

DOI: 10.11766/trxb202107290386

郝辰恺, 孙孝林, 王会利. 广义线性地统计模型在典型亚热带丘陵区数字土壤制图中的应用[J]. 土壤学报, 2023, 60 (4): 993–1006.
HAO Chenkai, SUN Xiaolin, WANG Huili. Application of Generalized Linear Geostatistical Model for Digital Soil Mapping in a Typical Subtropical Hilly Area[J]. Acta Pedologica Sinica, 2023, 60 (4): 993–1006.

广义线性地统计模型在典型亚热带丘陵区数字土壤制图中的应用^{*}

郝辰恺¹, 孙孝林^{1†}, 王会利²

(1. 中山大学地理科学与规划学院, 广州 510275; 2. 广西壮族自治区林业科学研究院, 南宁 530002)

摘要: 数字土壤制图在当前的应用越来越广泛, 其方法主要包括环境相关模型、空间自相关模型, 以及这些模型的混合模型。理论上, 混合模型相对单一模型具有明显的优势。广义线性地统计模型 (GLGM) 也是一种混合模型, 相对于最常用的混合模型——回归克里格 (RK), 又具有能加入随机效应来解决土壤变异的非平稳性等优势。然而, GLGM 因计算繁琐等缺点, 在国内外应用较少。本文以广西南宁高峰林场内一小面积 (3.03 km^2) 丘陵为研究区, 以 14 个地形因子为预测变量, 使用广义线性混合模型 (GLMM) 及其与普通克里格 (OK) 相结合的模型 (即 GLGM), 对土壤有机碳 (SOC)、pH、黏粒和阳离子交换量 (CEC) 的空间分布进行预测, 并与常用的多元线性回归 (MLR)、地理加权回归 (GWR)、回归森林 (RF)、OK、RK 和广义可加模型 (GAM) 进行比较。结果表明: GLMM 在预测黏粒上准确度较高; GLMM 和 GLGM 在预测 CEC 上准确度中等, 但在预测 SOC 和 pH 上准确度较低。综合线性回归模型的调整决定系数、块金效应和全局莫兰指数, 本文认为, 当土壤属性与环境变量具有较低的线性回归调整决定系数 (即小于 5%)、土壤属性具有较弱的空间自相关性 (即块金效应大于 71%) 和较强的局部空间变异 (即全局莫兰指数小于 0.09) 时, GLMM 和 GLGM 具有较高的适用性, 例如本文中的黏粒。反之, GLMM 和 GLGM 的适用性不好, 例如 SOC 和 pH。鉴于土壤空间变异的高度异质性和多尺度性, GLMM 和 GLGM 具有较好的应用前景。但是, 今后研究还需进一步探讨如何提高 GLMM 和 GLGM 的模拟效率。

关键词: 数字土壤制图; 土壤计量学; 广义线性地统计模型

中图分类号: K903 **文献标志码:** A

Application of Generalized Linear Geostatistical Model for Digital Soil Mapping in a Typical Subtropical Hilly Area

HAO Chenkai¹, SUN Xiaolin^{1†}, WANG Huili²

(1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; 2. Guangxi Forestry Research Institute, Nanning 530002, China)

Abstract: 【Objective】Digital Soil Mapping is receiving more attention and becoming widely used. Its methods mainly include

* 国家自然科学基金项目(42071062, 41771246)资助 Supported by the National Natural Science Foundation of China (Nos. 42071062, 41771246)

† 通讯作者 Corresponding author, E-mail: sun_xiaolin@yahoo.com

作者简介: 郝辰恺(1997—), 男, 广西桂林人, 硕士研究生, 主要从事数字土壤制图研究。E-mail: haochk@mail2.sysu.edu.cn

收稿日期: 2021-07-29; 收到修改稿日期: 2021-11-08; 网络首发日期 (www.cnki.net): 2022-03-21

environmental correlation-based models, spatial auto-correlation based models, and a mixture of these two kinds of models. The mixed model is expected to be advantageous over the single models. A generalized linear geostatistical model (GLGM) is a kind of mixed model. Compared with the commonly used mixed model, i.e., regression kriging (RK), GLGM has advantages such as having random effects to account for the non-stationarity of soil variability. However, GLGM is seldomly used due to its major disadvantages, i.e., complicated computation. 【Method】 In this study, within a small hilly area (3.03 km^2) in Gaofeng Forest of Nanning, Guangxi, generalized linear mixed model (GLMM) and its combination with ordinary kriging (OK), i.e., GLGM, were used to predict the spatial distribution of soil organic carbon (SOC), pH, clay and cation exchange capacity (CEC). Performances of the two models were then compared with commonly used models, including multivariable linear regression (MLR), geographically weighted regression (GWR), regression forest (RF), OK, RK and generalized additive model (GAM). 【Result】 The results showed that GLMM had higher accuracy in predicting clay, while GLMM and GLGM had medium accuracy in predicting CEC. Further, GLMM and GLGM had lower accuracy in predicting SOC and pH. 【Conclusion】 Based on the adjusted R^2 of the linear regression model, nugget effect and global Moran's I, it is concluded that GLMM and GLGM are appropriate when there is a low adjusted R^2 of linear soil-landscape regression (less than 5%), weak spatial auto-correlation of soil (nugget-to-sill ratio large than 71%), and strong local spatial variability of soil (Moran's I less than 0.09), e.g., clay in this paper. Otherwise, GLMM and GLGM are not appropriate, e.g., for SOC and pH in this paper. For the high spatial heterogeneity and multi-scale variability of soil, we think that GLMM and GLGM are promising for DSM, although more studies are needed to improve the efficiency of GLMM and GLGM modelling.

Key words: Digital soil mapping; Pedometrics; Generalized linear geostatistical model

数字土壤制图 (Digital soil mapping, DSM) 的应用越来越广泛。例如, Hengl 等^[1]建立了 SoilGrids 全球数字土壤图系统; Poggio 和 Gimona^[2]应用 DSM 估算土壤有机碳 (SOC) 储量; Dai 等^[3]将 DSM 产品应用于区域地面过程模拟。

DSM 的模型模拟方法主要有: 基于土壤发生学理论的环境相关模型; 基于地理学第一定律的空间自相关模型; 前两者的混合模型^[4]。环境相关模型包括多元线性回归 (Multivariable linear regression, MLR)、广义可加模型 (Generalized additive model, GAM)、判别分析、分类回归树 (Classification and regression tree, CART)、随机森林 (Random forest, RF)、人工神经网络 (Artificial neural network, ANN) 等; 空间自相关模型如反距离加权 (Inverse distance weighted, IDW) 和地统计学等; 混合模型主要是回归克里格 (Regression kriging, RK), 即 MLR 与普通克里格 (Ordinary kriging, OK) 的结合。

理论上, 混合模型由于同时将环境变量与土壤属性的空间自相关纳入考虑, 因此相对其他单一模型而言, 具有明显的优势。例如, 混合模型适用于复杂地区制图、可以降低制图成本等^[5]。很多研究表明混合模型的准确度较好。例如, 李启权等^[6]研究发现径向基函数神经网络 (Radial basis function

neural network model, RBFNN)-OK 混合模型和 RK 在预测精度上优于 OK; Ma 等^[7]研究发现决策树 (Cubist) -简单克里格 (Simple kriging, SK) 混合模型在较大尺度上准确度高于单一的决策树模型。然而, 也有很多研究表明, 混合模型的准确度不如单一模型^[5, 8]。这些研究进一步分析后认为, 混合模型的准确度与环境变量和土壤属性之间的关系、土壤属性空间自相关程度和样本间距等有关^[5]。但是, 这些因素对混合模型准确度的影响难以判断^[5, 8]。

广义线性地统计模型 (Generalized linear geostatistical model, GLGM) 也是一种混合模型。它是广义线性混合模型 (Generalized linear mixed model, GLMM) 与地统计学的混合, 由 Diggle 等^[9]提出。GLGM 与 RK 类似, 均对回归模型的残差项进行空间分析, 但与 RK 不同的是, 它包含了随机项, 并引入连接函数从而适应响应变量的不同随机分布形式。因此, GLGM 相对于广泛使用的 RK 而言具有一定的优势。最明显的优势是它通过随机项解决了土壤变异的非平稳性。尤其是当前, 土壤变异的多尺度性逐渐在 DSM 研究中受到重视^[10]。例如 Sun 等^[11]探索了多尺度土壤制图方法。另一个优势是 GLGM 可以适应响应变量的不同分布形式, 尽管这个优势在 DSM 中并不特别明显, 因为土壤属性

数据一般符合或经过转化后符合正态分布。综上, GLGM 在 DSM 中的应用具有一定的优势, 在一些研究中得到了应用, 例如 Kempen 等^[12]利用 GLGM 更新荷兰泥炭地区的土壤图; Poggio 和 Gimona^[2]利用与 GLGM 类似的 GAM-地统计混合三维模型 (Hybrid GAM-geostatistical 3D model) 模拟苏格兰的土壤有机碳垂直与横向分布。

另一方面, GLGM 也具有明显的劣势, 主要是形式复杂多样导致的计算繁琐问题, 以及参数和变异函数的最优选择等问题^[12]。一般研究人员要依赖较强的专业知识或者大量的计算, 从众多的环境变量和多样的模型组合 (即固定变量、随机变量、空间变异函数) 中筛选出最优的模型及其参数。因此, GLGM 较难使用, 在国内外的 DSM 应用中较少。截至目前, 还没有研究表明其准确度是否优于常用的 DSM 方法。

本文以广西南宁高峰林场林业示范基地内一小面积 (3.03 km^2) 丘陵为研究区, 利用地形因子, 建立 SOC、pH、黏粒和阳离子交换量 (CEC) 的 GLMM 及其与 OK 相结合的模型 (GLGM), 并与 MLR、GWR、回归森林 (Regression Forest, RF)、OK、RK 和 GAM 等常用的 DSM 方法进行准确度比较, 探讨产生准确度差异的原因, 评价 GLMM 和 GLGM 在 DSM 中的使用效果, 从而分析 GLMM 和 GLGM 在 DSM 中的适应性, 为提高 DSM 的准确度提供参考。

1 材料与方法

1.1 研究区概况

本文的研究区位于广西南宁高峰林场林业示范基地内, 经纬度范围为 $22^{\circ}57'8''$ — $22^{\circ}58'41''\text{N}$, $108^{\circ}20'57''$ — $108^{\circ}21'54''\text{E}$, 南北距离约为 2 820 m, 东西距离约为 1 710 m, 面积约为 3.03 km^2 。本区高程约为 126~301 m, 属于典型的丘陵地形 (图 1)。本区气候为亚热带湿润季风气候, 年平均气温约为 21.6°C , 年平均降水量约为 1 300 mm。土壤母质以古生代的泥岩、泥质页岩、砂页岩等沉积岩系的坡积物为主; 主要土壤类型为赤红壤, 相当于中国土壤系统分类中的强育湿润富铁土^[13]。本区自有记录以来一直用于林业, 近 20 年以种植桉树人工林为主。

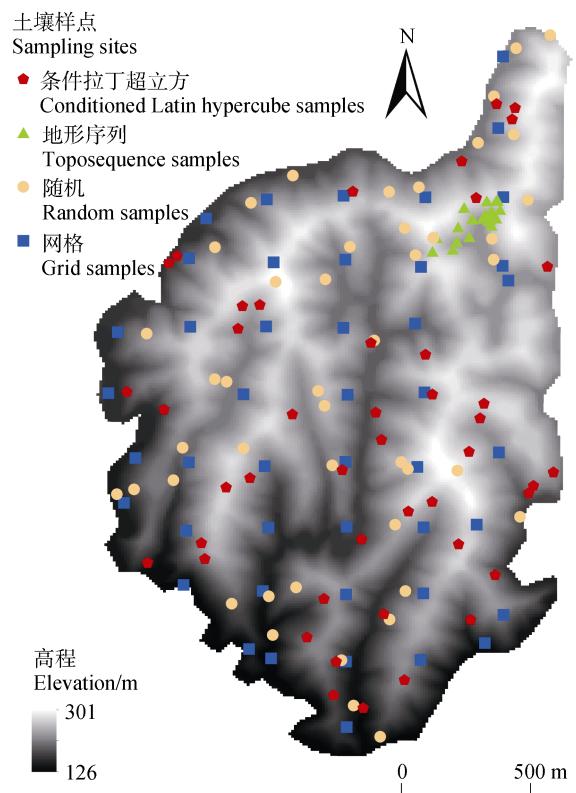


图 1 研究区的数字高程图与土壤样点分布
Fig. 1 Digital elevation map of the study area and distributions of sampling sites

1.2 土壤采样与测定

本文的土壤采样设计共使用了 4 种方法, 包括条件拉丁超立方抽样、网格采样、地形序列和随机采样。条件拉丁超立方抽样以六个地形因子: 高程、坡度、坡向、平面曲率、剖面曲率、地形湿度指数为条件变量, 共采集 45 个样点; 网格采样法以 250 m 为间距, 共采集 45 个样点; 地形序列采样是在研究区的一个山谷中, 沿与等高线相平、近似“Z”字形的道路上, 选择等高线凸和凹的顶点处采样, 共获得满足条件的 20 个地形序列样点; 随机采样法即从研究区所有位置中, 随机抽取 45 个样点。地形序列样点是在杨谦等^[14]的基础上增加的, 其余样点均相同。这些地形序列样点间隔较近, 有助于反映土壤属性在小范围内的变异, 因而有助于本文研究 GLGM。本文的土壤采样是在以样点为中心的 $1 \text{ m} \times 1 \text{ m}$ 正方形内, 取中心点和四个顶角上 0~20 cm 的土样, 充分混合后形成的土壤样本。这些土壤样本经过风干、研磨过筛后, 用重铬酸钾氧化法测定 SOC 含量; 用 pH 计 (1: 2.5 土水比) 测定 pH; 用比重计方法测定黏粒含量; 用交换法测定 CEC。本文以

条件拉丁超立方抽样、网格采样和地形序列三种采样方法所得的 110 个样点作为训练集, 以随机采样所得的 45 个样点作为验证集, 用于验证 DSM 准确度。

1.3 环境变量

本文的研究区面积较小, 植被均一, 气候和成土母质等环境要素也基本一致。因此, 本文仅使用地形因子作为环境变量进行制图。本文在 ArcGIS10.2 和 SAGA 软件中, 使用 10 m 分辨率的数字高程模型, 提取出 14 个地形因子, 包括: 高程 (m)、坡度 (%)、坡向 (°)、偏北度 (°)、偏东度 (°)、偏北指数、偏东指数、太阳辐射 ($10^3 \text{ KW} \cdot \text{m}^{-2}$)、平面曲率、剖面曲率、比汇水面积 (m^2)、地形湿度指数、径流强度指数和地形位置指数。

1.4 制图模型

本文共使用 8 种制图模型: MLR、GWR、RF、OK、RK、GAM、GLMM 和 GLGM。其中, MLR、GWR、RF、OK、RK 是 DSM 常用的模型, 已有很多研究对它们进行了详细介绍, 例如文献[15]。因此, 本文仅对其他三种模型在下文中进行详细介绍。在 MLR 模型中, 本文首先依据方差膨胀因子 (Variance inflation factor, VIF) 剔除具有共线性的环境变量, 再用逐步回归方法进行变量筛选和最优模型拟合, 即当模型中所有环境变量达到显著 ($P < 0.05$), 且赤池信息准则 (Akaike information criterion, AIC) 最小时, 为最优变量组合模型。在 OK 模型中, 本文采用残差最大似然法 (Residual

maximum likelihood, REML) 进行半方差函数模拟, 从球状、高斯和 Matérn 函数中选择 AIC 值最小的作为最优半方差函数。其中 Matérn 函数的 kappa 系数以 0.1 为间隔, 由 0.1 递增至 3。在 RK 模型中, 线性回归部分由 MLR 模型组成, 残差的半方差函数选择方法与 OK 相同, 整个模型使用 REML 方法模拟^[14]。

广义可加模型 (Generalized additive model, GAM) 由 Hastie 和 Tibshirani 提出^[16]。它是在广义线性模型 (Generalized linear model, GLM) 的基础上扩展而来, 用每个环境变量的非线性光滑函数取代 GLM 中具体的参数来表达相应响应变量的函数, 且光滑函数具有可加性。GAM 形式为:

$$g(E(y)) = \alpha + \sum_{j=1}^p f_j(x_j) + \varepsilon \quad (1)$$

式中, $E(y)$ 为响应变量 y 的期望值, $g()$ 为连接函数, α 为截距, $f_j(x_j)$ 表示第 j 个预测变量 x_j 的非线性光滑函数, ε 为随机误差。光滑函数可以是样条函数、loess 函数和核函数等。在实际研究中, 光滑函数多使用样条函数进行拟合。本文使用样条函数中广泛使用的三次样条函数进行拟合, 因为三次样条函数的光滑性和拟合效应均较好, 也能良好适应数据的变化趋势。三次样条函数是由一系列节点分隔开的分段连续三次多项式组成:

$$f_j(x_j) = \begin{cases} S_{j1}(x_j) = a_{j1}x_j^3 + b_{j1}x_j^2 + c_{j1}x_j + d_{j1}, t_1 \leq x_j \leq t_2 \\ S_{j2}(x_j) = a_{j2}x_j^3 + b_{j2}x_j^2 + c_{j2}x_j + d_{j2}, t_2 \leq x_j \leq t_3 \\ \dots \\ S_{j(n-1)}(x_j) = a_{j(n-1)}x_j^3 + b_{j(n-1)}x_j^2 + c_{j(n-1)}x_j + d_{j(n-1)}, t_{n-1} \leq x_j \leq t_n \end{cases} \quad (2)$$

式中, $f_j(x_j)$ 为第 j 个环境变量 x_j 在一系列递增序列 t_1, t_2, \dots, t_n 所划分的区间 $[t_1, t_2], [t_2, t_3], \dots, [t_{n-1}, t_n]$ 上的三次样条函数, $a_{j1}, a_{j2}, \dots, a_{j(n-1)}$ 、 $b_{j1}, b_{j2}, \dots, b_{j(n-1)}$ 、 $c_{j1}, c_{j2}, \dots, c_{j(n-1)}$ 、 $d_{j1}, d_{j2}, \dots, d_{j(n-1)}$ 为这些多项式中的待定系数, t_1, t_2, \dots, t_n 称为节点, $S_{j1}(x_j)$, $S_{j2}(x_j), \dots, S_{j(n-1)}(x_j)$ 为这些区间中的三次多项式。在模型拟合时, 三次样条函数要求这些位于不同区间的分段多项式在节点处保持连续, 同时也保证两个相邻的多项式在节点处的一阶导数和二阶导数连

续^[16]。本文将 MLR 模型中 VIF 筛选出的环境变量进行所有可能组合后, 用于 GAM 建模。为了选择合适的 n 值, 本文在模型拟合时, 对每个环境变量光滑项的 n 值由小到大进行了试验, 最后选取所有光滑项显著 ($P < 0.05$)、光滑度指数 (k-index) 大于 1 并小于 1.1, 且 AIC 值最小的模型为最优模型。

GLMM 在形式上是 MLR 与 GAM 结合而成, 其中 MLR 作为固定项表达全局变异, 而 GAM 作为随机项表达局部变异。GLGM 则是 GLMM 与 OK 的混合。两者的形式分别为:

$$g(E(y(u_0))) = \alpha + X^T \beta + \sum_{j=1}^p f_j(x_j) + \varepsilon \quad (3)$$

$$g(E(y(u_0))) = \alpha + X^T \beta + \sum_{j=1}^p f_j(x_j) + \sum_i \omega_i Z(u_i) + \varepsilon \quad (4)$$

式中, $g()$ 为连接函数, $y(u_0)$ 为待估点 u_0 上响应变量的值, $E(y(u_0))$ 为 $y(u_0)$ 的期望值, α 为截距, X 表示作为固定项的环境变量, β 为 X 的参数, $f_j(x_j)$ 表示作为随机项的环境变量, $Z(u_i)$ 为采样点 u_i 上的 GLMM 残差, ω_i 为采样点 u_i 上的权重 (由半方差函数估计后计算得到), ε 为随机误差。与 GAM 模型模拟一样, 本文采用 VIF 筛选后的环境变量进行所有可能的固定项与随机项的组合后, 进行 GLMM 模拟, 并选择 AIC 值最小的模型为最优模型。GLMM 残差项的 OK 建模与上述 OK 建立过程相同。

1.5 模型准确度评价

在验证集的 45 个土壤样点上, 计算土壤制图预测值的平均误差 (Mean error, ME)、平均绝对误差 (Mean absolute error, MAE)、均方根误差 (Root mean square error, RMSE)、一致性相关系数 (Lin's concordance correlation coefficient, CCC) 和拟合优度 (Goodness-of-fit, R^2), 用于评价土壤制图的准

确度。ME 越接近 0, MAE 和 RMSE 越小, 以及 CCC 和 R^2 越接近 1, 表明模型预测的准确度越高, 反之则越低。

2 结果与讨论

2.1 土壤属性描述性统计特征

表 1 列出了土壤属性测定结果的统计特征。通过比较训练集、验证集的统计特征可以看到, 各土壤属性在训练集和验证集上的统计特征相近, 表明本文所采用的训练集与验证集对本研究区的各土壤属性均具有较好的代表性。训练集与验证集的变异系数表明, 除 pH 属于弱变异外, SOC、黏粒和 CEC 均属中等变异。八分位偏度 (Octile skew) 结果表明训练集各土壤属性均符合正态分布。

2.2 模型模拟结果

2.2.1 MLR、GWR 和 RF 建模 SOC、pH、黏粒和 CEC 的 MLR 模型及其决定系数 (R^2) 和调整决定系数 (Adjusted R^2) 如表 2 所示。调整决定系数结果表明, MLR 模型分别能解释 SOC、pH、黏粒和 CEC 空间变异的 7.5%、19%、4.9% 和 18%, 说明各土壤属性与地形因子之间的线性关系不强, 以及 MLR 模型对各土壤属性空间变异的解释程度不高。在杨谦等^[14]的研究中, SOC、pH 和黏粒的

表 1 研究区土壤属性的描述性统计特征

Table 1 Statistics of soil properties in the study area

	最小值 Min	最大值 Max	平均值 Mean	中值 Median	标准差 Standard deviation	偏度 Skewness	峰度 Kurtosis	变异系数/% Coefficient of variation	八分位偏度 Octile skew
训练集 (样本数 $n=110$) Calibration dataset ($n=110$)									
SOC / (g·kg ⁻¹)	11.58	40.11	21.62	21.20	5.34	0.67	3.58	24.72	0.07
pH	3.95	5.10	4.41	4.41	0.20	0.49	3.80	4.55	-0.06
黏粒 Clay /%	28.00	52.00	38.74	39.00	4.56	-0.08	2.70	11.77	-0.01
CEC / (cmol·kg ⁻¹)	11.40	26.80	16.15	16.00	2.44	1.00	5.80	15.13	0.01
验证集 (样本数 $n=45$) Validation dataset ($n=45$)									
SOC / (g·kg ⁻¹)	10.72	31.64	20.35	19.63	4.18	0.55	3.56	20.54	0.38
pH	4.01	5.04	4.42	4.41	0.22	0.63	3.69	4.97	0.09
黏粒 Clay /%	25.80	52.00	37.26	36.80	6.39	0.38	2.53	17.16	0.05
CEC / (cmol·kg ⁻¹)	10.20	20.40	15.52	15.40	2.54	-0.13	2.26	16.34	0.02

表2 各土壤属性的MLR模型

Table 2 MLR models for soil properties

土壤属性 Soil property	MLR 方程 MLR formulae	决定系数 R^2	调整决定系数 Adjusted R^2
SOC / (g·kg ⁻¹)	SOC = 31.41 - 0.04 × 高程 + 83.88 × 剖面曲率	0.092**	0.075
pH	pH = 4.79 - 4.55 × 10 ⁻³ × 坡度 - 3.35 × 剖面曲率 - 4.30 × 平面曲率 + 5.81 × 10 ⁻⁶ × 径流强度指数 - 0.05 × 地形湿度指数	0.227***	0.189
黏粒 Clay / %	Clay = 38.55 - 1.79 × 偏北指数	0.058*	0.049
CEC / (cmol·kg ⁻¹)	CEC = 16.44 + 0.89 × 偏东指数 + 49.29 × 平面曲率	0.193***	0.178

注: * 表示显著性水平为 0.05, ** 表示显著性水平为 0.01, *** 表示显著性水平为 0.001。下同。Note: * represents significant at the 0.05 level, ** represents significant at the 0.01 level, *** represents significant at the 0.001 level. The same as below.

MLR 模型的调整决定系数分别为 2.0%、20% 和 9.8%。可见,由于本文多使用了 20 个地形序列样点进行模型模拟, SOC、pH 和黏粒的 MLR 模型的调整决定系数分别增强、略降低和降低。Lai 等^[17]的研究也表明,同一个研究区上不同数量的土壤采样点会影响模型的模拟结果及土壤制图的准确性。

使用各土壤属性 MLR 模型中的环境变量建立各自的 GWR 模型, 经最小 AIC 选择, 得到各自的 GWR 模型。模型的最优带宽分别处于 266~880 m、538~1 143 m、415~1 012 m、787~1 799 m 区间范围内。所有带宽均处于训练集样点最小间距(4 m)与最大间距(2 652 m)之间, 模型的 R^2 分别为 0.293 ($P<0.001$)、0.343 ($P<0.001$)、0.126 ($P<0.001$) 和 0.222 ($P<0.001$), 调整决定系数分别为 0.280、0.311、0.118 和 0.208。

使用所有环境变量进行各土壤属性的 RF 建模, 结果表明, SOC、pH、黏粒和 CEC 的 RF 模型的 R^2 分别为 0.564 ($P<0.001$)、0.540 ($P<0.001$)、0.536 ($P<0.001$) 和 0.542 ($P<0.001$), 调整决定系数分别为 0.499、0.471、0.466 和 0.475。

综上,对于所有土壤属性, MLR、GWR 和 RF 模型拟合的调整决定系数的排序均为 RF>GWR>MLR, 说明非线性模型拟合程度优于线性模型, 且局部回归的线性模型拟合程度优于全局回归的线性模型。这与郭澎涛等^[18]研究中 RF 的相关系数大于逐步线性回归(Stepwise linear regression, SLR)、赵明松等^[19]研究中 GWR 的调整决定系数大于 MLR 等的结果一致。

2.2.2 OK 和 RK 建模 SOC、pH、黏粒和 CEC 的半方差分析结果表明, 块金效应分别为 41.9%、

71.3%、100% 和 0%, 表明除黏粒外, 其他三种土壤属性分别具有中等、中等和很强的空间自相关性, 而黏粒不具有空间自相关性。本文中 SOC、pH 和黏粒的半方差模型的块金效应分别大于、小于和大于杨谦等^[14]研究(即分别为 19.3%、85.8% 和 91.1%)。如前所述, 这主要是因为本文较杨谦等^[14]多使用了 20 个地形序列采样点。在 Sun 等^[20]的研究中, 在同一区域内不同采样结果上建立的半方差模型也具有大小不同的块金效应。

RK 的半方差分析结果表明, SOC、pH、黏粒和 CEC 经过 MLR 模拟之后, 残差的块金效应分别为 44%、80%、100% 和 0%, 分别表明具有中等、很弱、无和很强的空间自相关性。其中 SOC 与 pH 的 RK 模型块金效应相对于 OK 均有提高, 表明 SOC 与 pH 的空间变异适合 MLR 和地统计的混合模型; CEC 的 RK 块金效应与 OK 相比无变化, 表明 CEC 的空间变异不适合使用 MLR 和地统计的混合模型模拟。

2.2.3 GAM 建模 GAM 的模型模拟结果如表 3 所示。其中 CEC 的 GAM 模型中存在 3 个自由度为 1 的光滑项, 即 s (偏东指数)、 s (太阳辐射) 和 s (平面曲率)(表 3), 表明这 3 个光滑项近似于线性项。图 2 以 SOC 的 GAM 模型中部分光滑项为例, 反映了土壤属性与环境变量的非线性光滑项趋势。在小于 180 m 处, SOC 与高程呈负相关关系, 可能是因为: 在山脚海拔较低处, 积水较多, 植物生长茂盛, 归还量高, 因而 SOC 含量一般较高; 随着海拔升高, 积水减少, 森林经营活动较为剧烈, 植物生长略差, 归还量降低, 因而 SOC 含量降低; 随着海拔继续升高至 180 m~200 m 处, SOC 与高程呈

表 3 各土壤属性的 GAM 模型

Table 3 GAM models for soil properties

土壤属性 Soil property	GAM 方程 GAM formulae	决定系数 R^2	调整决定系数 Adjusted R^2
SOC / (g·kg ⁻¹)	$21.62 + s(\text{高程}, 6.07) + s(\text{坡度}, 11.34) + s(\text{偏东度}, 12.73) + s(\text{剖面曲率}, 6.32)$	0.625***	0.610
pH	$4.41 + s(\text{高程}, 3.04) + s(\text{偏北度}, 4.97) + s(\text{剖面曲率}, 1.64) + s(\text{地形湿度指数}, 4.73)$	0.422***	0.400
黏粒 Clay /%	$38.74 + s(\text{偏北指数}, 13.01)$	0.273***	0.266
CEC / (cmol·kg ⁻¹)	$16.15 + s(\text{坡度}, 12.94) + s(\text{偏东指数}, 1.00) + s(\text{太阳辐射}, 1.00) + s(\text{剖面曲率}, 6.24) + s(\text{平面曲率}, 1.00) + s(\text{径流强度指数}, 12.27)$	0.631***	0.610

注: $s()$ 表示 GAM 模型中的光滑项。括号内数字表示该光滑项的估计自由度, 接近 1 时表明曲线形状接近直线, 值越大表明曲线形状的波动越大。Note: $s()$ represents smooth terms within GAM formulae. Numbers in the bracket represent the estimated degree of freedom (EDF) of the smooth term. When EDF is close to 1, it indicates that the curve is close to a straight line. The larger the EDF, the more flexible the curve.

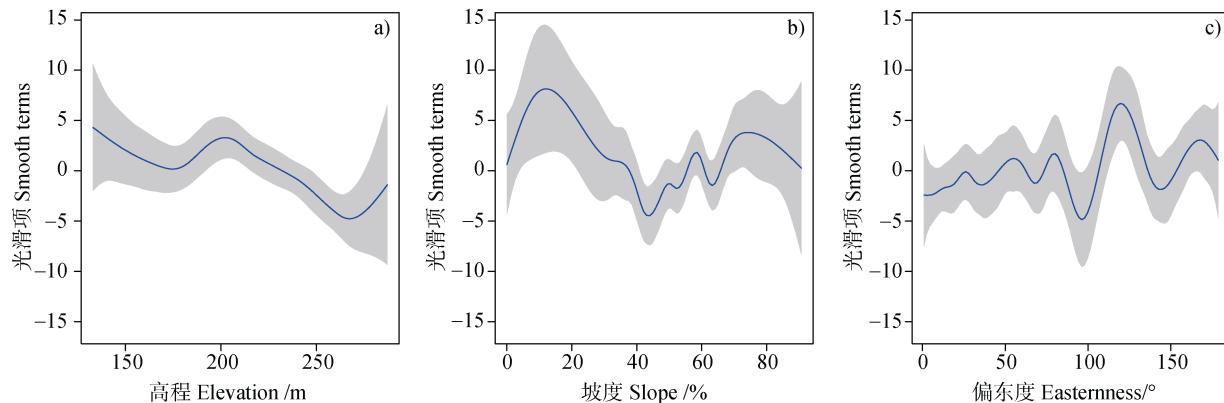


图 2 SOC 的 GAM 模型中部分光滑项. 图 a-c 的估计自由度分别为 6.07、11.34、12.73, 见表 3; 阴影部分表示光滑项估计的 ± 2 倍标准差范围.

Fig. 2 Some smooth terms of GAM for SOC. The degree of freedom in a-c are 6.07, 11.34 and 12.73, respectively, as shown in Table 3. The shaded area represents two times of standard deviations.

正相关关系, 可能是因为: 森林经营活动逐渐减少, 使土壤遭受的侵蚀降低, 因而 SOC 含量有所提升; 随着海拔继续升高至 200 m~260 m 处, SOC 与高程呈负相关关系, 可能是因为: 随着海拔提升, 汇水面积逐渐减小, 土壤水分较低, 植物生长较差, 归还量降低, 因而 SOC 含量降低; 随着海拔继续升高至大于 260 m 处, SOC 与高程呈正相关关系, 可能是因为: 海拔再升高时, 林下草本植物逐渐增加, 因而有利于 SOC 的累积^[21]。表 3 中的调整决定系数表明 GAM 模型分别能解释 SOC、pH、黏粒和 CEC 空间变异的 61%、40%、27%、61%, 均大于对应的 MLR 模型。这表明 GAM 模型中各土壤属性与地形

因子的非线性拟合较好。

2.2.4 GLMM 和 GLGM 建模

GLMM 的模型模拟结果如表 4 所示。调整决定系数表明 GLMM 模型分别能解释 SOC、pH、黏粒和 CEC 空间变异的 40%、51%、30%、67%, 均大于对应的 MLR 模型。并且, 除 SOC 之外, GLMM 的调整决定系数均大于 GAM 模型。SOC 的 GLMM 模型与 GAM 模型具有相同的环境变量, 但 GAM 模型存在 2 个自由度>10 的环境变量, 而 GLMM 模型中不存在自由度>10 的环境变量, 且有 1 个环境变量为固定项。因此, SOC 的 GLMM 模型拟合程度较 GAM 低, 在训练集上的调整决定系数较低。

表 4 各土壤属性的 GLMM 模型

Table 4 GLMM models for soil properties

土壤属性 Soil property	GLMM 方程 GLMM formulae	R^2	决定系数	调整决定系数
			Adjusted R^2	
SOC / (g·kg ⁻¹)	$19.99 + s(\text{高程}, 5.88) + s(\text{坡度}, 7.62) + s(\text{剖面曲率}, 5.13) + 0.02 \times \text{偏东度}$	0.422***	0.399	
pH	$4.59 + s(\text{高程}, 3.64) + s(\text{偏北度}, 4.87) + s(\text{剖面曲率}, 6.27) + s(\text{地形湿度指数}, 6.74) - 3.48 \times 10^{-3} \times \text{坡度}$	0.532***	0.509	
黏粒 Clay /%	$32.20 + s(\text{高程}, 3.68) + s(\text{坡向}, 9.42) - 58.39 \times \text{偏北指数} - 66.60 \times \text{剖面曲率}$	0.328***	0.302	
CEC / (cmol·kg ⁻¹)	$16.24 + s(\text{坡度}, 13.14) + s(\text{坡向}, 4.56) + s(\text{剖面曲率}, 6.45) + s(\text{径流强度指数}, 12.52) + 41.68 \times \text{平面曲率}$	0.682***	0.667	

注: $s()$ 表示 GLMM 模型中的光滑项。括号内数字表示该光滑项的估计自由度, 接近 1 时表明曲线形状接近直线, 值越大表明曲线形状的波动越大。Note: $s()$ represents smooth terms within GLMM formulae. Numbers in the bracket represent the estimated degree of freedom (EDF) of the smooth term. When EDF is close to 1, it indicates that the curve is close to a straight line. The larger the EDF, the more flexible the curve.

综上, 对于所有土壤属性的 MLR、GWR、RF、GAM 和 GLMM 模型, 整体而言 GLMM 的调整决定系数最大, 其次是 RF、GAM 和 GWR, MLR 最小, 表示线性与非线性结合的模型拟合程度最高, 其次是纯非线性模型, 线性模型拟合程度最低。

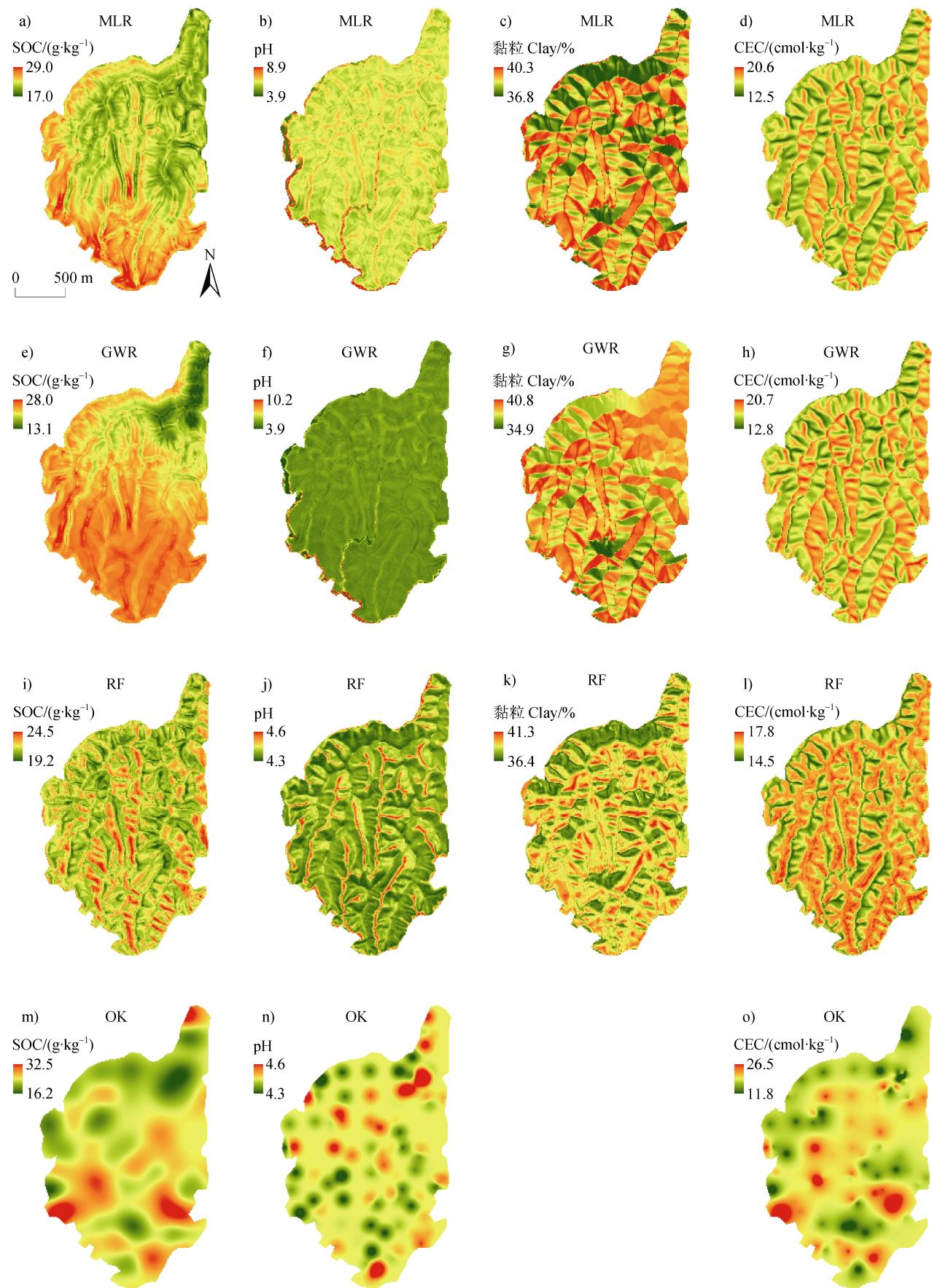
如前文所述, GLGM 为 GLMM 趋势项加上 GLMM 残差的 OK 插值。由于黏粒在 GLMM 模拟后残差的空间自相关性很弱, 只能建立纯块金效应的半方差函数, 因此不能建立黏粒的 GLGM 模型。其他三种土壤属性的 GLGM 模型的块金效应分别为 84%、96% 和 4.0%, 表明这三种土壤属性在 GLMM 模拟之后的残差分别具有很弱、很弱和很强的空间自相关性。而且, 这三种土壤属性的 GLGM 模型的块金效应均较 RK 和 OK 高, 表明 GLMM 更适合模拟这些土壤属性变异。

2.3 土壤制图与模型验证

上述模型预测 SOC、pH、黏粒和 CEC 的空间分布图如图 3 所示。MLR 由于使用了土壤属性与地形因子之间的线性关系, 因而预测图直接反映了研究区中所使用的地形因子分布。例如, SOC 的 MLR 方程中使用了高程(表 2), 因而预测图与 DEM 相似(图 3a)。由于 pH 的 MLR 模型中使用了较多的地形因子(表 2), 因而其预测图对地形因子分布的反映较不明显(图 3b)。GWR 使用的变量与 MLR 一致, 且建模原理相似, 因此各土壤属性的 GWR 预测图(图 3e~图 3h)与 MLR 模型的较为相似(图

3a~图 3d), 但在高低值区域有较为集中的趋势, 且空间邻近值之间的变化较为平滑。RF 的预测图也显示出了地形分布, 但由于使用了所有地形因子的非线性项, 因此 RF 预测图与地形因子分布的差别较 MLR 大, 并且细节更丰富, 值域范围相对于其他模型较窄(图 3i~图 3l)。OK 的预测图中, 值的变化较 GWR 更为平滑, 且高、低值分别呈集中环状分布, 反映了土壤属性分布的大致趋势(图 3m~图 3o)。RK 由于结合了 MLR 与 OK, 因此既反映了地形因子分布细节, 又反映了土壤属性空间相关性结构的大致分布格局(图 3p~图 3r)。

GAM 由于使用了非线性光滑项, 因此相较于其他模型细节更丰富, 更详细地反映了土壤属性的局部变异(图 3s~图 3v)。由于 pH 与 CEC 的 GAM 模型使用了自由度接近于 1 的项, 因此二者的 GAM 模型预测图均显示了地形信息。反之, SOC 与黏粒的 GAM 预测图对地形信息的反映较不明显。GLMM 与 GAM 建模方式类似, 因此两者的预测图也类似。但由于 SOC 和 pH 的 GLMM 与 GAM 模型相近, 而黏粒和 CEC 的 GLMM 与 GAM 模型差别较大, 因此 SOC 和 pH 的 GLMM 与 GAM 的预测图更相近, 而黏粒和 CEC 的 GLMM 与 GAM 预测图差别较大。此外, 由于 SOC、pH 的 GLGM 中块金效应较高, 而 CEC 的 GLGM 中块金效应较低, 因而 SOC、pH 的 GLMM 与 GLGM 的预测图差别不大, 而 CEC 的 GLMM 与 GLGM 预测图则差别较大。



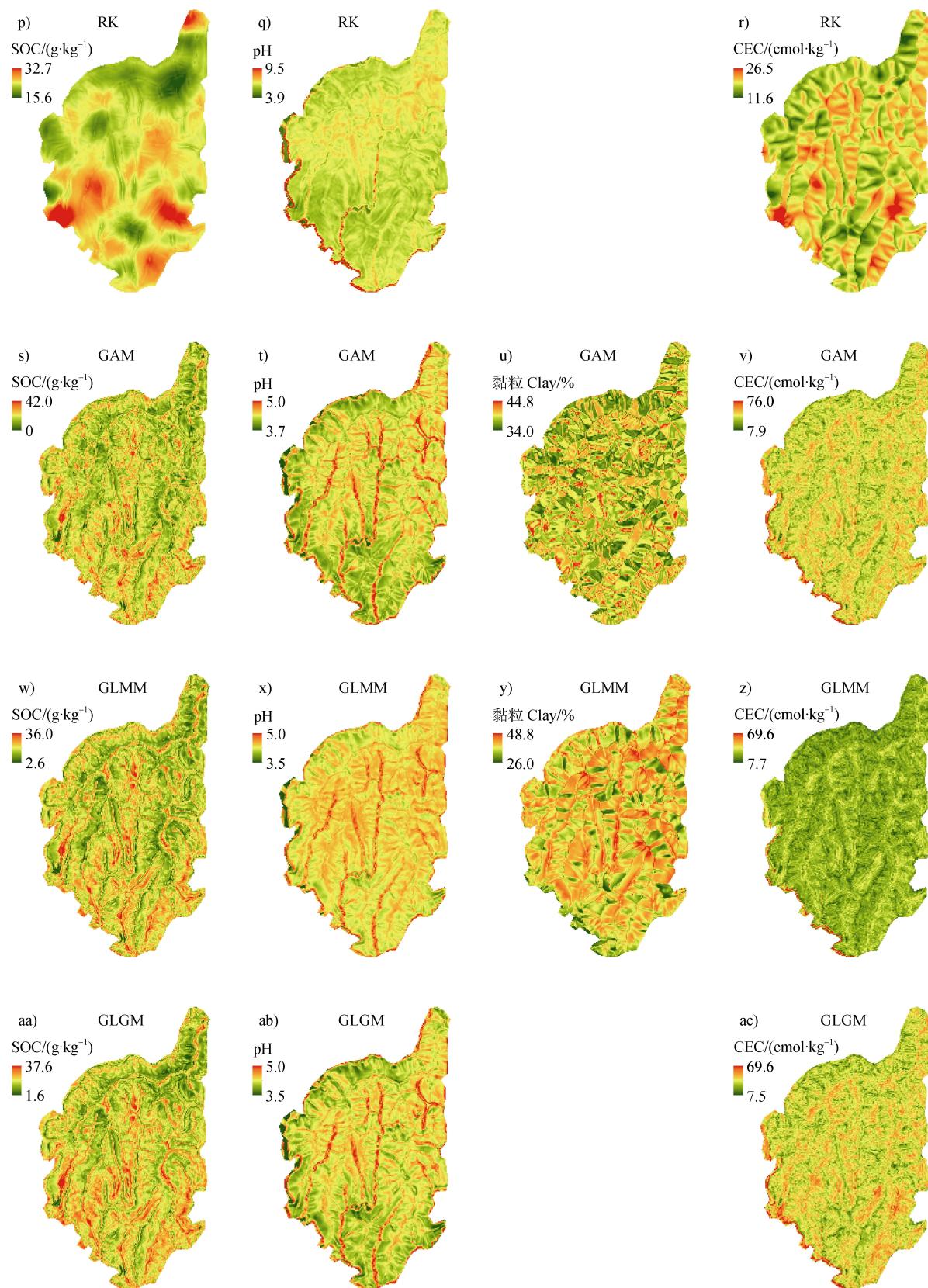


图3 基于不同模型模拟的各土壤属性的空间分布图: (a~d) MLR; (e~h) GWR; (i~l) RF; (m~o) OK; (p~r) RK; (s~v) GAM; (w~z) GLMM; (aa~ac) GLGM

Fig. 3 Spatial distribution of soil properties simulated using different models: (a~d) MLR; (e~h) GWR; (i~l) RF; (m~o) OK; (p~r) RK; (s~v) GAM; (w~z) GLMM; (aa~ac) GLGM

基于 45 个随机验证样点的预测结果准确性检验结果如表 5 所示。各指标均没有一致地表明哪种方法最优。以 SOC 为例, RMSE 表明, 最优模型为

RF, 最差模型为 GAM, 但 CCC 表明, 最优模型为 OK, 最差模型为 MLR。根据 Khaledian 和 Miller^[22] 分析, RMSE 与 ME 和 MAE 相似, 而 CCC 包含了

表 5 不同模型预测验证集土壤属性的准确度

Table 5 Accuracy of soil property prediction models based on the validation dataset

Soil property	Model	ME	MAE	RMSE	CCC	R^2
SOC / (g·kg ⁻¹)	多元线性回归 (MLR)	1.36	3.73	4.86	-0.11	0.035
	地理加权回归 (GWR)	1.20	4.00	5.05	-0.099	0.016
	回归森林 (RF)	1.51	3.56	4.34	0.059	0.030
	普通克里格 (OK)	1.32	3.68	4.62	0.086	0.013
	回归克里格 (RK)	1.29	3.60	4.63	0.058	0.0063
	广义可加模型 (GAM)	1.91	4.51	5.81	-0.0053	3.65×10^{-5}
	广义线性混合模型 (GLMM)	1.94	4.42	5.52	-0.041	0.0025
pH	广义线性地统计模型 (GLGM)	1.97	4.54	5.72	-0.021	5.75×10^{-4}
	多元线性回归 (MLR)	-0.012	0.159	0.21	0.137	0.046
	地理加权回归 (GWR)	-0.0096	0.163	0.221	0.10	0.023
	回归森林 (RF)	0.0018	0.170	0.222	-0.012	0.0013
	普通克里格 (OK)	-0.026	0.168	0.220	0.0022	8.98×10^{-5}
	回归克里格 (RK)	-0.0153	0.15	0.218	0.139	0.039
	广义可加模型 (GAM)	0.0084	0.18	0.25	0.031	0.0012
黏粒 Clay /%	广义线性混合模型 (GLMM)	0.0148	0.1960	0.2637	0.0034	1.34×10^{-5}
	广义线性地统计模型 (GLGM)	0.013	0.1958	0.2642	0.0041	1.93×10^{-5}
	多元线性回归 (MLR)	1.39	5.25	6.23	0.097	0.10
	地理加权回归 (GWR)	1.44	5.19	6.21	0.109	0.11
	回归森林 (RF)	1.54	5.26	6.20	0.111	0.17
	普通克里格 (OK)					
	回归克里格 (RK)					
CEC / (cmol·kg ⁻¹)	广义可加模型 (GAM)	1.70	5.62	6.60	0.088	0.023
	广义线性混合模型 (GLMM)	1.48	4.89	5.73	0.35	0.24
	广义线性地统计模型 (GLGM)					
	多元线性回归 (MLR)	0.97	2.111	2.703	0.13	0.043
	地理加权回归 (GWR)	0.99	2.105	2.702	0.14	0.047
	回归森林 (RF)	0.81	2.016	2.49	0.15	0.126
	普通克里格 (OK)	0.59	1.86	2.41	0.23	0.134

R^2 。然而 ME 中正误差和负误差会相互抵消，使误差变小；MAE 中所有残差的权重相等； R^2 不能捕捉模型预测中的偏差。因此本文综合使用 RMSE 和 CCC 作为主要的准确度指标进行分析。这两个指标表明，预测 SOC 准确度最好的模型是 RF 和 OK，其次是 RK，最差是 MLR、GWR、GAM、GLMM、GLGM；预测 pH 准确度最好的模型是 MLR 和 RK，其次是 GWR，再次是 OK 和 GAM，最差是 RF、GLMM、GLGM；在黏粒上，排序为 GLMM>RF>GWR>MLR>GAM；在 CEC 上，排序为 OK>RK、GLMM、GLGM>RF>GAM>GWR>MLR。这些结果表明，GLMM 与 GLGM 在黏粒上表现较好，在 CEC 上表现中等，在 SOC、pH 上表现较差，而 GAM 整体表现均较差。

为了分析预测制图准确度的影响因素，图 4 展示了各土壤属性在模型模拟中的调整决定系数和块金效应，以及主要的准确度指标——RMSE 和 CCC。MLR 模型的调整决定系数均在 0.2 以下，说明这些土壤属性的空间变异仅有少部分能由全局单一的线性模型解释。因此，MLR 模型的准确度一般较差，

仅在 pH 上因为块金效应较大（即较低的空间自相关性）且调整决定系数也较大时表现出较好的准确度。OK 模型的块金效应表明，SOC 和 CEC 具有较好的空间自相关性，而 pH 的空间自相关性较差。因此，OK 在 SOC 和 CEC 上的准确度较好，而在 pH 上的准确度较差。同理，可以总结出其他模型在不同土壤属性上准确度较好或较差的原因。例如，RF 在 SOC 上的决定系数较高，因而其在 SOC 上的准确度较高。

然而，上述规律也并不能完全解释所有结果。例如，GAM 在四个土壤属性上均具有较高的调整决定系数，但其准确度总是较差。进一步分析认为，这些模型的准确度差异还与模型自身的特点有关。尤其是 GWR、GAM、GLMM 和 GLGM 等四个模型并不是全局模型，而是局部模型，即描述局部土壤空间变异的模型。这些模型一般具有较好的决定系数（GLGM 具有与 GLMM 一样的决定系数），但它们并不一定具有较好的准确度。这是因为这些模型容易过度拟合，导致模型模拟结果实际上并不准确。例如，GAM 将所有变量视为随机项，因而非常容易

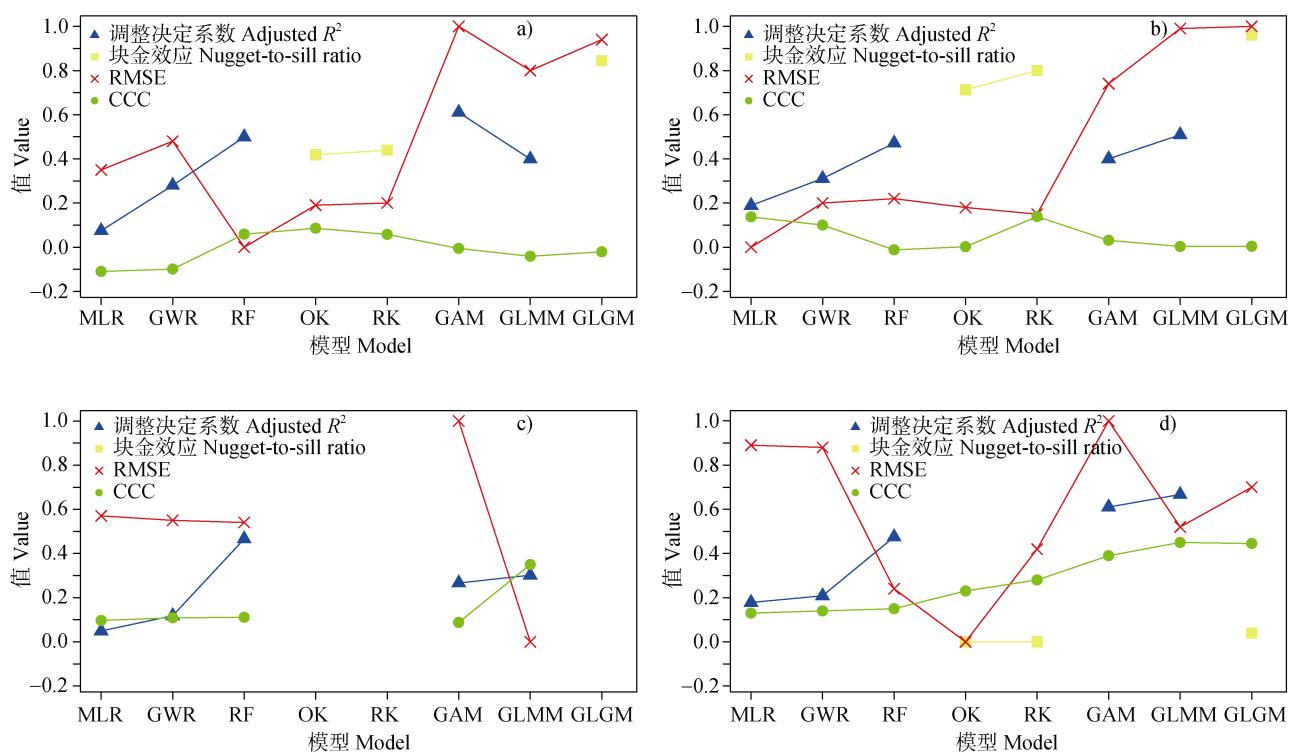


图 4 各土壤属性模型拟合的调整决定系数和块金效应与主要的准确度指标：(a) SOC；(b) pH；(c) 黏粒；(d) CEC。
RMSE 经 0~1 规范化转换

Fig. 4 Accuracy indices with fitting adjusted R^2 and the nugget-to-sill ratio for soil properties: (a) SOC, (b) pH, (c) clay, (d) CEC. RMSE has been standardized by MinMax normalization

过度拟合而使模型结果失真。为此,本文计算了各土壤属性的全局莫兰指数^[23],用以反映土壤属性的局部变异强弱。SOC 的莫兰指数为 0.31(极显著); pH 为 0.09(不显著); 黏粒为 -0.004(不显著); CEC 为 0.18(极显著)。这些结果表明: SOC、CEC 和 pH 的局部变异性较弱,而黏粒具有较强的局部变异性。因此,GAM 虽然在 SOC 上具有最高的调整决定系数,但这可能是由于模型过度拟合,因而预测准确度实际较差。反之, GLMM、GLGM 不仅用随机项来描述局部变异,同时也用固定项来描述全局变异,因而具有一定的优势,例如, GLMM 在黏粒预测上具有最好的准确度。

综合以上分析,本文认为当土壤属性与环境变量具有较低的线性回归调整决定系数(约小于 5%),且土壤属性具有较弱的空间自相关性(块金效应约大于 71%)和较强的局部空间变异(全局莫兰指数约小于 0.09)时, GLMM 及其与地统计的混合模型——GLGM 具有较高的适用性。对比本文中的多个模型,本文还认为,在选用模型进行 DSM 时,应根据土壤属性的实际变异情况来选用合适的模型。例如,Sun 等^[5]在大量模拟数据的基础上,系统比较了 OK 和 RK,也认为当模型能比较真实地反映土壤空间变异时,才表现出较好的预测准确度。

尽管本文初步分析得到了 GLMM 和 GLGM 在何种情况下准确度较好,但是使二者准确度高的线性回归调整决定系数、块金效应和莫兰指数等指标的具体定量范围仍有待进一步研究。而且本文的研究区面积较小,而在较大区域尺度中,由于气候、植被、母质等成土因素的尺度和作用强度不同,土壤属性的变异包含了多个尺度,以及不同的强度。因此, GLMM 和 GLGM 在较大区域尺度中可能具有较好的适用性。值得一提的是,本文研究区的地形为丘陵,土壤属性变异较为明显,也有较为丰富的地形变量用于 GLMM 和 GLGM。在地形平缓区,土壤属性的变异比较平缓,而可用于 DSM 的环境变量并不多^[24]。因此, GLMM 和 GLGM 在大面积平缓区的应用还有待深入分析。此外,目前还缺乏高效地为 GLMM 和 GLGM 筛选合适的环境变量作为固定项和随机项的方法。因此,本文在建立 GLMM 和 GLGM 模型时,运用全变量组合方法来海量寻找最优的 GLMM 和 GLGM 模型,未来需进一步研究如何高效地为 GLMM 和 GLGM 筛选合适的环境变

量作为固定项和随机项。

3 结 论

本文以广西南宁高峰林场林业示范基地内的小面积典型丘陵为研究区,以 14 个地形因子为预测变量,比较了 GLMM、GLGM 与 DSM 常用模型(即 MLR、GWR、RF、OK 和 RK)及 GAM 模型在 SOC、pH、黏粒和 CEC 空间分布预测制图上的准确性。结果表明,当土壤属性与环境变量具有较低的线性回归调整决定系数(约小于 5%),且土壤属性具有较弱的空间自相关性(块金效应约大于 71%)和较强的局部空间变异(全局莫兰指数约小于 0.09)时, GLMM 和 GLGM 具有较高的适用性。同时, GLMM 和 GLGM 预测的土壤图较其他常规模型具有更丰富的细节,更详细地反映了土壤属性的局部变异。鉴于土壤空间变异的高度异质性和多尺度性,本文认为 GLMM 和 GLGM 具有较好的应用前景,尤其是在包含了不同尺度、不同强度变异的较大区域上。这将为 DSM 研究中模型方法的选取提供有价值的参考。此外, GLMM 和 GLGM 在模型模拟时的效率很低,未来还需研究如何高效地为 GLMM 和 GLGM 筛选合适的环境变量作为固定项和随机项,以提高效率。

参考文献 (References)

- [1] Hengl T, Mendes de Jesus J, Heuvelink G B M, et al. Soil Grids 250m: Global gridded soil information based on machine learning[J]. PLoS One, 2017, 12(2): e0169748.
- [2] Poggio L, Gimona A. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation—An example from Scotland[J]. Geoderma, 2014, 232/233/234: 284—299.
- [3] Dai Y J, Shangguan W, Wei N, et al. A review of the global soil property maps for Earth system models[J]. Soil, 2019, 5 (2): 137—158.
- [4] Sun X L, Zhao Y G, Liu F, et al. Digital soil mapping and advance in research[J]. Chinese Journal of Soil Science, 2013, 44 (3): 752—759. [孙孝林,赵玉国,刘峰,等.数字土壤制图及其研究进展[J].土壤通报,2013,44(3): 752—759.]
- [5] Sun X L, Yang Q, Wang H L, et al. Can regression determination, nugget-to-sill ratio and sampling spacing determine relative performance of regression kriging over ordinary Kriging?[J]. Catena, 2019, 181: 104092.
- [6] Li Q Q, Wang C Q, Zhang W J, et al. Prediction of soil

- nutrients spatial distribution based on neural network model combined with geostatistics[J]. Chinese Journal of Applied Ecology, 2013, 24 (2): 459—466. [李启权, 王昌全, 张文江, 等. 基于神经网络模型和地统计学方法的土壤养分空间分布预测[J]. 应用生态学报, 2013, 24 (2): 459—466.]
- [7] Ma Y X, Minasny B, Wu C F. Mapping key soil properties to support agricultural production in Eastern China[J]. Geoderma Regional, 2017, 10: 144—153.
- [8] Keskin H, Grunwald S. Regression kriging as a workhorse in the digital soil mapper's toolbox[J]. Geoderma, 2018, 326: 22—41.
- [9] Diggle P J, Tawn J A, Moyeed R A. Model-based geostatistics[J]. Journal of the Royal Statistical Society: Series C: Applied Statistics, 2002, 47 (3): 299—350.
- [10] Zhang G L, Zhu A X, Shi Z, et al. Progress and future prospect of soil geography[J]. Progress in Geography, 2018, 37 (1): 57—65. [张甘霖, 朱阿兴, 史舟, 等. 土壤地理学的进展与展望[J]. 地理科学进展, 2018, 37 (1): 57—65.]
- [11] Sun X L, Wang Y D, Wang H L, et al. Digital soil mapping based on empirical mode decomposition components of environmental covariates[J]. European Journal of Soil Science, 2019, 70 (6): 1109—1127.
- [12] Kempen B, Brus D J, Heuvelink G B M. Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland[J]. Geoderma, 2012, 189/190: 540—553.
- [13] Chinese Soil Taxonomy Research Group, Institute of Soil Science, Chinese Academy of Sciences, Cooperative Research Group on Chinese Soil Taxonomy. Keys to Chinese Soil Taxonomy[M]. 3rd ed. Hefei: University of Science and Technology of China Press, 2001. [中国科学院南京土壤研究所土壤系统分类课题组, 中国土壤系统分类课题研究协作组. 中国土壤系统分类检索[M]. 第3版. 合肥: 中国科学技术出版社, 2001.]
- [14] Yang Q, Wang X Q, Sun X L, et al. Comparing prediction accuracies of ordinary kriging and regression kriging with REML in soil properties mapping[J]. Chinese Journal of Soil Science, 2018, 49 (2): 283—292. [杨谦, 王晓晴, 孙孝林, 等. 基于 REML 的普通克里格和回归克里格在土壤属性空间预测中的比较[J]. 土壤通报, 2018, 49 (2): 283—292.]
- [15] Somarathna P D S N, Minasny B, Malone B P. More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon[J]. Soil Science Society of America Journal, 2017, 81 (6): 1413—1426.
- [16] Wood S N. Generalized Additive Models: An Introduction with R[M]. 2nd ed. Boca Raton: Chapman and Hall/CRC, 2017: 1—476.
- [17] Lai Y Q, Wang H L, Sun X L. A comparison of importance of modelling method and sample size for mapping soil organic matter in Guangdong, China[J]. Ecological Indicators, 2021, 126: 107618.
- [18] Guo P T, Li M F, Luo W, et al. Prediction of soil total nitrogen for rubber plantation at regional scale based on environmental variables and random forest approach[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31 (5): 194—200, 201. [郭澎涛, 李茂芬, 罗微, 等. 基于多源环境变量和随机森林的橡胶园土壤全氮含量预测[J]. 农业工程学报, 2015, 31 (5): 194—200, 201.]
- [19] Zhao M S, Liu B Y, Lu H L, et al. Spatial modeling of soil organic matter over low relief areas based on geographically weighted regression[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35 (20): 102—110. [赵明松, 刘斌寅, 卢宏亮, 等. 基于地理加权回归的地形平缓区土壤有机质空间建模[J]. 农业工程学报, 2019, 35 (20): 102—110.]
- [20] Sun X L, Wang H L, Forristal D, et al. Limited spatial transferability of the relationships between kriging variance and soil sampling spacing in some grasslands of Ireland: Implications for sampling design[J]. Pedosphere, 2019, 29 (5): 577—589.
- [21] Fan X H, Wang D C, Sun X L, et al. Relationships between soil organic carbon density of *Eucalyptus* plantations and terrain factors in Nanning[J]. Acta Ecologica Sinica, 2016, 36 (13): 4074—4080. [范晓晖, 王德彩, 孙孝林, 等. 南宁市桉树人工林土壤有机碳密度与地形因子的关系[J]. 生态学报, 2016, 36 (13): 4074—4080.]
- [22] Khaledian Y, Miller B A. Selecting appropriate machine learning methods for digital soil mapping[J]. Applied Mathematical Modelling, 2020, 81: 401—418.
- [23] Cliff A D, Ord J K. Spatial process: Models and Applications[M]. London: Plon, 1981: 1—266.
- [24] Liu F, Rossiter D G, Song X D, et al. An approach for broad-scale predictive soil properties mapping in low-relief areas based on responses to solar radiation[J]. Soil Science Society of America Journal, 2020, 84 (1): 144—162.

(责任编辑: 檀满枝)