质预测值空间分布的局部变异性随样点数减少而减少。因此, 不同采样方式下合理样点集明显影响有机质预测精度, 但预测结果分布趋势相似, 仍可完整表征土壤有机质空间分布的空间格局。

关键词: 合理样点数; 地形单元划分; 地理加权回归模型; 空间格局; 地形起伏度最佳统计单元

中图分类号: S151.9 **文献标识码:** A **文章编号:** 0564-3945(2021)03-0515-12

DOI: 10.19336/j.cnki.trtb.2020051901

李广月, 毕如田, 朱洪芬, 陈卓. 基于土壤—地形关系的合理样点对土壤有机质空间预测的影响 [J]. 土壤通报, 2021, 52(3): 515 - 526

LI Guang-yue, BI Ru-tian, ZHU Hong-fen, CHEN Zhuo. Spatial Prediction for Soil Organic Matter Affected by Appropriate Samples Based on Soil-Terrain Relationship[J]. Chinese Journal of Soil Science, 2021, 52(3): 515 - 526

土壤有机质是表征土壤肥力状况的重要指标之一, 在人类活动及土壤成土过程中具有较强的空间异质性。目前研究多采用地统计学, 结合土地利用类型、土壤类型、成土因素等对不同模型 (地统计学模型、半变异模型、决策树、土壤属性推理) 的模拟结果对比分析。随着土壤科学的发展, 土壤学家 Jenny 和 Huggett 提出了土壤—景观模型理论, 依据空间统计及应用数理统计, 并结合如神经网络^[1]、决策树^[2] 等不同数据挖掘算法, 准确提取土壤知识。基于已有研究成果, 本研究将以协同环境变量为辅助变量, 利用土壤成土因素并结合矢量化土壤图以共同表述土壤—地形关系模式, 主要以相关地形因子及土壤有机质的空间多尺度关系为属性源挖掘土壤地形知识以便进行精细化数字土壤制图及预测^[3-5]。

样点属性信息是进行数字土壤制图的主要数据来源, 而合理样点数量的确定及空间格局是土壤采样

策略的核心^[6]。Tobler、Miller^[7-8] 以地理学第一定律为参考, 揭示了空间距离对地理现象及土壤样点空间分布的影响, 即空间位置相关性越大的区域, 土壤属性及其交互强度具有更高的相似性^[9-12], 而过高的相似性与相关性将导致样点信息重合, 对样点有效信息产生影响, 因此采样点合理布局比简单增加采样点数量更为重要^[13]。

目前国内外学者对土壤属性变异及样点数量的研究主要基于实地采样点, 采用的样点布设方法有: 简单随机采样、格网尺度采样、分层随机采样、逐次及高效的补样方案等^[14]。已有研究表明: 不同地形单元、不同土壤养分指标、不同格网尺度下的合理采样数差异明显, 且基于不同地形因子的土壤养分变异性及合理样点数的研究, 不仅提高了采样的工作效率和降低了实验成本, 且影响着取样结果的准确性, 加快了土壤精细化管理的脚步^[15]。

收稿日期: 2020-05-19; 修订日期: 2021-03-13

基金项目: 国土资源部公益性行业项目 (201411007) 资助

作者简介: 李广月 (1994-), 女, 青海西宁人, 硕士研究生。主要研究方向为土地信息与技术。E-mail: 2601613838@qq.com

*通讯作者: E-mail: brt@sxau.edu.cn

综上, 基于土壤—地形关系, 以地形单元分区采样、地形起伏度最佳统计单元采样及系统均匀格网多尺度采样三种采样模式设置了不同样点个数和空间分布的八个土壤样点集, 进而确定合理样点数, 通过比较基于不同采样模式下不同样点系列空间分布的土壤有机质空间预测精度的差异, 探讨基于土壤—地形关系的样点数量空间分布变化对土壤属性空间预测影响的表征能力, 为后期获取空间信息及土壤属性更新奠定基础。

1 材料与方法

1.1 研究区概况

万荣县隶属于山西省运城市, 位于山西省西南部, 地处黄河与汾河交汇处 (110°25′~110°59′E、35°13′~35°31′N, 图 1)。万荣县下辖 14 个乡镇, 是山西省重要的粮食生产基地, 耕地以旱地为主, 总耕地面积 53031.29 hm²。万荣县属温带大陆性季风

气候, 四季分明、雨热同期。境内山峦连绵, 三面环山, 整体地势呈东南、西北向倾斜, 海拔位于 265~1386 m 间, 地形破碎复杂, 以低山丘陵区、山前倾斜平原区、中低山区为主。土壤类型包括潮土、褐土、粗骨土、石质土、新积土五大土类, 七个亚类, 九个土属。土壤呈微碱性, 侵蚀较为严重。

1.2 数据来源

本研究收集的数据包括: 万荣县 2017 年 1:10000 土地利用变更调查数据库、耕地质量等级更新成果 (2017 年)、土壤采样点信息及养分指标等, 以上数据均经空间数据格式转换、校正并投影转化为 CGCS2000 坐标系。地形因子包括基础地形因子 (高程、坡度、坡向、平面曲率、剖面曲率、地形起伏度) 和复杂地形因子 (比汇水面积、汇流动力指数、地形湿度指数、地形位置指数、相对位置指数) 等因子, 通过 ArcGIS 软件及 SimDTA^[16] 软件提取。技术路线如图 2 所示。

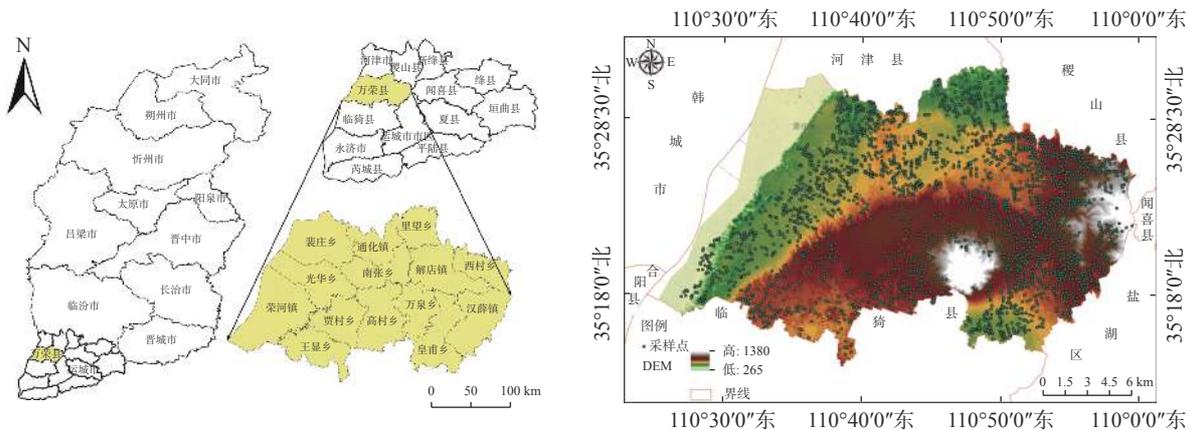


图 1 研究区概况

Fig.1 Basic information of the Study area

1.3 土壤—地形关系模型建立

在土壤—环境单元中, 各环境因素以某种特定的模式相互作用, 基于各组合地形因子的空间分布确定土壤空间特征, 并结合典型对应排序方法 (Canonical correlation analysis, CCA)^[17]。即在对应分析的迭代过程中将获取的地形因子与有机质相关程度进行多元逐步回归, 反映地形因子组合及地形重要值对有机质的影响, 以对应排序值与多元线性逐步回归组合构建土壤—地形模型 (式 1)。运用 IBM SPSS 25.0 对各采样方式下各样点系列的地形因子与有机质进行相关分析, 采用相关系数衡量相关性水平。地形因子分类公式为:

$$y_i = a + \sum_{i=1}^{i=n} b_i x_i + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (1)$$

式中: y_i 为有机质含量, a 为常量, b_i 为回归系数, ε_i 是残差项。

1.4 不同采样方法及合理样点确定

样本代表性指所有样本的平均代表程度, 很多学者采用样本代表性确定合理样点^[18-19], 本研究应用统计学中检验样本代表性系数确定合理采样点数。以平均数代表性检验系数以 $\pm 2\%$ 或 $\pm 3\%$ 以内作为限制范围, 结构代表性差异率以 $\pm 5\%$ 以内作为检验应控制的范围。应用不同采样方式获得的不同样点子集与全局之间同时满足检验系数与差异率的精度要求, 确定的样点子集对全局具有足够代表性^[20]。进

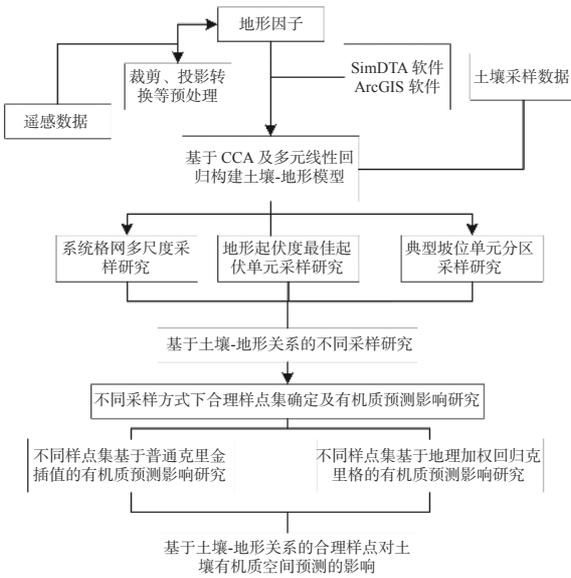


图2 技术流程图

Fig.2 Technology roadmap of the research

一步对土壤有机质含量空间预测精度的影响研究提供依据^[21]。

1.4.1 系统均匀网格多尺度采样 利用 ArcEngine 二次开发软件, 以研究区域行政界线外接矩形为输入范围分别创建生成 2000 m × 2000 m (G1, 233 个样点)、1500 m × 1500 m (G2, 390 个样点)、1000 m × 1000 m (G3, 762 个样点)、500 m × 500 m (G4, 1720 个样点), 共 4 个尺度的格网数据集, 将样点与格网进行属性连接并赋予唯一标识码, 获取不同尺度格网内样点属性^[22]。利用 ArcGIS 近邻分析取距格网中心最近的样点作为代表研究区域内采用点属性信息的代表性样点, 分别形成含 233、390、762、1720 个样点的子集。

1.4.2 地形起伏度最佳统计单元采样 以地形起伏因子为代表, 分析其最佳统计单元内合理样点集的有机质空间分布规律, 运用典型均值变点分析法^[23]确定最佳变点, 获取研究区地形起伏度最佳统计单元^[24-25]。以拟合系数 0.8 和均方根误差绝对值相差 0.200 g/kg 为标准, 依据二分原则, 正确符合采样间断小于 20 个样点的要求形成含有 2708、1724、1656、1379 个样点的样点子集。

均值变点法确定变点过程: (1) 利用焦点统计及邻域分析计算各像元邻域单元的平均起伏度 (t_i), 并统计相对应像元的邻域面积 (S_i), 获取平均起伏度 (T_i), 其值为平均起伏度与相应邻域面积的比值 ($i=2, 3, \dots, 30$); (2) 在不同统计窗口下, 取平

均起伏度的对数, 获取序列 X , X 是关于 t 的集合, $\{x_t, t=1, 2, 3, \dots, 29\}$; (3) 根据相应公式计算获取统计量、均值 ($\bar{X}_{i1}, \bar{X}_{i2}$)、方差 (S); (4) 计算 $S-St$ 的值 T_i , 得到归一化值 G_i , 曲线上平均地形起伏度随统计面积的增大而增大的趋势由陡变缓的拐点为最佳变点, 如图 3 所示, 表达式如下:

$$\bar{X} = \sum_{i=1}^n x_i/n \tag{2}$$

$$S = \sum_{i=1}^n (x_i - \bar{X})^2 \tag{3}$$

$$S_t = \sum_{i=1}^{t-1} (x_i - \bar{X}_{i1})^2 + \sum_{i=t}^n (x_i - \bar{X}_{i2})^2 \tag{4}$$

$$\bar{X}_{i1} = \sum_{i=1}^{t-1} X_i/n \quad \bar{X}_{i2} = \sum_{i=t}^n X_i/n \tag{5}$$

$$G_i = (T_i - T_{min}) / (T_{max} - T_{min}) \tag{6}$$

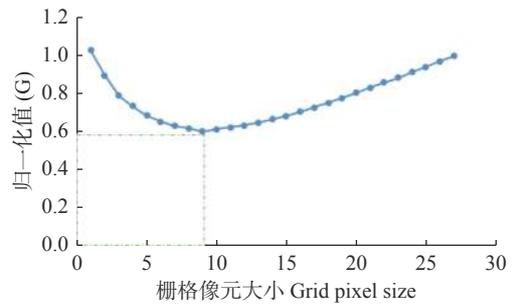


图3 Gt 随栅格像元大小变化曲线

Fig.3 Variation curve of Gt with grid pixel size

1.4.3 地形单元分区采样 本研究结合研究区实际情况及不同地形因子组合, 按照相关标准分级并赋值, 采用区域统计工具分区统计不同级别内的数据并叠加, 划分为若干个小单元, 以确定地形单元 (表 1)。

不同地形条件下各地形因子分布规律不同, 将背坡、沟谷、陡坡按山脉走向及河流流向划分为西南、东北侧, 结合实际情况共划分为 13 种地形单元类型, 依据地形单元划分 (图 4-b) 获取相对应地形单元的采样点。

1.5 样点属性空间预测优化

地理加权回归克里格 (GWRK) 是地理加权回归的扩展, 是对空间中确定的合理样点做线性局部拟合, 用邻近实际值进行局部加权回归得到相关系数, 并将局部拟合残差进行普通克里格插值的一种方法^[26]。其模型为:

$$y_i = \beta_0(ui, vi) + \sum_k \beta_k(ui, vi) x_{ik} + \epsilon_i, i = 1, 2, \dots, n \tag{7}$$

表 1 地形单元划分参数设置
Table 1 Parameters of the classification for topographic units

地形单元类型 Terrain unit type	相对位置指数 RPI	剖面曲率 K _v	平面曲率 K _h	坡度 Slope	坡向 Aspect	修正地形位置指数 TPI* (SD)	修正地形湿度指数 TWI*
背坡 (N/EN)	[0.3, 0.6]	[-0.25, 0.03]	-	[6°, 25°]	0° ~ 135°	0.3 ~ 0.7	-
背坡 (S/WS)					135° ~ 315°		
沟谷 (N/EN)	≤ 0.1	[-0.008, 0.1]	-	[-1°, 6°]	0° ~ 135°	≤ 0.3	-
沟谷 (S/WS)					135° ~ 315°		
陡坡 (S/WS)	-	-	-	[25°, 65°]	0° ~ 135°	-	-
陡坡 (S/WS)					135° ~ 315°		
坡脚	[0.1, 0.2]	≤ -0.008	-	-	-	≤ 0.3	-
山脊	≥ 1	-	≥ 0.5	-	-	> 1	-
坡肩	[0.6, 0.7]	≥ 0	-	-	-	0.7 ~ 1	-
旱平地	-	-	-	-	-	0.3 ~ 0.7	< 2.9
平顶峰	-	-	-	-	-	> 1	-
湿平地	-	-	-	-	-	0.3 ~ 0.7	≥ 2.9
河漫滩	-	-	-	-	-	-	≥ 2.9

注: N/EN-北/东北侧, S/WS-南/西南侧, TPI* (SD): Topographic position index-修正地形位置指数标准差

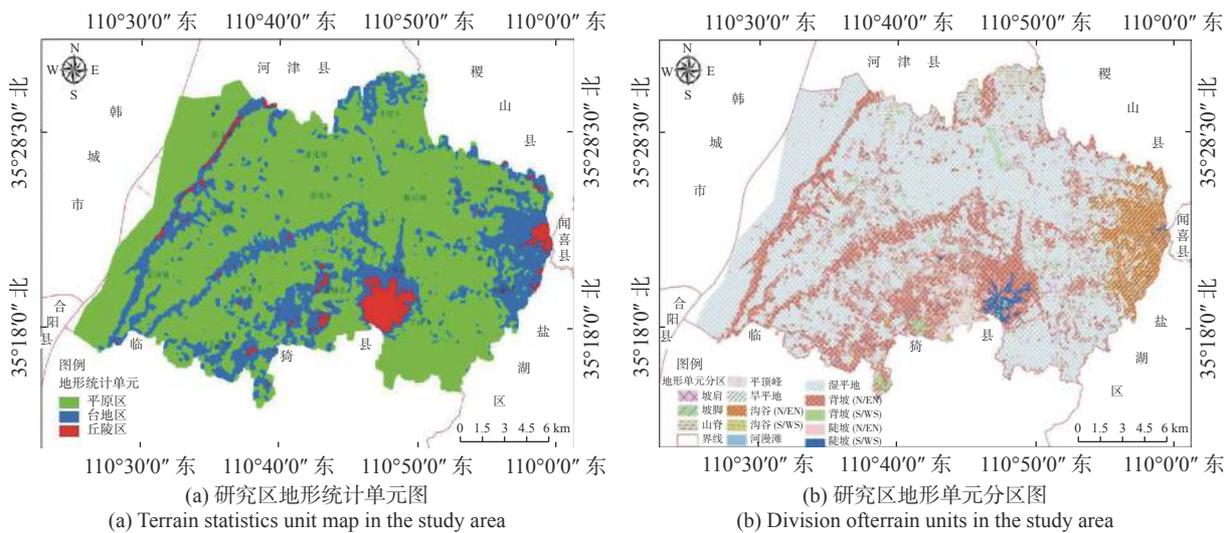


图 4 研究区地形单元
Fig.4 Terrain unit in the study area

$$Y_{GWRK}(ui,vi) = Y_{GWR}(ui,vi) + \epsilon_{OK}(ui,vi) \quad (8)$$

式中: y_i 为预测值, (ui,vi) 是第*i*个样点地理坐标; $\beta_k(ui,vi)$ 为*i*点的第*k*个局部预测系数, x_{ik} 为有机质含量 x_k 在*i*点的含量值, ϵ_i 是残差项。 $Y_{GWRK}(ui,vi)$ 为土壤有机质在*i*点处的预测值, $Y_{GWR}(ui,vi)$ 是*i*点处的地理加权回归结果, $\epsilon_{OK}(ui,vi)$ 为残差插值结果。

2 结果与分析

2.1 研究区地形、土壤属性描述性统计

土壤有机质描述性统计分析见表 2、表 3, 由表 2 可知有机质含量变化范围在 5.61 ~ 20.39 g kg⁻¹ (2008 年)、6.00 ~ 26.70 g kg⁻¹ (2017 年), 平均含量为 11.693 ± 0.0307 g kg⁻¹。根据全国土壤养分分

级标准^[27]: 有机质含量在 III 级间的样点数最多, 占全部样点数的 70.1%, 在 IV 级间的样点占 18.14%, II 级间的样点占 10.50%, 变异系数在 18.90% ~ 20.61% 之间。变异程度随样点数的增多而越大, 空间自相关性随之增强。采用地统计模块中探索性数据分析, 经对数转换各系列样点属性值均符合正态分布, 且满足地统计学对样点数据的要求。有机质在湿平地为最大 (23.88 g kg⁻¹), 山脊处为在 7.82 ~ 19.4 g kg⁻¹, 平顶峰处为 6.93 ~ 18.73 g kg⁻¹, 表明坡度较大的区域 (山脊、平顶峰) 内有机质变化趋势相似, 各地形单元内有机质变异程度均为中等变异, 相邻地形单元内有机质分布范围大致相同, 从 2008 年到 2017 年, 有机质含量明显升高。

表 2 同采样方式下各样点集有机质统计特征
Table 2 Descriptive statistics analysis for soil organic matter at grid scales in 2008

样点个数 Sample numbers	最小值 (g kg ⁻¹) Minimum	最大值 (g kg ⁻¹) Maximum	平均值 (g kg ⁻¹) Mean	标准差 (g kg ⁻¹) Standard deviation	中值 (g kg ⁻¹) Medium	变异系数 (%) Coefficient of variation	峰度 Kurtosis	偏度 Skewness
2708	5.38	24.58	12.08	2.46	11.82	20.36	4.83	0.82
2167	5.38	24.58	12.09	2.47	11.83	20.43	4.78	0.81
1724	5.38	24.58	12.10	2.44	11.87	20.17	4.60	0.73
1656	5.75	24.58	12.01	2.47	11.74	20.57	5.04	0.84
1379	5.38	24.58	12.08	2.45	11.82	20.28	4.83	0.79
1103	5.38	24.58	12.06	2.45	11.77	20.32	4.84	0.82
882	5.38	24.58	12.07	2.44	11.74	20.22	4.99	0.84
762	6.30	19.68	11.95	2.30	11.75	19.22	3.25	0.45
390	5.85	19.54	11.85	2.35	11.63	19.84	3.47	0.55
233	6.15	20.39	11.86	2.44	11.59	20.61	4.07	0.77

表 3 2008 年地形单元分区内有机质描述性统计特征
Table 3 Descriptive statistics analysis for soil organic matter at terrain unit in 2008

地形单元 Terrain unit	样点个数 Sample numbers	最小值 (g kg ⁻¹) Minimum	最大值 (g kg ⁻¹) Maximum	平均值 (g kg ⁻¹) Mean	标准差 Standard deviation	中值 (g kg ⁻¹) Medium	变异系数 (%) Coefficient of variation	峰度 Kurtosis	偏度 Skewness
旱平地	24	8.52	17.79	12.89	2.35	12.65	18.23	2.25	0.17
湿平地	1914	5.61	23.88	12.11	2.47	11.83	20.39	4.48	0.76
山脊	39	7.82	19.4	12.36	2.95	11.50	23.87	2.94	0.90
坡肩	23	6.75	17.13	11.3	2.40	11.2	21.24	3.48	0.30
平顶峰	61	6.93	18.73	11.23	2.15	10.80	19.15	4.39	0.75
背坡 (N/EN)	481	5.75	20.13	11.99	2.28	11.82	19.02	3.95	0.57
背坡 (S/WS)	38	5.38	17.58	11.26	2.70	11.96	23.98	2.87	0.0007
沟谷 (N/EN)	167	7.59	24.58	12.28	2.58	12.17	21.01	8.66	1.65
沟谷 (S/WS)	19	7.82	14.9	11.53	1.86	11.42	16.13	2.91	0.02
陡坡 (N/EN)	9	8.76	15.71	11.75	2.09	11.59	17.79	3.32	0.67
陡坡 (S/WS)	3	8.61	10.78	9.84	1.12	10.15	11.38	1.5	-0.46
坡脚	28	7.31	23.67	12.18	3.43	11.73	28.16	6.10	1.54
全局地形	2806	5.38	24.58	12.07	2.46	11.82	20.38	4.79	0.81

2.2 土壤属性信息表达

2.2.1 土壤属性空间异质性分析 土壤属性的最优半变异函数模型由表 4 所示, R^2 由大到小依次为 G4、G3、G2、G1, G4 下样点间距小, 样点空间分布密集, 相关程度大, 相关系数越小则样点空间分布越广, 变程值越大。因此, 从 G1 到 G4 变程的有效变化波动较大, 块金值呈下降趋势, 块基比均小于 25%, 表明不同尺度格网内样点变异强度受模型随机因素的影响较小, 受结构因素影响较大。该结果证明了基于地形因子确定合理样点的可行性各个相关参数均表征样点土壤属性空间变异结构性特征, 且对样点变异性分析精度有间接影响。

各地形单元类型的半变异模型中, R^2 最大的地形单元是沟谷 (N/EN) 的球状模型。沟谷 (南/西南侧)、坡肩、旱平地的理论模型均为高斯模型, 陡坡 (北/东北侧)、坡脚单元理论模型为线性模型, 陡坡 (南/西南侧)、湿平地为球状模型, 且 R^2 值仅次于沟谷, 山脊、平顶峰的理论模型均为指数模型 (表 5)。

2.2.2 土壤—地形关系表达 高程、地形起伏度、相对位置指数、地形位置指数等对有机质空间变异及预测影响较大, 坡度、剖面曲率、地形湿度指数对其空间变异影响较小 (表 6)。高程、坡向、地形位置指数、相对位置指数、地形起伏度等因子是影响土壤有机质空间变异的主要因素可解释有机质空间变异的 69.2%。综上, 土壤—地形线性回归模型可表征土壤属性的 20%~88% 的变异^[28]。

2.3 三种采样方式下合理样点系列确定及有机质空间分布

采用样本代表性评价方法确定基于系统格网采样、地形起伏最佳统计采样、地形单元分区采样方式下的合理样点数。

2.3.1 基于系统格网多尺度合理采样系列确定及有机质空间分布 将样点与格网属性连接并赋予唯一标识码, 得到不同尺度格网中所有样点的信息, 利用 ArcGIS 的邻近功能, 取得距格网中心最近样点作为代表性样点, 分别形成 233、390、762、1720 个

表 4 有机质多尺度半变异函数模型参数表
Table 4 Variogram parameters model of soil organic matter in multiscale grids

尺度代码 Scale code	模型 Model	块金值 Nugget/ (C_0)	基台值 Sill/ (C_0+C)	变程 (km) Range	基底效应 (%) $C_0 / (C_0+C)$	R^2	残差 RSS
G1	E	0.76	6.06	2130.00	12.54	0.40	0.84
G2	E	0.78	5.69	1890.00	13.72	0.30	0.82
G3	E	0.64	5.31	990.00	12.04	0.77	0.08
G4	E	0.67	5.21	1500.00	12.36	0.82	0.05

注: E-指数模型 (Exponential model), 以下同。

表 5 各地形单元有机质半变异函数模型参数表
Table 5 Parameters of semi- variogram model of soil organic matter at terrain units

地形单元 Terrain unit	模型 Model	块金值 Nugget/ (C_0)	基台值 Sill/ (C_0+C)	变程 (km) Range	基底效应 (%) $C_0 / (C_0+C)$	R^2	残差 RSS
背坡 (N/EN)	E	0.72	5.31	1650.00	13.56	0.58	0.39
背坡 (S/WS)	E	1.26	7.20	1320.00	17.50	0.11	0.60
沟谷 (N/EN)	G	2.54	26.07	3521.99	9.74	0.84	0.81
沟谷 (S/WS)	G	0.01	3.95	3602.67	0.26	0.22	0.36
陡坡 (N/EN)	L	0.64	3.65	1508.76	17.50	0.19	0.38
陡坡 (S/WS)	S	0.96	3.68	3810.00	25.97	0.14	0.13
坡脚	L	2.80	12.80	3076.86	21.89	0.56	0.44
坡肩	G	3.58	7.17	1195.15	49.93	0.44	0.21
山脊	E	6.74	13.49	3980.01	49.96	0.81	0.12
旱平地	G	3.06	6.47	4451.37	47.30	0.11	0.14
湿平地	S	0.48	6.25	1380.00	7.69	0.82	0.09
平顶峰	E	3.65	9.07	3960.00	40.25	0.27	0.18

注: L-线性模型 (Linear model), E-指数模型 (Exponential model), S-球状模型 (Spherical model) G-高斯模型 (Gaussian model), 以下同。

表 6 不同地形因子与有机质 CCA 相关性分析
Table 6 CCA correlation analysis between soil organic matter and topographic factors

	地形起伏度 Relief amplitude	高程 Dem	坡度 Slope	坡向 Aspect	地形湿度指数 TWI	剖面曲率 k_p	相对位置指数 RPI	地形位置指数 TPI
Pearson相关性	0.439*	0.842**	-0.060	-0.119**	0.104	-0.038	0.023*	0.231*

注: *在0.05级别 (双尾), 相关性显著, **在0.01级别 (双尾), 相关性显著, RPI-Relative Position Index, 以下同。

样点子集 (图 5)。

由表 7 可知: 当样点数为 233、390 时, OK 预测的有机质变化范围为 $5.85 \sim 20.39 \text{ g kg}^{-1}$, 由 GWRK 获取的有机质范围为 $7.09 \sim 12.44 \text{ g kg}^{-1}$, 均匀格网尺度较大时, 有机质整体分布与 OK 基本一致。当样点数量为 762、1720 时, 由 OK 预测的有机质变化范围为 $5.61 \sim 19.93 \text{ g kg}^{-1}$, 由 GWRK 法预测的有机质变化范围为 $6.22 \sim 18.56 \text{ g kg}^{-1}$ 。由系统格网采样分析得, 当格网尺度达 500 m 时, 均值代表性检验精度满足要求, 格网尺度小于 1 km 时, 结构代表性差异精度满足要求, 因此, 万荣县格网尺度最大不超过 1 km, 此时合理采样点数量为 762。

2.3.2 基于地形起伏度最佳统计单元合理采样系列确定及有机质空间分布 经均值变点法分析得变点系列为 10 像元大小为最佳地形统计像元, 将统计单元划分为 7 种类型: 平原、台地、丘陵区、小起伏山地、中起伏山地、大起伏山地和极大起伏山地^[29], 按

以上标准划分研究区地势为平原区 ($0 \sim 23 \text{ m}$)、台地区 ($23 \sim 59 \text{ m}$)、丘陵区 ($59 \sim 176 \text{ m}$) (图 4-a 所示)。

当样点不少于 1724 个样点时, 均方根误差呈上升趋势, 变化范围均 $\leq 0.200 \text{ g kg}^{-1}$, 拟合系数达 80.3% (表 8), 表明其结果可预测整个研究区不同地形统计单元上有机质空间分布。当样点数量降至 1379 时, 均方根误差变化呈上升趋势, 超出阈值范围, 合理采样数区间定为 (1379, 1724)。当样点数量达 1656 时, 有机质均满足平均代表性检验精度和结构代表性差异精度, 将研究区最佳地形统计单元内有机质的合理样点数量定为 1656, 经研究区地势划分, 平原区样点分布最密集, 台地次之, 样点分布较少区为丘陵区。

2.3.3 基于典型坡位单元分区采样及有机质空间分布 土壤有机质高值区较多分布于沟谷、旱/湿平地等地形单元, 低值区主要分布在山脊、背坡等地形

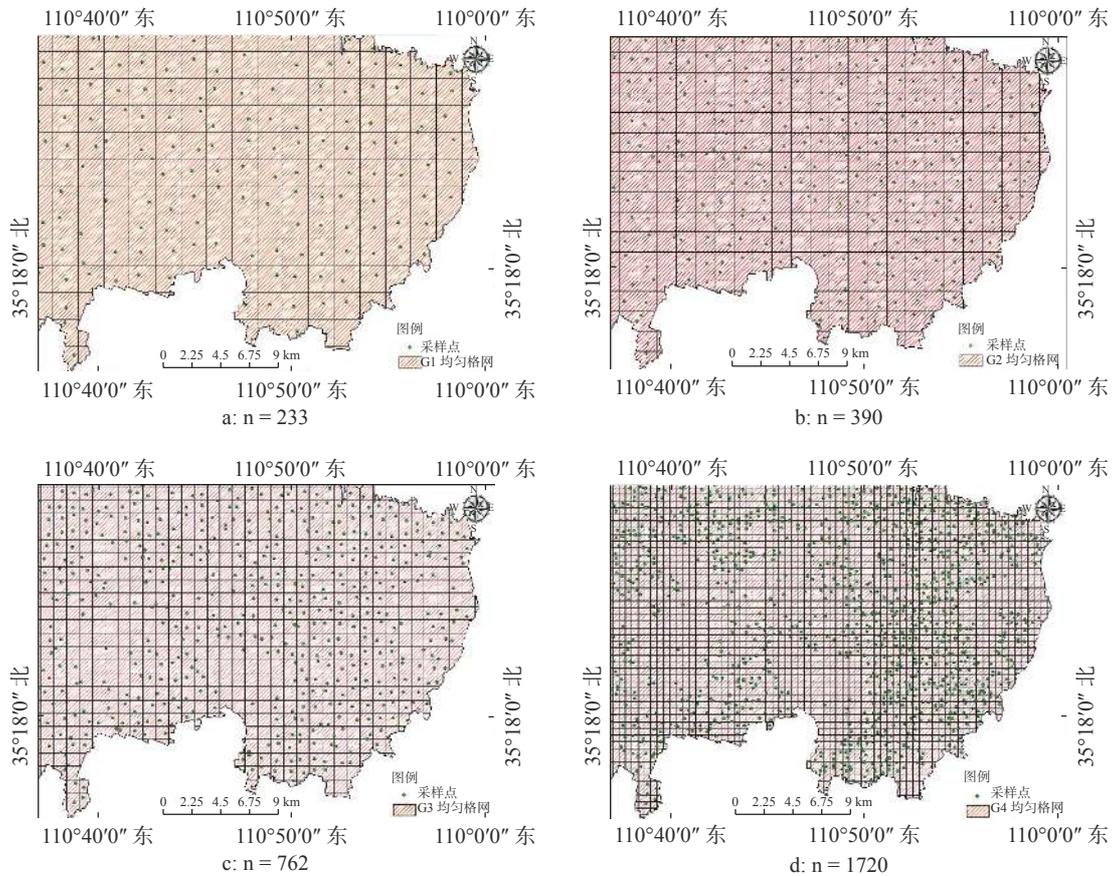


图 5 不同尺度系统格网及采样点分布图

Fig.5 Grid of different scale systems and distribution of sampling points

表 7 不同尺度有机质空间分布统计及样点代表性评价

Table 7 Statistical analysis for the spatial distribution of soil organic matter at different scales

格网尺度 Grid scale	样点个数/个 Samples	实测值范围 (g kg ⁻¹) Observed value range	预测值范围 (g kg ⁻¹) Predicted value range	残差值范围 (g kg ⁻¹) Residual value range	检验系数 (%) Inspection coefficient	差异率 (%) Difference ratio
G1	233	6.15 ~ 20.39	8.10 ~ 16.42	-2.04 ~ 4.41	1.74	0.81
G2	390	5.85 ~ 19.54	7.09 ~ 12.44	-2.45 ~ 2.63	1.82	4.47
G3	762	6.3 ~ 19.68	6.78 ~ 17.75	-2.46 ~ 3.19	0.99	6.50
G4	1720	5.61 ~ 19.93	6.22 ~ 18.56	-2.64 ~ 3.25	0.25	0.81

表 8 最佳地形统计单元下有机质显著性检验及样点代表性评价

Table 8 The significance test and representative evaluation of relief amplitude on the best statistical unit

样点个数/个 Sample numbers	分布类型 Distribution	均方根误差 Root mean of squared error	拟合系数 Determination coefficient	显著性 (LSD检验) Sig	检验系数 (%) Inspection coefficient	差异率 (%) Difference ratio
2708	正态	2.395	0.824	0.256	0.08	2.44
2167	正态	2.402	0.812	0.281	0.17	0.41
1724	正态	2.419	0.803	0.396	0.33	0.81
1656	正态	2.537	0.790	0.387	1.24	5.28
1379	正态	2.647	0.745	0.377	0.08	0.41
1103	正态	2.658	0.733	0.299	0.08	0.41
882	正态	2.746	0.729	0.472	0.06	0.81

单元, 该单元内地形凹凸不平, 土壤侵蚀严重, 有机质积累较少, 山脊与沟谷之间的过渡地形区呈明显渐变规律, 有机质由山脊向沟谷递增。当样点分布在不同地形单元内时, 可由土壤—地形关系表征

研究区内有机质的空间结构性。

全局地形单元平均检验系数为 2.85% (表 9)。整体上, 各地形单元检验系数随样点数减少而增加。山脊、背坡 (N/EN)、陡坡 (N/EN)、坡脚等单元

内有机质同时满足平均数代表性检验精度和结构代表性差异精度, 因此万荣县采用地形单元分区布设

样点至少应达到 39 (山脊)、481 (背坡)、9 (陡坡)、28 (坡脚)、沟谷 (167)。

表 9 地形单元分区样点代表性评价
Table 9 The representative evaluation of division samples at terrain unit in 2008

地形单元 Terrain unit	样点个数 Sample numbers	平均值 (g kg^{-1}) Mean	标准差 Standard deviation	差异率 (%) Difference ratio	检验系数 (%) Inspection coefficient
旱平地	24	12.89	2.35	4.47%	-6.79%
湿平地	1914	12.11	2.47	-0.41%	-0.33%
山脊	39	12.36	2.95	-19.92%	-2.40%
坡肩	23	11.3	2.40	2.44%	6.38%
平顶峰	61	11.23	2.15	12.61%	6.96%
背坡 (N/EN)	481	11.99	2.28	7.32%	0.66%
背坡 (S/WS)	38	11.26	2.70	-9.76%	6.71%
沟谷 (N/EN)	167	12.28	2.58	-4.88%	-1.74%
沟谷 (S/WS)	19	11.53	1.86	24.39%	4.47%
陡坡 (N/EN)	9	11.75	2.09	15.04%	2.65%
陡坡 (S/WS)	3	9.84	1.12	54.47%	18.48%
坡脚	28	12.18	3.43	-39.43%	-0.91%
全局地形	2806	12.07	2.46	-	-

2.4 三种采样方式下合理样点集土壤有机质空间预测影响及精度分析

不同采样方式下合理样点的土壤有机质普通克里金及地理加权回归克里格的空间预测如图 6、图 7 所示。有机质空间分布与原始空间分布基本相同。有机质空间分布呈由中部向周边逐渐升高的趋势。从制图效果分析, 经 GWRK 插值有机质空间分布趋势更明显, 与地形起伏度、地形位置指数等地形因子的空间分布趋势更贴近, 结构性因素 (地形、气候等) 是导致有机质变异的主导因素。

不同样点数量 (除 $n = 390$ 时) 的 GWRK 验证结果差异较小: MSE 介于 $-0.0096 \sim 0.0076 \text{ g kg}^{-1}$, ASE 介于 $0.2953 \sim 2.4750 \text{ g kg}^{-1}$, 均小于全部样点建模的 OK 精度, ME 介于 $-0.0096 \sim 0.0040 \text{ g kg}^{-1}$ (表 10), 表明 GWRK 结果存在全局高估现象, 当样点数量减少至 762 时, RMSE 先增大后下降, 不同采样模式的样点空间分布对预测精度的影响存在明显差异, 采用 GWRK 进行有机质预测时, 系统均匀格网采样模式下样点数对建模精度的影响较小, 表明较少样点数也可得到较好的制图结果, 样点空间分布对预测精度的影响随着样点数的减少而增强, 如 1379 个样点的预测精度高于 1656 个样点的结果。

与 OK 插值方法比较发现, 在样点数最多时, GWRK 建模精度提高幅度有限, 不同样点集、不同地形单元分区内样点的 GWRK 模型的调整决定系数 R^2 均大于相应的 OK 模型的调整系数, 较相应的模型增加 0.15 ~ 0.20 之间 (表 11、12), 表明对于地形单元内相同样点集, GWRK 模型对有机质的解释

能力大于空间插值模型。对于不同样点集有机质的 GWRK 模型, R^2 介于 0.85 ~ 0.53, 总体上随样点集数的减少而略微降低, 说明采用较少样点数进行 GWRK 建模时, 也可满足一定的精度需求 (如样点 $n = 1656$ 、 $n = 2806$ 、 $n = 1724$ 的模型拟合度较接近)。

3 结论与讨论

本文基于地形单元分区、系统格网多尺度采样、基于地形起伏度最佳统计单元三种采样模式, 采用土壤-地形关系模式确定不同采样点对土壤有机质预测的影响, 结果如下:

(1) 山脊、背坡、陡坡、坡脚等典型坡位单元内、地形统计平原区域内 ($n = 1656$)、系统格网尺度 G4 系列下 ($n = 1720$) 时, 土壤有机质实测范围同预测范围基本一致, 预测精度均优于全局地形单元土壤有机质预测。

(2) 基于土壤-地形关系的地形单元分区因结合有不同地貌单元的结构差异致使采样误差降低, 因此, 可采用有限合理样点实现有机质预测制图的小误差表达, 而对于样点数量较少的地形单元内合理样点的确定亟待进一步研究。

(3) 万荣县的土壤样点主要分布在耕地集中连片性强、耕地质量好的平原和坡位区域, 符合当地现状和国家相关规程要求。经有机质回归系数分析表明 6 ~ 7 个地形因子组合的效果能较好预测各个地形单元内有机质含量, 地形因子越多, 土壤-地形关系模型的 R^2 值越大。

(4) 在地形较复杂的地区, 土壤有机质具有较

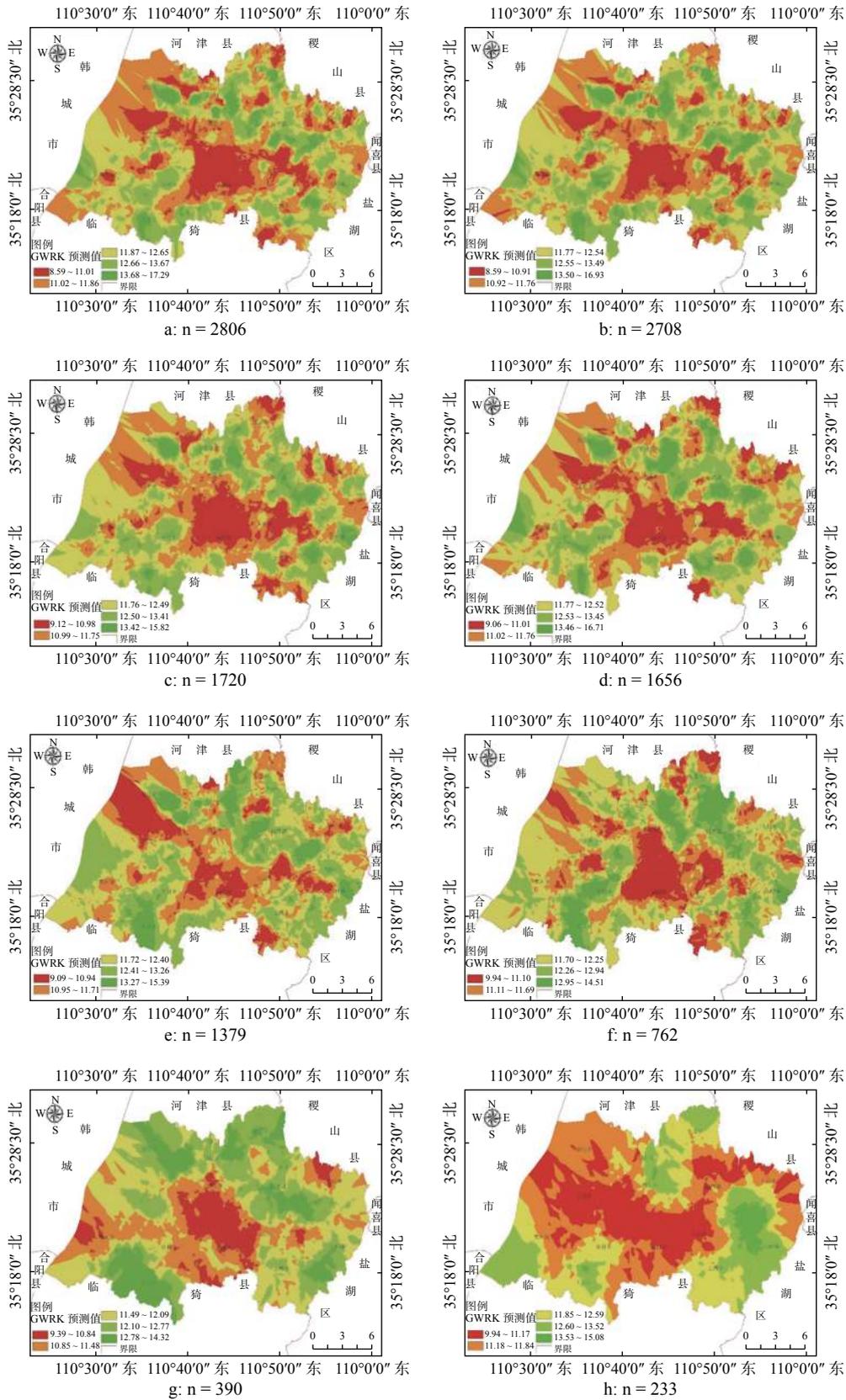


图 6 有机质空间分布预测图

Fig.6 Spatial distribution of predicted soil organic matter content

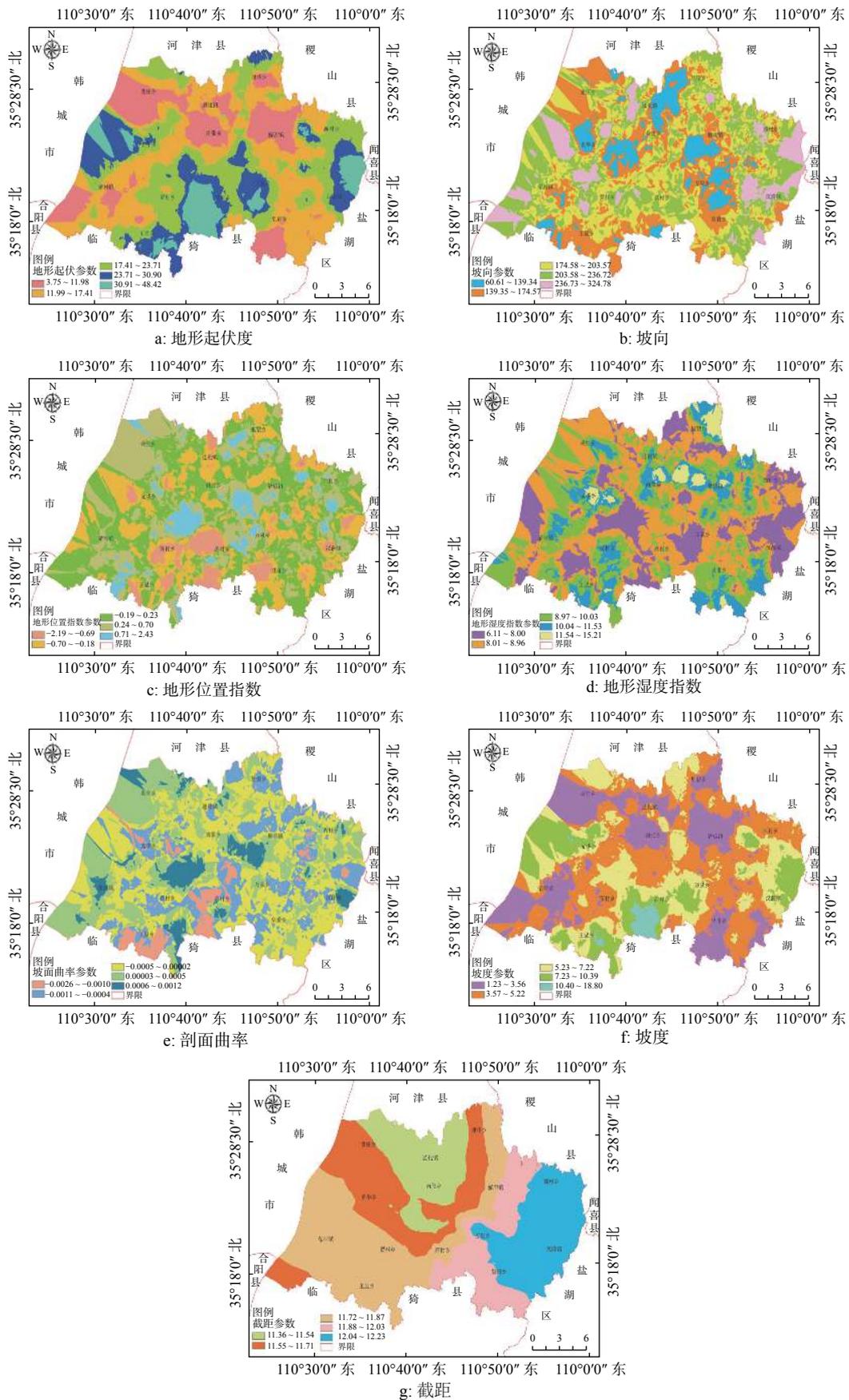


图 7 有机质回归系数空间分布图(n = 1656)

Fig.7 Spatial distribution of regression coefficient related to soil organic matter prediction (n = 1656)

表 10 不同样点集下有机质 GWRK 建模交叉验证结果
Table 10 Result of validation of GWRK model at different sampling series

	N = 2806	N = 2708	N = 1914	N = 1724	N = 1656	N = 1379	N = 762	N = 390	N = 233	N = 2806
MSE (g kg ⁻¹)	-0.0096	-0.0011	-0.0030	0.0014	-0.0001	-0.0015	-0.0031	0.0076	0.0070	0.0068
ASE (g/kg)	2.4715	2.4750	0.3123	2.4557	2.4711	2.5223	0.2953	0.4031	0.5704	2.8654
RMSE (g/kg)	2.3968	3.3950	-0.0030	2.4190	2.5120	2.6470	0.2904	0.4236	0.5658	3.9762
RMSSE (g/kg)	0.9711	0.9682	1.0160	0.9809	0.9760	0.9705	0.9842	1.0500	0.9821	0.9921
ME (g/kg)	-0.0096	-0.0026	-0.0006	0.0032	-0.0005	-0.0038	-0.0009	0.0006	0.0040	0.0056

注: MSE-标准预测误差, ASE-平均标准差, RMSE-均方根预测误差, RMSSE-均方根标准差, ME-平均预测误差

表 11 不同样点集有机质 GWRK 建模与 OK 结果比较
Table 11 Comparison between validation of GWRK and OK models at different sampling series

	N = 2806		N = 2708		N = 1724		N = 1656		N = 1379		N = 762		N = 390		N = 233	
	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK
AIC	245.6	1383.3	185.3	1125.6	134.7	923.3	110.9	768.3	94.8	500.3	88.6	470.1	75.6	397.9	58.9	298.39
R ²	0.85	0.61	0.82	0.60	0.81	0.60	0.79	0.59	0.75	0.43	0.73	0.44	0.66	0.42	0.53	0.39

注: AIC- Akaike information criterion (带宽), G-K: GWRK (地理加权回归克里格), 以下同。

表 12 不同地形单元内有机质 GWRK 建模与 OK 结果比较
Table 12 Comparison between validation of GWRK and OK models at different terrain units

	背坡 Back slope		沟谷 Channel		陡坡 Steep slope		坡脚 Footslope		坡肩 Slope shoulder		山脊 ridge		旱平地 Dry ground		湿平地 Wet ground		平顶峰 Flat peak	
	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK	G-K	OK
AIC	298.8	1140.2	150.34	780.37	241.8	350.2	421.6	503.5	400.6	490.8	99.34	525.9	411.3	491.6	364.8	1135.4	110.14	649.3
R ²	0.54	0.32	0.48	0.32	0.3	0.2	0.36	0.22	0.34	0.19	0.38	0.25	0.34	0.2	0.64	0.44	0.4	0.3

注: AIC- Akaike information criterion (带宽), G-K: GWRK (地理加权回归克里格), 以下同。

高的空间异质性, 可导致区域有机质预测存在较大的不确定性^[30]。在一样点分布较少的地形单元内合理样点数量的确定存在一定的误差。

(5) 基于地形选取样点时, 因受限于在该地形单元所采集的样点数, 故得到的结论仅针对确定的单元, 应用于其他单元需要仍需依据实际情况确定。同时, 系统格网采样方法中是否属于规则格网, 有待进一步深入研究。

参考文献:

- [1] 买买提·沙吾提, 塔西甫拉提·特依拜, 丁建丽, 张 飞 BP神经网络在渭河流域土壤盐渍化预测中的应用[J]. 新疆农业科学, 2013, 50(4): 774 - 779.
- [2] 栗丽华, 吉根林. 决策树分类技术研究[J]. 计算机工程, 2004, 30(9): 94 - 105.
- [3] 朱洪芬, 南 锋, 徐占军, 荆耀栋, 段永红, 毕如田. 黄土高原盆地土壤有机质与影响因子的空间多尺度关系[J]. 生态学报, 2017, 37(24): 8348 - 8360.
- [4] Zhu H, Bi R, Duan Y, et al. Scale-location specific relations between soil nutrients and topographic factors in the Fen River Basin, Chinese Loess Plateau[J]. Frontiers of Earth Science, 2017, 11(2): 397 - 406.
- [5] Zhu H, Hu W, Jing Y, et al. Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform[J]. Geoderma, 2018, 330: 9 - 18.
- [6] 巫振富, 赵彦锋, 程道全, 陈 杰. 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.
- [7] Tobler W. A computer movie simulating urban growth in the Detroit region[J]. Economic Geography, 1970, 46(2): 234 - 234.
- [8] Miller H J. Tobler's First Law and spatial analysis[J]. Annals of the Association of American Geographers, 2004, 94(2): 284 - 284.
- [9] 刘 瑜, 龚 俐, 童庆禧. 空间交互作用中的距离影响及定量分析[J]. 北京大学学报(自然科学版), 2014, 50(3): 526 - 534.
- [10] 李丽琴. 省际边界区域城市空间相互作用的时空演变[D]. 山西师范大学, 2018.
- [11] 陈彦光. 空间相互作用模型的形式、量纲和局域性问题探讨[J]. 北京大学学报: 自然科学版, 2009, 45(2): 333 - 333.
- [12] 孙 俊, 潘玉君, 和瑞芳, 刘海琴, 常楠静, 刘树芬, 李会仙. 地理学第一定律之争及其对地理学理论建设的启示[J]. 地理研究, 2012, 31(10): 1749 - 1763.
- [13] 赵业婷, 常庆瑞, 李志鹏等. 基于Co-Kriging的耕层土壤全氮空间特征及采样数量优化研究[J]. 土壤学报, 2014, (2): 415 - 422.
- [14] 张淑杰, 朱阿兴, 刘 京, 杨 琳. 整合已有土壤样点的数字土壤制图补样方案[J]. 地理科学进展, 2012, 31(10): 1318 - 1325.
- [15] 王子龙, 陈伟杰, 付 强等. 土壤优化采样策略研究进展[J]. 水土保持通报, 2017, (5): 211 - 218.
- [16] 秦承志, 卢岩君, 包黎莉, 朱阿兴, 邱维理, 程维明. 简化数字地形分析软件(SimDTA)及其应用-以嫩江流域鹤山农场区的坡位模糊分类为例[J]. 地球信息科学学报, 2009, 11(6): 737 - 743.
- [17] Zhao F, Qiao L, Shi F, et al. Feature fusion via hierarchical

- supervised local CCA for diagnosis of autism spectrum disorder[J]. *Brain Imaging and Behavior*, 2017, 11(4): 1 – 11.
- [18] 李文华. 社会调查研究中样本的代表性问题探讨[J]. *统计与决策*, 2006, (17): 157 – 159.
- [19] 杨琳, 朱阿兴, 张淑杰, 安艺明. 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[J]. *土壤学报*, 2015, 52(1): 28 – 37.
- [20] 杨琳, 朱阿兴, 秦承志, 李宝林, 裴 韬. 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): 46 – 54.
- [21] 张贝尔, 黄 标, 赵永存, 张晓光, 孙维侠, 胡文友, 杨劲松. 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[J]. *土壤*, 2013, 45(3): 540 – 547.
- [22] Franzen D W, Peck T R. Field soil sampling density for variable rate fertilization[J]. *Journal of Production Agriculture*, 1996, 8(4): 568 – 574.
- [23] 丁贤法. 基于SRTM DEM与变点分析法的云南省富宁县地貌形态研究[J]. *测绘与空间地理信息*, 2014, 37(11): 98 – 100.
- [24] 杨 晓, 黎 武, 冉 红, 贺春明, 徐 珍, 蒋 婷. 基于DEM数据的祁县地形起伏度分析[J]. *南方农业*, 2016, 10(22): 28 – 31.
- [25] 陈学兄, 张小军, 常庆瑞. 陕西省地形起伏度最佳计算单元研究[J]. *水土保持通报*, 2016, 36(3): 265 – 270+370.
- [26] 戚玉娇. 大兴安岭森林地上碳储量遥感估算与分析[D]. 哈尔滨: 东北林业大学, 2014.
- [27] ANSELIN L. Local Indicators of Spatial Association-LISA[J]. *Geographical Analysis*, 1995, 27(2): 93 – 115.
- [28] 孙孝林, 赵玉国, 赵 量, 李德成, 张甘霖. 应用土壤-景观定量模型预测土壤属性空间分布及制图[J]. *土壤*, 2008, (5): 837 – 842.
- [29] 中国科学院地理研究所. 中国 1 : 1000000地貌图制图规范(试行)[M]. 北京: 科学出版社, 1987: 33-34.
- [30] ZHU M, FENG Q, QIN Y, et al. The role of topography in shaping the spatial patterns of soil organic carbon[J]. *Catena*, 2019, 176: 296 – 305.

Spatial Prediction for Soil Organic Matter Affected by Appropriate Samples Based on Soil-Terrain Relationship

LI Guang-yue, BI Ru-tian*, ZHU Hong-fen, CHEN Zhuo

(College of Resource and Environment, Shanxi Agricultural University, Taigu 030801, China)

Abstract: For the problems of reasonable sampling point number and spatial prediction optimization in fine soil management system, this paper explored the possibility of finite sample points with different sampling methods based on soil-terrain relationship. Specially, the effects of spatial distribution of optimal sampling points on the prediction of soil organic matter (SOM) were studied based on the sampling methods of grid system, terrain unit partition expression and optimal statistical unit for relief degree, combined with the geo-statistics, soil-topography relationships, and geographically weighted regression kriging model. The results showed that: (1) The optimal unit size for topographic relief was 10×10 pixels in the plain, hilly and platform areas, and the spatial distributions of sampling points were mostly dense in the statistical units of plain area. The optimal sampling point numbers was 1656 based on the evaluation of standard samples. (2) The main influencing factors on the spatial variation of SOM were elevation, slope direction, topographic position, relative position, topographic relief, and the other topographic factors, which could explain 69.2% of the spatial variation of SOM in the study area. (3) The accuracy of GWRK model was greater than that of OK interpolation model based on different sampling points or terrain units. The optimal number of sampling points were 39, 481, 9, and 28 for the positions of mountain ridge, backslope (N/EN), steep slope (N/EN), and slope base, respectively. The improvement of GWRK model on SOM predictions was limited under the maximum sampling points ($n = 2806$). The local variation of the spatial distribution of the predicted SOM content was relatively decreased with the sampling point number decreasing. Therefore, the effect of reasonable sampling points on predict SOM content was significantly different among different sampling methods, while the prediction trend was similar among them. The mapping of predicted SOM contents could demonstrate the basic spatial distribution pattern of SOM.

Key words: Appropriate sampling; Terrain unit division; Geo-weighted regression model; Spatial pattern; Optimum statistics unit of relief amplitude.

[责任编辑: 孙福军]