

典型柑橘种植区土壤有机质空间分布与含量预测

段丽君 张海涛 郭 龙 杜佩颖 陈 可 琚清兰

华中农业大学资源与环境学院, 武汉 430070

摘要 以湖北省宜都市红花套镇典型柑橘种植区采集到的 329 个土壤样本为研究对象, 设置土壤有机质 (SOM) 进行普通克里格 (OK) 插值的结果为参照, 借地理探测器选取与 SOM 相关性最大的前 5 种主要影响因素, 分别建立全局模型多元线性回归、偏最小二乘回归和局部模型地理加权回归 (GWR), 再深入分析模型残差的结构, 构造 GWR 扩展模型 GWR_{MLR} 、 GWR_{PLSR} , 讨论几种 SOM 预测模型的差异。结果表明: 使用 GWR_{PLSR} 模型预测研究区 SOM 含量的均方误差和均方根误差可分别降低到 9.834 和 3.136, 相对分析误差提高到 1.468, 实测值与预测值间的相关系数 (r) 达 0.743, 具有最高的预测精度, GWR_{MLR} 其次, 说明除 SOM 与主要影响因子间存在空间相关性, 分析模型残差可进一步消除预测的不平稳性。因此, 将模型残差项纳入考虑的局部扩展模型更适宜进行区域化 SOM 空间分布预测与数字土壤制图。

关键词 土壤有机质; 空间分层异质性; 地理探测器; 模型残差; GWR_{MLR} ; GWR_{PLSR}

中图分类号 S 159 **文献标识码** A **文章编号** 1000-2421(2019)01-0073-09

土壤有机质 (soil organic matter, SOM) 既是土壤中各种营养元素的重要来源^[1], 也是陆地生态系统碳循环的主要源与汇^[2-3]。由于 SOM 能促进土壤结构形成、改善土壤物理性质, 其含量成为衡量土壤质量和肥力的指标之一^[4-5]。湖北省是柑橘种植大省, 但目前全省柑橘园总体有机质含量偏低^[6], 直接影响柑橘树生长及柑橘产量。因此, 准确掌握区域内的 SOM 空间分布信息是控制柑橘园合理施肥的重要前提。

研究表明, 许多方法可有效借鉴或应用于区域尺度上 SOM 含量的预测及数字制图。多元线性回归 (multiple linear regression, MLR) 是早期最常用的方法之一, 它选取主要影响因子作为辅助变量建立 SOM 预测模型。黄安等^[7] 研究指出, MLR 与克里格法的预测结果在宏观上具有一致的空间分布趋势, 但 MLR 的前提假设是全局辅助变量对因变量作用相同, 忽略了空间非平稳性, 有可能产生误导信息^[8-9]。偏最小二乘回归 (partial least squares regression, PLSR) 作为集中了主成分分析、典型相关分析和线性相关分析等优势的非空间回归模型, 经常被用于解决影响因子间冗余和多重共线性问

题^[10]。于雷等^[11] 使用 PLSR 法对高光谱数据降维处理后选取了能最大程度揭示有机质含量变化的主控因子。侯艳军等^[12] 更是在比较了多种区域 SOM 高光谱估算模型后, 得出 PLSR 法稳健性和适应性优于 MLR 法的结论。然而, 由于应用过程中未考虑变量的空间结构性, PLSR 法在 SOM 预测上仍具一定局限^[13]。地理加权回归 (geographical weighted regression, GWR) 则是解释空间非平稳性的局部空间统计模型^[14], 与全局模型的最大不同之处是它可依据距离衰退公式在不同地理位置上赋予影响因子以变化系数, 在土壤和环境科学中有广泛的应用前景^[15-16]。杨顺华等^[17] 对比纳入环境因子的 GWR 模型和 OLS 模型, 发现 GWR 法对 SOM 预测的结果与实测值的拟合程度更好。Kumar 等^[18] 也通过对美国俄亥俄州的 SOM 估值与制图证明 GWR 在提高插值精度、捕捉空间非平稳性方面具有优越性。当然, GWR 的模型预测残差还可能与辅助变量间存在某些未解释的相关性^[19], 如果对其作进一步分析并将结果当成空间随机的一部分参与到预测趋势估计中, 可能会反映出更加真实的 SOM 空间分布情况。

收稿日期: 2018-01-12

基金项目: 国家自然科学基金面上项目 (41371227)

段丽君, 硕士研究生。研究方向: 地统计应用与数字国土。E-mail: duanlijun@webmail.hzau.edu.cn

通信作者: 张海涛, 博士, 副教授。研究方向: 土壤空间过程与模拟。E-mail: hzau_zht@163.com

本研究通过地理探测器选取影响 SOM 空间分布的主要影响因子,以空间非平稳性为出发点,比较全局模型(MLR、PLSR)和局部模型(GWR)在有机质含量预测方面的差异,尝试深入分析 GWR 模型残差,建立局部空间回归与回归残差线性插值相结合的 GWR 扩展模型(GWR_{MLR} 、 GWR_{PLSR}),以期提高区域尺度的 SOM 含量预测精度,从而对养分调控、合理施肥乃至指导农业生产布局提供科学依据和数据支持。

1 材料与方 法

1.1 研究区概况

研究区位于湖北省宜都市红花套镇(111°14'13"~111°26'38"E, 30°26'45"~30°35'59"N),

是鄂西山区向江汉平原过渡的典型柑橘种植区。该镇属湿润季风气候,雨热同期,四季分明,年均气温 16.8℃,年均降雨量 1 200 mm,全年无霜期 275 d。研究区内西北地势高,以丘陵山地为主,东南地势低,以冲积平原为主。土地利用类型中,园地、林地、耕地、其他用地分别占研究区总面积的 72.46%、0.59%、0.13%和 26.82%。土壤类型多样,以黄棕壤、水稻土为主,也分布有少量潮土、紫色土和石灰土。

1.2 数据采集与提取

通过 1:10 000 比例尺的红花套镇地形图和全国第二次土壤普查数据,设定代表红花套镇典型柑橘种植区地形地势、土地利用格局和土壤类型的研究区域(图 1)。在样区内布置采样点时,总体上采

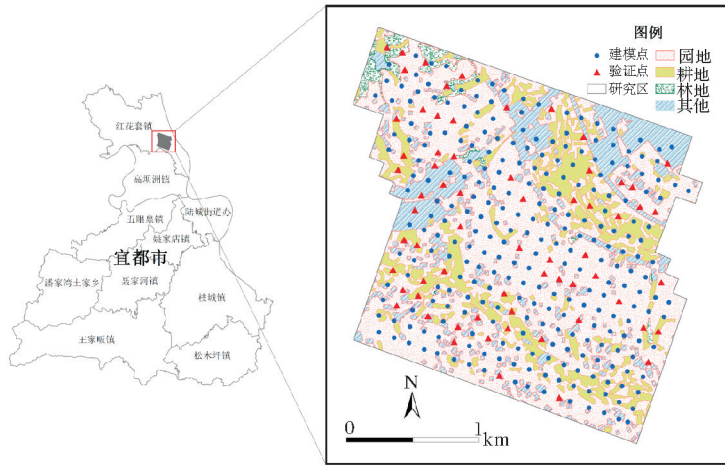


图 1 研究区概况及土壤样点分布

Fig.1 Study area and distribution of soil samples

用密度为 120 m×120 m 的规则网格选取采样中心,地势不平坦地块适当增加布设密度,在每个样点中心按 10 m 间距的十字交叉法采集混合土样 1.0 kg。使用差分式全球定位系统(DGPS)记录中心点位的经纬度及高程,共获得 329 个表层土样本。取回土样后对其风干、磨碎、过筛,留作化学分析。

由于 SOM 含量分布受其他土壤属性影响,所以根据研究区土壤养分分布状况及主要农作物所需土壤营养元素等实际情况,参考文献[20]的实验方法测定包括土壤有机质(SOM)在内的 14 种土壤属性数据:酸碱度(pH)、阳离子交换量(CEC)、全氮(TN)、全磷(TP)、全钾(TK)、全铜(TC)、全铅(Pb)、碱解氮(AN)、速效磷(AP)、有效铁(AI)、有效锌(AZ)、交换性钙(EC)、交换性镁(EM)。

SOM 含量分布还与地形因素有关^[21]。因此,

从地理空间数据云下载 GDEMDEM 的 30 M 分辨率 DEM 数据中,提取基本的和衍生的土壤地形数据共 14 种:坡度(slope)、坡向(aspect)、坡度变率(SOS)、坡向变率(SOA)、曲率(curvature)、平面曲率(plane-curvature)、剖面曲率(profile-curvature)、地形起伏度(relief)、地形粗糙度(roughness)、地表切割深度(surface cutting depth)、高程变异系数(elevation variation coefficient)、地形湿度指数(TWI)、汇流动力指数(SPI)、沉积物运移指数(STI)。

使用 SPSS 软件进行 3 倍标准差法剔除异常值,再通过 ArcMap 地统计工具中的子要素集功能对样点数据集进行均匀随机划分。最终获得建模集数据 263 个(80%),验证集数据 66 个(20%)。

1.3 地理探测器

地理探测器(GeoDetector)是探测空间分层异质性以及揭示其背后驱动因子的一种新的统计学方法^[22],它假设地理事物总是存在于特定的空间位置,影响其变化的因子在空间上具有差异性,若某影响因子与地理事物的变化在空间上具有显著一致性,则该影响因子对地理事物的发生与发展具有决定意义^[23]。由于该模型受到假设条件的制约较少^[24-25],因此,广泛应用于地理学科研究中。地理探测器包括风险因子探测、交互探测、风险区探测以及生态探测这4部分,其中风险因子探测可通过 q 统计量识别自变量中的影响因子对因变量的解释程度判断二者间的相关性大小,从而选取出主要辅助变量参与因变量的预测建模。本研究使用地理探测器中的风险因子探测寻找影响SOM含量变化的主要影响因子:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (1)$$

式(1)中, q 为因变量的空间异质性, L 为自变量的分层总数, h 为自变量所在层, N_h 和 N 分别为第 h 层和全层的单元数, σ_h^2 和 σ^2 分别为第 h 层和全层的因变量方差。 q 的值域为 $[0,1]$,值越大说明因变量的空间分异性越强,反之空间分布的随机性越强。如果分层是由某自变量生成,则 q 值越大表示该自变量对因变量的影响越大,反之越弱。当 $q=0$ 时,表示研究对象不存在空间异质性;当 $q=1$ 时,表示自变量完全控制了因变量的空间分布。

1.4 MLR模型和PLSR模型

MLR可通过设置逐步线性回归,建立因变量与多个自变量间剔除多重共线性冗余后的线性关系。PLSR则是在尽可能提取更多自变量信息且不剔除样本点的基础上,保证自变量与因变量间的最大相关性。两个模型的表达式分别为:

$$Y_{\text{MLR}}(S_0) = \sum_{k=1}^p \beta_k(S_0) \cdot X_k(S_0) \quad (2)$$

$$Y_{\text{PLSR}}(S_0) = \sum_{k=1}^p \beta_k(S_0) \cdot X_k(S_0) \quad (3)$$

式(2)、(3)中, $Y_{\text{MLR}}(S_0)$ 为在位置 S_0 处的MLR预测值, p 为自变量的个数, $X_k(S_0)$ 为第 k 个自变量在位置 S_0 处的实测值, β_k 为对应第 k 个自变量的全局回归系数。 $Y_{\text{PLSR}}(S_0)$ 为在位置 S_0 处的PLSR预测值,其余参数同理,不再赘述。

1.5 GWR模型

GWR是将数据的空间位置嵌入到回归参数

中,利用局部加权最小二乘法进行逐点参数估计。通过各地空间位置上的参数估计值随地理空间位置的变化情况,可直观探测空间关系的非平稳性^[26]:

$$Y_{\text{GWR}}(S_0) = \beta_0 + \sum_{k=1}^p \beta_k(S_0) \cdot X_k(S_0) + \epsilon \quad (4)$$

式(4)中, $Y_{\text{GWR}}(S_0)$ 为在位置 S_0 处的GWR预测值, β_0 为方程的常数项, p 为自变量的个数, $X_k(S_0)$ 为第 k 个自变量在位置 S_0 处的实测值, $\beta_k(S_0)$ 为对应第 k 个自变量在位置 S_0 处的局部变化回归系数, ϵ 为模型的随机误差。

1.6 GWR_{MLR}模型和GWR_{PLSR}模型

GWR_{MLR}和GWR_{PLSR}是GWR的扩展模型,它在考虑GWR局部地理变化系数的同时,对模型残差分别进行MLR和PLSR建模并赋予其全局固定系数。此模型理论上既可利用空间分异性进一步缩小预测值与实测值间的差异,又能同时实现自变量对因变量和残差的解释目的。

$$Y_{\text{GWRMLR}}(S_0) = Y_{\text{GWR}}(S_0) + \epsilon_{\text{MLR}}(S_0) \quad (5)$$

$$Y_{\text{GWRPLSR}}(S_0) = Y_{\text{GWR}}(S_0) + \epsilon_{\text{PLSR}}(S_0) \quad (6)$$

式(5)、(6)中, $Y_{\text{GWRMLR}}(S_0)$ 和 $Y_{\text{GWRPLSR}}(S_0)$ 分别是位置 S_0 处的GWR_{MLR}和GWR_{PLSR}预测值, $Y_{\text{GWR}}(S_0)$ 是式(4)中经GWR插值获得的预测值,在该模型中作为趋势项部分出现, $\epsilon_{\text{MLR}}(S_0)$ 和 $\epsilon_{\text{PLSR}}(S_0)$ 分别代表经MLR和PLSR建模的模型残差。

1.7 模型验证

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \{y_i - y_i^*\}^2} \quad (7)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \quad (8)$$

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(y_i^* - \bar{y}_i^*)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (y_i^* - \bar{y}_i^*)^2}} \quad (9)$$

式(7)~(9)中, n 代表样本数, y_i 和 y_i^* 分别代表样本的实测值和预测值, \bar{y}_i 和 \bar{y}_i^* 分别代表样本实测平均值和预测平均值。RMSE是均方根误差,作为插值精度的量度越小说明方法越精确;RPD是相对分析误差,依据样本标准差和均方根误差的比值判断模型预测能力。RPD <1.4 时模型无法对样品进行预测;1.4 \leq RPD ≤ 2 时,模型效果一般,可以用来对样品进行粗略评估;RPD >2 时模型具有极好的预测能力。 r 是取值在 $-1\sim 1$ 之间的Pearson相关系数,当 $0 < |r| < 1$ 时,表明 y_i 和 y_i^* 之间存在

一定的线性相关关系。若 $r > 0$, 表示 X 和 Y 正相关; 若 $r < 0$ 时, 表明 X 和 Y 负相关。

2 结果与分析

2.1 土壤有机质含量基本统计

由土壤有机质含量的基本统计特征(表 1)可知, 全部数据集的有机质含量均值为 16.970 g/kg, 范围为 6.070~29.180 g/kg, 最大和最小值间差异

表 1 土壤有机质含量的基本统计特征

Table 1 Basic statistical characteristics of soil organic matter content

g/kg

数据集 Dataset	样本数量 Size	平均值 Mean	最小值 Min	最大值 Max	标准差 SD	变异系数/% CV	正态分布检验 K-S
全部数据 All	329	16.970	6.070	29.180	4.221	24.873	0.800
建模集 Calibration	263	17.237	7.180	29.180	4.086	23.705	0.949
验证集 Validation	66	15.909	6.070	28.670	4.603	28.933	0.960

中, 前 5 种影响因子的解释力(q 统计量)由高到低依次为: 碱解氮(0.574)、有效铁(0.226)、交换性镁(0.111)、全氮(0.086)、阳离子交换量(0.034), 而其他因子 q 值太低, 不作为主要影响因子参与后续建模。同时, 交互探测的结果显示这 5 种主要影响因子交互作用后 q 值均呈现非线性增强, 说明主要影响因子间的交互作用都大于各单因子对 SOM 含量的影响。其中, SOM 是土壤氮素的重要来源, 有机质含量高的区域, 氮素转化为有机态并保蓄在土壤中的量就多, 因此, 碱解氮和全氮与 SOM 的空间变

显著且变异系数为 24.87%, 均属中等变异, 说明 SOM 在该研究区范围内具有一定的变异性, 适合空间局部估计。建模集中的有机质数据通过 K-S (Kolmogorov-Smirnov Z) 检验, 渐进显著性 P 值为 0.949, 符合正态分布, 可直接用于插值建模。

2.2 主要影响因子提取

研究区中与 SOM 含量相关性较大的主要影响因子及其 q 值如表 2 所示。风险因子探测的结果

异规律大致相似^[27]。SOM 与土壤有效铁的关系密切, 是由于有机质中多种酸性功能团解离使负电荷增多, 能吸附铁离子等金属阳离子, 故对铁离子具有较强的络合和富集能力^[28]。SOM 则与镁螯合形成富啡酸镁, 溶解性较大, 易遭淋失将镁带到土壤下层, 使得表层土壤交换性镁含量降低, 导致 SOM 与交换性镁两者间呈一定的负相关关系^[29]。此外, SOM 对胶体阳离子交换量的相对贡献率较高, 还能够间接协助反映土壤保蓄、供应和缓冲阳离子养分的能力^[30]。

表 2 研究区主要影响因子风险探测及交互作用

Table 2 Risk detection and interaction of major influence factors in the study area

主要影响因子 Main influence factors	因子探测 Factor detector		交互探测 Interaction detector			
	q 统计量 q statistic	碱解氮 AN	有效铁 AI	交换性镁 EM	全氮 TN	阳离子交换量 CEC
碱解氮 AN	0.574					
有效铁 AI	0.226	0.633				
交换性镁 EM	0.111	0.593	0.290			
全氮 TN	0.086	0.618	0.339	0.186		
阳离子交换量 CEC	0.034	0.599	0.309	0.145	0.148	

2.3 土壤有机质预测模型构建

利用主要影响因子对有机质含量进行 MLR 建模, 结果见表 3。其中, t 值是主要影响因子影响显著性的统计指标, 碱解氮($t = 17.838$)、有效铁($t = 5.645$)和常数项($t = 3.739$)的绝对值较大, 表明通过相关性检验进行模型拟合因子筛选能够保证它们对模型有较大的贡献率。各方差膨胀因子(VIF)

均小于 7.5, 说明模型中没有冗余影响因子信息且变量间不存在局部共线性。

选择主要影响因子对 SOM 含量进行 PLSR 建模(表 4)。从表 4 可以看出, 随主成分数量增加, X 和 Y 的累计百分比不断变大, 均方根误差逐渐缩小, 决定系数显著增大。在同时使用 5 种主要影响因子参与建模时, 均方根误差降低到 3.295, 决定系

表 3 多元线性回归模型预测参数统计

Table 3 Prediction parameter statistics of multivariate linear regression model

变量 Variables	系数 Coefficients	标准误差 Standard error	t 值 t-value	方差膨胀因子 VIF
截距 Intercept	4.804	1.285	3.739	—
碱解氮/(mg/kg) AN	0.129	0.007	17.838	1.351
有效铁/(mg/kg) AI	0.003	0.000	5.645	1.260
交换性镁/(cmol/kg) EM	-0.083	0.199	-0.418	1.596
全氮/(g/kg) TN	1.860	0.651	2.858	1.112
阳离子交换量/(cmol/kg) CEC	-0.092	0.087	-1.047	1.434

表 4 偏最小二乘回归模型预测参数统计

Table 4 Prediction parameter statistics of partial least squares regression model

主成分数量 PC	X 累计量/%	Y 累计量/%	均方根误差	决定系数 R_p^2
	Cumulative percent of X	Cumulative percent of Y	RMSE _p	
1	99.41	76.40	8.512	0.144
2	100.00	98.08	4.020	0.476
3	100.00	98.25	3.678	0.475
4	100.00	98.30	3.691	0.459
5	100.00	98.30	3.295	0.537

注: X 累计百分比表示主要环境因子参与建模的贡献度, Y 累计百分比表示能用模型获得准确因变量预测值的比例。Note: The cumulative percentage of X indicates the contribution of main environmental factors to the modeling, and the cumulative percentage of Y indicates the proportion of the predicted value of the accurate dependent variable that can be obtained by the model.

数升高到 0.537, 获得最好预测效果。因此, 本研究的最佳 PLSR 模型需使用碱解氮、有效铁、交换性镁、全氮及阳离子交换量共同建立。

使用主要影响因子对有机质含量进行 GWR 建模, 得到相关统计参数(表 5)。与只能概括 SOM 空间分布变化全貌的方法相比, GWR 对不同位置上的有机质与对应影响因子间赋予不同系数, 能够反

映局部变化特征^[31]。参与 GWR 建模的主要影响因子系数有正有负, 其中碱解氮、有效铁和全氮的系数中值分别为 0.130、0.002 和 1.890, 与 SOM 含量呈正相关, 而交换性镁和阳离子交换量的系数中值分别为 -0.053、-0.138, 与 SOM 含量呈负相关。此外, 5 个主要影响因子与 SOM 显著相关 ($P < 0.05$), Moran 指数范围在 0.968~0.990 之间,

表 5 地理加权回归模型预测参数统计

Table 5 Prediction parameter statistics of geographically weighted regression model

变量 Variables	最小值 Min	最大值 Max	中值 Median	Moran 指数 Moran's index	P 值 P-value
截距 Intercept	3.618	7.230	5.140	0.986	0.000
碱解氮/(mg/kg) AN	0.104	0.143	0.130	0.985	0.000
有效铁/(mg/kg) AI	0.002	0.003	0.002	0.980	0.000
交换性镁/(cmol/kg) EM	-0.344	0.043	-0.053	0.968	0.000
全氮/(g/kg) TN	0.485	3.842	1.890	0.990	0.000
阳离子交换量/(cmol/kg) CEC	-0.191	0.008	-0.138	0.989	0.000

表现出强烈的空间自相关性。

2.4 土壤有机质预测模型残差分析

通过 MLR、PLSR 以及 GWR 模型预测, 得到 SOM 残差半方差变异函数(表 6)。依据均方根误差最小的原则, 分别选择指数模型、指数模型和高斯模型对 3 组预测残差进行插值。MLR 残差、PLSR 残差和 GWR 残差的块基比分别是 34.61%、39.38% 和 68.42%, 均具有中等的空间相关性, 说明

SOM 预测残差的空间异质性受影响因子影响的程度比随机因素要大, 可以进一步分析残差部分与主要影响因子的关系。图 2 中, A、B、C 分别代表建立 MLR、PLSR 和 GWR 模型后获得的 SOM 残差制图, 它们的空间分布特征整体相似。SOM 大于 2.71 g/kg 的残差正值主要集中在研究区的西北部, 该地区为人迹罕至的茂密林地, 富集大量枯枝落叶, 便于有机质积累, 因此, 造成实测值高于预测值的结果。

而 SOM 小于 -1.05 g/kg 的残差负值分布在研究区中心的大面积范围, 此处的土地利用类型为园地和耕地, 高估实测值说明农作物的生长吸收了土壤中的有机质, 需要定期施用有机肥料以保证当地农产品的产量和品质。

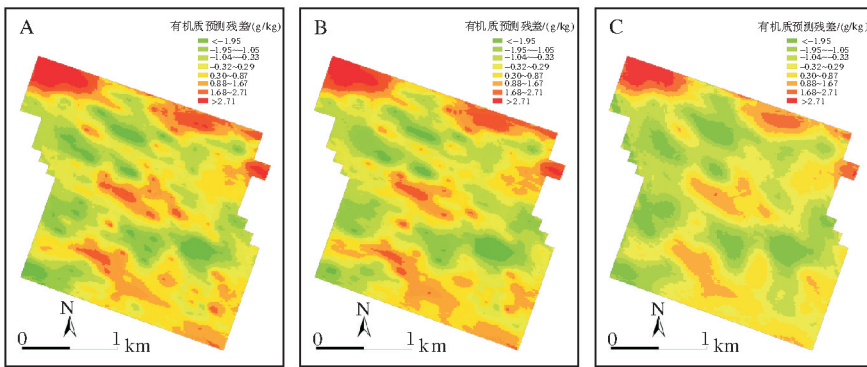
其中, 经 MLR 法和 PLSR 法预测获得的 SOM 残差分布在 $1.68 \sim 2.71 \text{ g/kg}$, 比 GWR 残差插值得到的面积更多且空间分布更加破碎, 而 GWR 的大部分有机质残差分布在 $-1.04 \sim 0.87 \text{ g/kg}$ 这个范围, 预测结果更接近于真实值。出现这样的空间分

布结果是由于使用 MLR 和 PLSR 建模时, 仅分析建模集中 SOM 数据与已知影响因子数据间的全局数量关系, 忽略了临近样点间的局部权重。GWR 建模则兼顾了 SOM 的空间异质性和依赖性, 因此, 整体预测残差低于 MLR 和 PLSR。基于以上研究基础, 分别对仍具有一定空间自相关性的 GWR 模型残差进行 MLR、PLSR 建模, 尝试用全局模型解释局部模型中未能识别的残差项及主要影响因子间的相互关系, 最终将结果与作为趋势项的 GWR 预测部分叠加, 以期获得 2 种结合全局与局部模型优

表 6 土壤有机质预测残差半方差模型参数

Table 6 Parameters of soil organic matter prediction residual semi-variance models

插值内容 Content	拟合模型 Model	模型参数 Model parameters				预测误差 Prediction error	
		块金值 C_0 Nugget	拱高 C Partial sill	基台值 ($C_0 + C$)	块基比/% [$C_0 / (C_0 + C)$]	变程/m Range	均方根误差 RMSE
MLR 残差 MLR residuals	指数模型 Exponential	1.651	3.119	4.770	34.612	31.860	2.148
PLSR 残差 PLSR residuals	指数模型 Exponential	1.985	3.056	5.041	39.377	33.333	2.199
GWR 残差 GWR residuals	高斯模型 Gaussian	2.957	1.365	4.322	68.417	33.839	2.071



A、B 和 C 分别代表经 MLR、PLSR 和 GWR 建模获得的 SOM 残差空间分布图。A, B, and C represent the spatial distribution of SOM residuals obtained by MLR, PLSR and GWR modeling, respectively.

图 2 不同模型预测获得的土壤有机质残差空间分布图

Fig.2 Spatial distribution of soil organic matter residuals predicted by different models

势的 GWR 扩展模型。

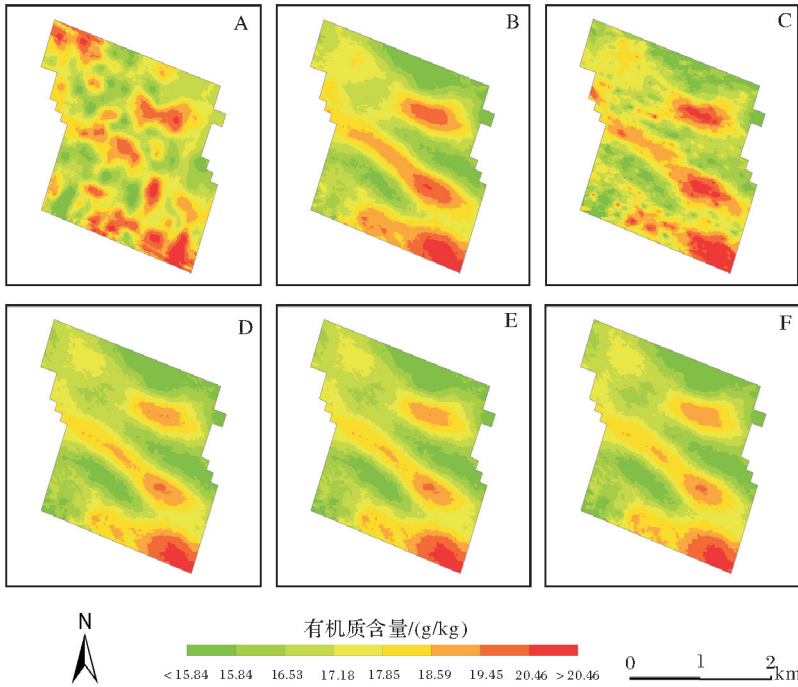
2.5 土壤有机质空间分布与预测模型精度比较

参照对建模集 SOM 实测值进行普通克里格插值 (ordinary Kriging, OK) 的结果, 分别观察对 SOM 及其主要影响因子间建立 MLR、PLSR、GWR、 GWR_{MLR} 和 GWR_{PLSR} 预测模型后的 SOM 空间分布差异 (图 3)。对比显示, 除对单一变量进行 OK 插值获得的 SOM 空间分布表现得十分破碎 (图 3A), 其余模型因受各主要影响因子的综合影响, 预测结果均表现出与土地利用类型对应的区域

化特征并自北向南伴随有高低值交叉分布的条带结构。高值区 ($>19.45 \text{ g/kg}$) 位置是大面积连续分布的园地和耕地, 土地利用类型单一, 空间变异较小。低值区 ($<16.53 \text{ g/kg}$) 位置则对应于研究区中夹杂分布有其他土地利用类型的土地, 频繁的人为活动降低了 SOM 积累。其中, 经 MLR 和 PLSR 建模制图的中部、南部有机质含量 $>18.59 \text{ g/kg}$ 的面积比其他模型预测分布得更多 (图 3B, C), 这部分也是图 2 中模型残差差异最明显的位置。而因距离衰退公式在不同地理位置上赋予主要影响因子以变化系

数,使用 GWR、GWR_{MLR} 及 GWR_{PLSR} 模型获得的 SOM 预测结果过渡更平滑,且在空间分布上具有一致性(图 3D-F)。因此,使用全局模型与局部模型预

测 SOM 空间分布的差异情况可通过制图结果对比得到,但局部扩展模型预测优势并未体现,需进一步分析精度验证指标。



A、B、C、D、E 和 F 分别代表经 OK、MLR、PLSR、GWR、GWR_{MLR} 和 GWR_{PLSR} 建模获得的 SOM 空间分布图。A、B、C、D、E 和 F represent the SOM spatial distributions obtained through OK、MLR、PLSR、GWR、GWR_{MLR} and GWR_{PLSR} modeling, respectively.

图 3 不同模型预测获得的土壤有机质空间分布图

Fig.3 Spatial distribution of soil organic matter predicted by different models

使用验证集数据验证由建模集构建的几种 SOM 预测方法的精度指标。由表 7 可知,只对 SOM 建立 OK 模型的实测值与预测值间拟合程度最差,证实复杂环境下加入主要影响因子对辅助预测 SOM 的必要性。与 MLR 和 PLSR 相比,GWR 的均方根误差 RMSE 降低到 3.143,相对分析误差 RPD 和相关系数 r 分别升高到 1.465 和 0.744,说明使用考虑了局部空间位置关系模型的预测效果比只考虑全局数量关系模型的要好。而细化分析了

GWR 模型残差空间非平稳性的 2 种局部扩展模型 GWR_{MLR} 和 GWR_{PLSR},除在相关系数 r 上无法突破 GWR,其均方误差 MSE、均方根误差 RMSE 均有降低,相对分析误差 RPD 均有上升,预测精度更是较最初的 OK 模型分别提高了 35.53% 和 35.57%,其中 GWR_{PLSR} 的模型解释能力最佳。由此可见,除 SOM 与主要影响因子间存在空间相关性,模型残差也在预测有机质含量中起重要作用,因此,不应被忽略。

表 7 土壤有机质预测方法精度指标对比

Table 7 Comparison of precision indexes of soil organic matter prediction methods

方法 Method	均方误差 MSE	均方根误差 RMSE	相对分析误差 RPD	相关系数 r
OK	23.701	4.868	0.945	0.099
MLR	10.603	3.256	1.414	0.727 **
PLSR	10.857	3.295	1.397	0.733 **
GWR	9.878	3.143	1.465	0.744 **
GWR _{MLR}	9.845	3.138	1.467	0.743 **
GWR _{PLSR}	9.834	3.136	1.468	0.743 **

3 讨论

本研究使用地理探测器提取影响湖北省宜都市红花套镇典型柑橘种植区 SOM 空间分布的 5 种主要影响因子,以 OK 插值结果为参照,分别建立 SOM 及其主要影响因子间的 MLR、PLSR、GWR、 GWR_{MLR} 和 GWR_{PLSR} 模型,探讨它们对 SOM 含量及空间分布的预测效果。结果表明,对 SOM 及其主要影响因子间建立 MLR、PLSR 以及 GWR 模型的残差半方差函数块基比均在 25%~75% 之间,具有中等空间自相关,说明有机质预测残差仍受主要影响因子影响。而考虑了空间异质性的局部模型的预测结果更接近真实值。因此,可尝试进一步分析其残差项与主要影响因子间的相关性。在预测 SOM 含量的全局模型、局部模型及其扩展模型中, GWR_{PLSR} 的模型预测精度提高了 35.57%,解释能力最佳, GWR_{MLR} 其次,说明结合了全局与局部模型优势的 GWR 扩展模型既可在研究区的整体范围内考虑 SOM 的空间非平稳性,又能在存在预测残差的小面积区域上解释空间自相关性,更适宜进行典型柑橘种植区的 SOM 预测。此外,由于本研究仅在小尺度研究区内讨论了使用局部扩展模型对预测 SOM 空间分布及含量的研究意义,该模型是否适用于更大更复杂的环境中以及预测效果如何,还需深入挖掘。

参 考 文 献

- [1] 李玲,张少凯,吴克宁,等.基于土壤系统分类的河南省土壤有机质时空变异[J].土壤学报,2015,52(5):979-990.
- [2] 赵业婷,齐雁冰,常庆瑞,等.渭河平原县域农田土壤有机质时空变化特征[J].土壤学报,2013,50(5):1048-1053.
- [3] 刘艳芳,卢延年,郭龙,等.基于地类分层的土壤有机质光谱反演校正样本集的构建[J].土壤学报,2016,53(2):332-341.
- [4] 唐梦迎,丁建丽,夏楠,等.干旱区典型绿洲土壤有机质含量分布特征及其影响因素[J].土壤学报,2017,54(3):759-766.
- [5] 赵明松,张甘霖,吴运金,等.江苏省土壤有机质含量时空变异特征及驱动力研究[J].土壤学报,2014,51(3):448-458.
- [6] 鲁剑巍,陈防,王富华,等.湖北省柑橘园土壤养分分级研究[J].植物营养与肥料学报,2002(4):390-394.
- [7] 黄安,杨联安,杜挺,等.基于多元成土因素的土壤有机质空间分布分析[J].干旱区地理,2015,38(5):994-1003.
- [8] KUMAR S. Estimating spatial distribution of soil organic carbon for the Midwestern United States using historical database [J]. *Chemosphere*, 2015, 127: 49-57.
- [9] ZENG C Y, YANG L, ZHU A X, et al. Mapping soil organic

matter concentration at different scales using a mixed geographically weighted regression method [J]. *Geoderma*, 2016, 281: 69-82.

- [10] 付苗苗,刘梅英,牛智有,等.基于近红外光谱法的水稻秸秆可溶性糖快速检测[J].华中农业大学学报,2016,35(2):115-121.
- [11] 于雷,洪永胜,耿雷,等.基于偏最小二乘回归的土壤有机质含量高光谱估算[J].农业工程学报,2015,31(14):103-109.
- [12] 侯艳军,塔西甫拉提·特依拜,买买提·沙吾提,等.荒漠土壤有机质含量高光谱估算模型[J].农业工程学报,2014,30(16):113-120.
- [13] GUO L, ZHAO C, ZHANG H T, et al. Comparisons of spatial and non-spatial models for predicting soil carbon content based on visible and near-infrared spectral technology [J]. *Geoderma*, 2017, 285: 280-292.
- [14] 宁秀红,郭龙,张海涛.基于空间自回归和地理加权回归模型的不同尺度下土地利用程度研究[J].华中农业大学学报,2013,32(4):48-54.
- [15] KUMAR S, LAL R, LIU D S. A geographically weighted regression kriging approach for mapping soil organic carbon stock [J]. *Geoderma*, 2012, 189: 627-634.
- [16] 瞿明凯,李卫东,张传荣,等.地理加权回归及其在土壤和环境科学上的应用前景[J].土壤,2014,46(1):15-22.
- [17] 杨顺华,张海涛,陈家赢,等.平原丘陵过渡带土壤有机碳空间分布及环境影响[J].中国环境科学,2015,35(12):3728-3736.
- [18] KUMAR S, LAL R, LIU D S, et al. Estimating the spatial distribution of organic carbon density for the soils of Ohio, USA [J]. *Journal of geographical sciences*, 2013, 23(2): 280-296.
- [19] HARRIS P, FOTHERINGHAM A S, CRESPO R, et al. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets [J]. *Mathematical geosciences*, 2010, 42(6): 657-680.
- [20] 鲍士旦.土壤农化分析[M].北京:中国农业出版社,2000.
- [21] 杨顺华,张海涛,郭龙,等.基于回归和地理加权回归 Kriging 的土壤有机质空间插值[J].应用生态学报,2015,26(6):1649-1656.
- [22] 王劲峰,徐成东.地理探测器:原理与展望[J].地理学报,2017,72(1):116-134.
- [23] WANG J F, LI X H, CHRISTAKOS G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China [J]. *International journal of geographical information science*, 2010, 24(1): 107-127.
- [24] 湛东升,张文忠,余建辉,等.基于地理探测器的北京市居民宜居满意度影响机理[J].地理科学进展,2015,34(8):966-975.
- [25] HU Y, WANG J F, LI X H, et al. Geographical detector-based risk assessment of the under-five mortality in the 2008 Wenchuan earthquake, China [J]. *PLoS One*, 2011, 6(6): e21427.
- [26] 覃文忠.地理加权回归基本理论与应用研究[D].上海:同济大学,2007.
- [27] 于婧.基于 GIS 和地统计学方法的土壤养分空间变异及应用研究[D].武汉:华中农业大学,2007.

- [28] 徐小逊,张世榕,余妮娜,等.沱江中游土壤有效铁空间分布及其影响因素分析[J].西南农业学报,2012,25(3):977-981.
- [29] 张巧萍,张玉亭,聂胜委,等.信阳毛尖茶园土壤有机质和硫、镁含量研究[J].土壤通报,2015,46(1):153-156.
- [30] 张水清,黄绍敏,郭斗斗.河南三种土壤阳离子交换量相关性及其预测模型研究[J].土壤通报,2011,42(3):627-631.
- [31] 连纲,郭旭东,傅伯杰,等.基于环境相关法和地统计学的土壤属性空间分布预测[J].农业工程学报,2009,25(7):237-242.

Spatial distribution and content prediction of soil organic matter in typical citrus growing areas

DUAN Lijun ZHANG Haitao GUO Long DU Peiying CHEN Ke JU Qinglan

College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China

Abstract 329 soil samples were collected from the citrus growing areas in Honghuatao Town, Yidu City, Hubei Province. Based on the principle of spatial stratified heterogeneity, the top five major impact factors having the greatest correlation with soil organic matter (SOM) were selected with the GeoDetector software. Using the interpolation results of ordinary Kriging as control, the global model multiple linear regression (MLR), partial least squares regression (PLSR) and local model geographical weighted regression (GWR) were established by the soil organic matter and its main environmental factors. After analyzing the structure of the model residuals, GWR_{MLR} and GWR_{PLSR} were constructed as the extensions of GWR model. The results showed that the mean square error (MSE), root mean square error (RMSE), relative analysis error (RPD) and the correlation coefficient (r) between measured and predicted values of GWR_{PLSR} were 9.834, 3.136, 1.468, 0.743, respectively. The GWR_{PLSR} model had the highest prediction accuracy, followed by GWR_{MLR} . In summary, except for the spatial correlation between SOM and its major impact factors, analyzing model residuals can further eliminate the predicted instability. Therefore, taking the model residual terms into consideration is more suitable to predict the regional SOM spatial distribution and digital soil mapping.

Keywords soil organic matter; spatial stratified heterogeneity; GeoDetector; model residuals; GWR_{MLR} ; GWR_{PLSR}

(责任编辑:陆文昌)