

DOI: 10.11766/trxb201710240410

基于三种空间预测模型的海南岛土壤有机质空间分布研究*

姜赛平¹ 张怀志¹ 张认连¹ 李兆君¹ 谢良商² 徐爱国^{1†}

(1 中国农业科学院农业资源与农业区划研究所, 北京 100081)

(2 海南省农业科学院农业环境与土壤研究所, 海口 571100)

摘要 为探索适合热带地形复杂区土壤有机质 (SOM) 含量的空间预测方法, 以海南岛为研究区域, 结合地形因子、归一化植被指数、土壤类型、土地利用类型变量, 选用普通克里格法 (OK)、回归克里格法 (RK)、随机森林模型 (RF) 三种方法对训练集 128 个样点 SOM 含量的空间分布规律进行预测, 并通过验证集 32 个验证点比较了三种方法的预测精度。结果表明: (1) 0~5 cm 土层三种方法的平均预测误差 (ME) 均接近于 0, 从均方根预测误差 (RMSE) 来看, RF (0.8867) < RK (0.9104) < OK (0.9641), 从决定系数 (R^2) 来看, RF (0.2141) > RK (0.1715) > OK (0.0708)。综合以上三个参数, 该土层最优拟合模型为 RF。同理得出 0~20、20~40、40~60 cm 土层的最优拟合模型分别为 RF、RF、OK。RK 和 RF 能够更好地描述 SOM 含量局部变异信息; (2) 四个土层 SOM 含量的均值分别为 19.67、15.89、10.30、8.07 g kg⁻¹, 呈现出西南、东北高, 西部、东南沿海地区低的空间分布趋势。

关键词 海南岛; 土壤有机质; 随机森林模型; 空间分布

中图分类号 S159 **文献标识码** A

海南岛处于热带地区, 岛内地形复杂, 土地利用多样, 具有丰富的热带作物资源, 是我国重要的热带作物生产基地。研究该区土壤有机质 (SOM) 含量的空间变异规律, 对于了解热带气候条件下、不同热带作物种植方式下土壤肥力的空间分布状况具有重要意义。

近些年来, 随着“3S”等信息技术的快速发展, 数字土壤属性制图方法被广泛应用到土壤学领域中, 用以描述土壤属性的空间变异规律。常见的数字土壤属性制图方法有地统计学制图方法、机器学习方法和混合模型方法等。地统计学制图方法是应用最广泛也是最为成熟的方法。其中, 普通克里格法 (Ordinary kriging, OK) 和回归克里格法

(Regression-kriging, RK) 最具代表性。OK 因方法简单、容易操作而被普遍接受。但土壤属性的空间变异同时受地形因子、土地利用类型、土壤类型等众多因素的影响^[1-3]。由于未考虑到辅助变量对土壤属性的影响, 因而成图只能描述土壤属性的整体空间分布规律, 对局部信息描述不够详细, 且平滑效应严重^[4]。为了更好地揭示土壤属性空间变异规律, 结合辅助变量的 RK 被应用于土壤属性制图当中。赵永存等^[5]采用多元线性回归, 泛克里格法和 RK 分别对河北省土壤有机碳的空间分布规律进行预测, 结果表明, 结合地形属性的 RK 预测精度最高, 且对土壤有机碳局部变异信息描述地更加详细。连纲等^[3]采用该法结合地形因子与遥感指数对黄土高

* 国家科技部基础性工作专项 (2012FY112100) 资助 Supported by the Special Basic Research Funds of China (No. 2012FY112100)

† 通讯作者 Corresponding author, E-mail: xuaiguo@caas.cn

作者简介: 姜赛平 (1992—), 女, 辽宁省凌源市人, 硕士研究生, 研究方向为土壤资源与管理。E-mail: saipingjiang@163.com

收稿日期: 2017-10-24; 收到修改稿日期: 2018-01-28; 优先数字出版日期 (www.cnki.net): 2018-03-05

原丘陵沟壑区土壤属性的空间分布规律进行预测,结果表明,该法不但能够提高制图精度,且能消除部分平滑效应。杨顺华等^[6]选取相对高程和汇流动力指数为辅助变量采用该法对湖北省枝城镇SOM空间分布规律进行预测,亦取得较高的精度。RK能够较好地描述地形复杂区土壤属性的空间分布规律,但该法要求数据满足正态分布和内蕴假设^[7],给实际问题的处理带来一定的困难。支持向量机是一种基于统计学习理论的机器学习方法,相对于地统计学方法,它不需要任何前提假设,在解决小样本、高维数、非线性以及局部极小点等问题方面有着较高的性能,但其要求辅助变量为连续型变量,且需对辅助变量进行标准化处理,减小了辅助变量空间分布的异质性,影响辅助变量与土壤属性间的协同变化关系^[8]。目前该法在数字土壤属性制图中应用相对较少。神经网络模型、分类与回归树模型是近些年来发展较为迅速的机器学习方法,对于连续型变量和类别型变量均适用,且不需对数据进行预处理,能够很好地描述土壤属性与环境变量间的非线性关系,虽有研究表明二者能够提高制图精度^[9-10],但同时也存在着一些缺陷,例如:神经网络模型的参数较多,不易确定^[11],模型的预测结果也不易解释^[12]。分类与回归树模型的结构取决于样本数据,对于样本量较小的数据,训练样本的微小变化可能导致分类节点的不同^[13]。并且二者均易过度拟合^[14-15]。因此并未广泛应用至土壤属性制图当中。随机森林(Random Forest, RF)模型在分类与回归树模型基础上发展而来,模型参数相对较少,计算简单,且不易过度拟合^[16],被引入数字土壤属性制图当中,在地形相对复杂地区,能够取得较高的制图精度^[17-18]。

目前涉及海南全岛的SOM含量空间变异特征研究较少,结合土地利用类型、土壤类型等信息对不同模型的模拟结果进行对比,从而选出适合该区SOM含量空间变异模型的研究更是鲜见^[19-20]。本文根据地形复杂区模型适用情况,结合土地利用类型、土壤类型和地形因子等辅助变量,选用OK、RK和RF三种方法预测该区SOM含量的空间分布特征,通过精度验证确定出最优模型,从而为地处热带、较大区域尺度地形复杂区SOM含量空间变异制图方法的选取提供依据。

1 材料与方法

1.1 研究区概况

海南岛位于 $18^{\circ} 10' \sim 20^{\circ} 10' N$, $108^{\circ} 37' \sim 111^{\circ} 3' E$ 之间,是我国第二大岛,面积 $32\ 900\text{ km}^2$ 。该岛属于热带季风气候,年均气温 $23 \sim 25^{\circ}\text{C}$, $\geq 10^{\circ}\text{C}$ 年积温 $8\ 200 \sim 9\ 200^{\circ}\text{C}$,年平均降水量为 $1\ 720\text{ mm}$ 。岛内地形复杂,由山地、丘陵、台地、平原组成中间高耸、四周低平的环形层状地貌。成土母质有花岗岩、砂页岩、浅海沉积物、玄武岩等10种。土壤类型包括砖红壤、水稻土、赤红壤、燥红土、风沙土、紫色土等15个土类。土地利用主要为水田、旱地、林地,岛内以种植热带作物为主,种类繁多。

1.2 样品采集与分析

土壤样品的采集时间为2012年11月—2012年12月,采用近似网格法,综合考虑当地的地形地貌、土壤类型和土地利用类型等因素,以样点具有代表性、均匀分布性为原则,在全岛19个县市布点,采样间距平均为 10 km ,共采集样点163个,每个样点按 $0 \sim 5$ 、 $0 \sim 20$ 、 $20 \sim 40$ 、 $40 \sim 60\text{ cm}$ 分层采样,分别研究表层、耕层、中层和中下层SOM含量空间分布状况。其中, $0 \sim 5$ 和 $0 \sim 20\text{ cm}$ 分别取样, $0 \sim 5\text{ cm}$ 土钻取样点紧邻 $0 \sim 20\text{ cm}$ 土钻点,记录样点经纬度、地形和土地利用类型等信息。土壤类型包括砖红壤、水稻土、赤红壤、燥红土、火山灰土、紫色土、风沙土、石灰(岩)土等8个土类。土地利用类型包括水田、旱地、园地。

土壤样品SOM含量的测定采用重铬酸钾-硫酸消化法,具体步骤参见文献^[21]。

1.3 数据来源

辅助变量: RK和RF均在6个地形因子: 高程(x_1)、坡度(x_2)、坡向(x_3)、平面曲率(x_4)、剖面曲率(x_5)、地形湿度指数(x_6), 3个环境因子: 归一化植被指数(NDVI, x_7), 土地利用类型(x_8)、土壤类型(x_9)的基础上进行剔除。

辅助变量数据来源: (1) 海南岛 $1:50\ 000$ 等高线矢量图; (2) 遥感影像: 2013年10月26日和2013年12月6日的海南岛Landsat8 OLI_TIRS四景卫星影像,数据来源于中国科学院计算机网络信息中心国际科学数据镜像网站(<http://www.gscloud.cn>); (3) 海南岛 $1:200\ 000$ 土壤图; (4) 2010

年1:100 000遥感解译土地利用图,数据源自中国科学院资源环境科学数据中心。

1.4 模型方法

本研究采用OK、RK和RF对SOM含量的空间分布特征进行预测。其中,RF的基本原理参见文献[22]。树节点预选的变量个数(mtry)和随机森林中树的个数(ntree)是RF的两个重要参数。本文通过对不同个数的辅助变量逐次计算确定出最优的mtry值^[12]。设定ntree为500、1 000、1 500、2 000,结合mtry值,选择使模型袋外误差(OOB)最小的参数组合用于最终预测。

1.5 数据处理与分析

(1)数字高程模型的生成。采用1:50 000等高线矢量图在ArcGIS9.3中生成10 m分辨率的数字高程模型栅格影像。

(2)哑变量赋值。由于土壤类型和土地利用类型为类别变量,不能直接用于回归分析,因此采用哑变量赋值方法^[23]对其赋值,赋值结果如下:

$(x_{91}=1, x_{92}=0, x_{93}=0, x_{94}=0, x_{95}=0)$, $(x_{91}=0, x_{92}=1, x_{93}=0, x_{94}=0, x_{95}=0)$, $(x_{91}=0, x_{92}=0, x_{93}=1, x_{94}=0, x_{95}=0)$, $(x_{91}=0, x_{92}=0, x_{93}=0, x_{94}=1, x_{95}=0)$, $(x_{91}=0, x_{92}=0, x_{93}=0, x_{94}=0, x_{95}=1)$

分别代表砖红壤、水稻土、赤红壤、燥红土、火山灰土和紫色土。 $(x_{81}=1, x_{82}=0)$, $(x_{81}=0, x_{82}=1)$, $(x_{81}=0, x_{82}=0)$ 分别代表水田、旱地、园地。

(3)异常值处理与验证集的选取。采用阈值法将区间(平均值 $\pm 3 \times$ 标准差)外的数值视作异常值并剔除,并参见文献[2],将风沙土、石灰(岩)土样点剔除,剩余有效样点160个,从有效样点中随机选取128个样点作为训练集,用于空间数据分析,其余32个样点用于验证。全部保留样点土壤类型覆盖海南岛91.24%的陆域面积。样点分布见图1。

(4)模型精度验证。采用验证集的平均预测误差(ME)、均方根预测误差(RMSE)和决定系数(R^2)来评价预测的准确性。

(5)分析工具。地形因子的提取以及SOM含量空间制图在ArcGIS9.3软件中完成,遥感影像镶嵌和NDVI的提取在ENVI5.1软件中完成,地统计学分析和随机森林规则分别在R3.3.1的gstat包和randomForest包中完成,统计分析使用SAS9.2软件和Microsoft Excel 2010工具。

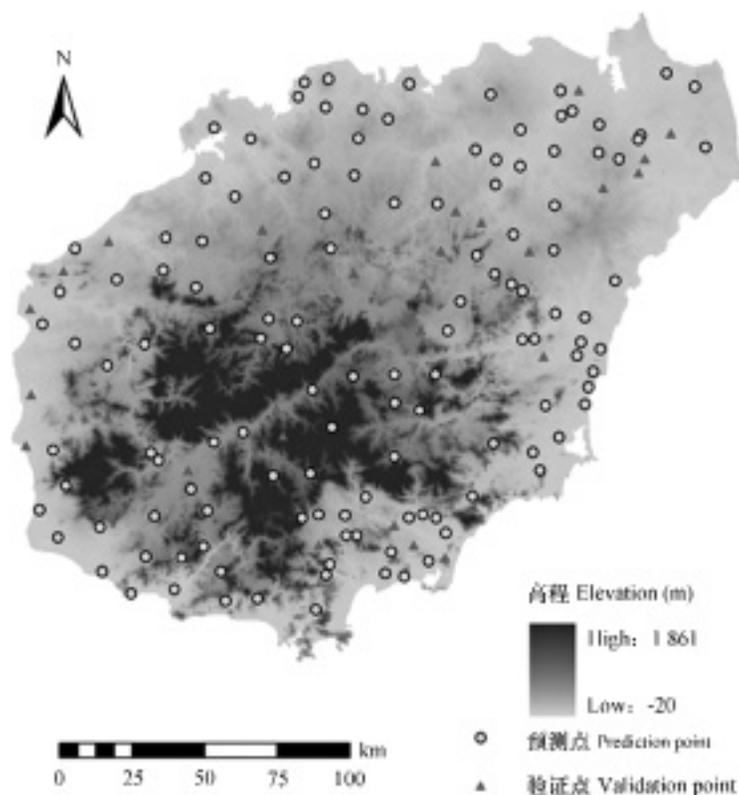


图1 研究区采样点分布图

Fig. 1 Distribution of sampling sites in the study area

2 结果与讨论

2.1 不同土层SOM含量的描述性统计特征

由表1可知,随着土层深度的增加,SOM含量的最小值、最大值逐渐减小,最小值由4.52减小至1.73 g kg⁻¹,最大值由45.83减小至21.53 g kg⁻¹。从表层(0~5 cm)到深层(40~60 cm),SOM含量

的均值分别为19.67、15.89、10.30、8.07 g kg⁻¹。按全国第二次土壤普查养分分级标准,除40~60 cm土层SOM含量的均值处于五级水平外,其他三个土层均处于四级水平。四个土层SOM含量的标准误差值均较小,表明各个土层数据波动不是很大,结果较可靠。从变异程度看,四个土层SOM含量的变异系数在47.16%~56.56%,均属于中等变异。

表1 SOM含量基本统计特征

Table 1 Statistical characters of SOM contents

土层 Depth (cm)	样本数 Sample size	最小值 Min (g kg ⁻¹)	最大值 Max (g kg ⁻¹)	极差 Range (g kg ⁻¹)	均值 Mean (g kg ⁻¹)	标准误差 Std err (g kg ⁻¹)	变异系数 CV (%)
0~5	162	4.52	45.83	41.31	19.67	0.74	47.90
0~20	162	3.16	38.12	34.96	15.89	0.59	47.16
20~40	162	1.89	23.88	21.99	10.30	0.43	53.75
40~60	162	1.73	21.53	19.80	8.07	0.36	56.56

2.2 基于OK的SOM含量空间分布

对训练集数据进行Kolmogorov-Smirnov检验发现,四个土层原始SOM含量数据 p 均小于0.01,均不符合正态分布。对其取算数平方根后,0~5、0~20、20~40和40~60 cm, $p > 0.05$,满足正态分布。

由图2可知,0~5 cm土层SOM含量共分为6个等级,按占全岛面积的比例进行排序如下:20~25 (32.63%) > 15~20 (25.83%) > 10~15 (19.36%) > 25~30 (12.99%) > 6~10 (8.74%) > 30~40 g kg⁻¹ (0.45%), SOM含量 > 10 g kg⁻¹的等级占全岛面积的91.26%。其中高值(25~30、30~40 g kg⁻¹)主要分布在岛内东北的琼山南部、文昌西南部、定安东部、琼海北部地区和岛中部的琼中地区,低值(6~10 g kg⁻¹)分布在岛西部和西南部的昌江、东方沿海一带。整体而言,该岛SOM含量呈现出西部地区低于东部地区的空间分布趋势,这与前人^[19-20]的研究结果相一致。0~20 cm土层SOM含量的空间变异规律与表层相似,但SOM含量等级数减少,>10 g kg⁻¹等级的土壤面积占全岛的84.53%,其中10~15 g kg⁻¹和15~20 g kg⁻¹等级所占比例相当,二者共占全岛面积的70.98%。20~40 cm土层>10 g kg⁻¹等级的土壤面积占全岛的44.54%。而40~60 cm土层SOM含量

>10 g kg⁻¹等级的土壤面积仅占全岛的2.85%,该土层主要分布在6~10 g kg⁻¹范围内,占全岛面积的72.12%。

2.3 基于RK的SOM含量空间分布

本文采用逐步线性回归方程拟合SOM含量与辅助变量之间的趋势,然后用简单克里格对剔除趋势后的残差进行估计,将二者结果叠加(即RK)作为最终预测结果。

由表2可知,四个土层拟合方程对总方差的解释率分别为:26.83%、17.12%、18.68%、18.04%,其回归系数相对较小,这是因为本研究区面积较大,SOM与辅助变量之间的关系较复杂,而实际进入方程的变量较少,故解释率相对较低^[17]。如果考虑气温、降水等因素,可能会获得更高的解释率。根据李燕丽等^[24]的统计,逐步线性回归模型对土壤属性变异的解释率在15%~82%,本研究结果在该范围内。且从概率来看,四个土层拟合模型均是极显著的,表明方程能够很好地描述SOM含量的变异规律。NDVI和土壤类型变量在四个土层中均进入方程,表明这两个变量是影响SOM含量的重要因素。有研究表明^[1],NDVI和SOM含量呈极显著正相关关系,即该指数越大,SOM含量越高。海南岛西部NDVI值低于东部,与西部较东部干旱,植

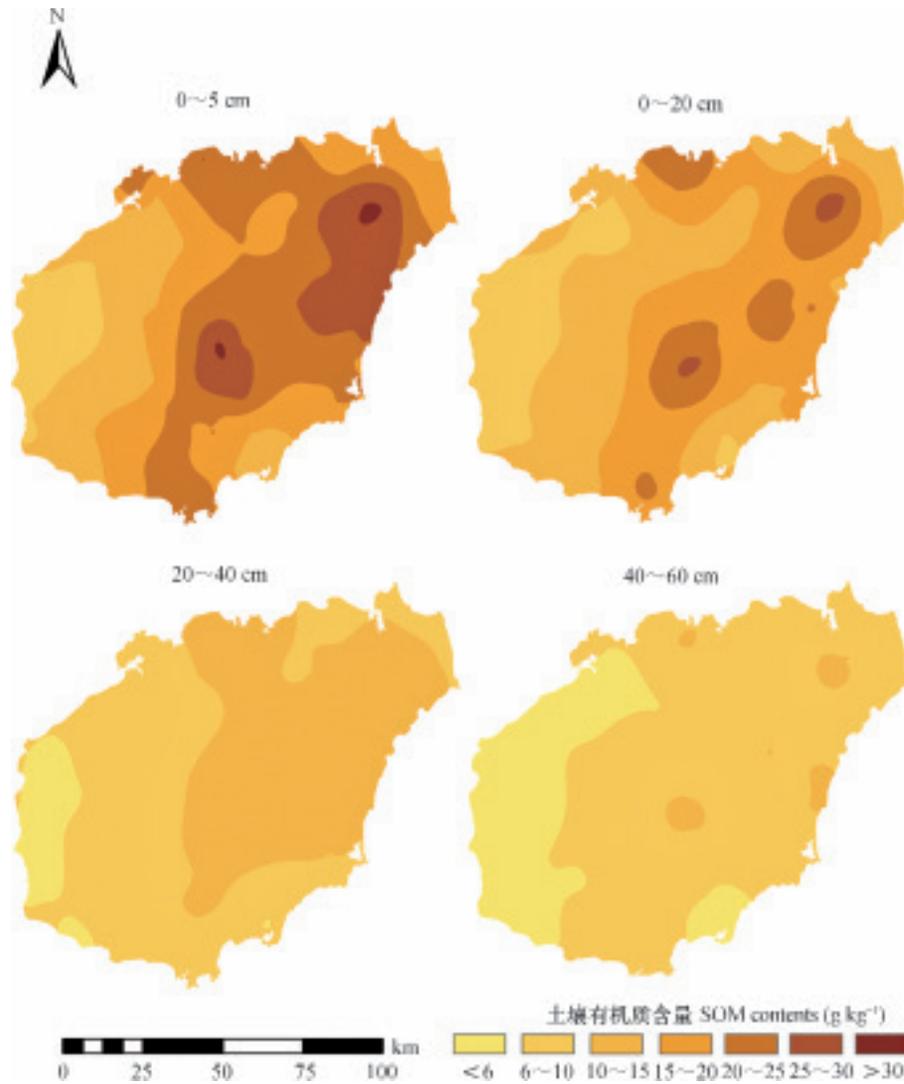


图2 OK预测SOM含量空间分布图

Fig. 2 OK-based SOM content spatial distribution map

表2 各土层逐步线性回归方程拟合结果

Table 2 Fitting of the soil layers with the stepwise linear regression equation

土层 Depth (cm)	方程 Equator	调整回归系数 Adj Rsq	概率 Probability
0~5	$sd5=1.783+0.002\ 7x_1+0.060\ 8x_6+3.335\ 8x_7+0.733\ 3x_{81}+1.864\ 8x_{95}$	0.268 3	<0.001
0~20	$sd20=3.164\ 4+2.229\ 5x_7-0.670\ 6x_{82}-1.015\ 7x_{94}+1.096\ 9x_{95}$	0.171 2	<0.001
20~40	$sd40=1.824\ 8+3.015\ 8x_7+1.140\ 5x_{93}+0.915\ 2x_{95}$	0.186 8	<0.001
40~60	$sd60=1.753\ 9+2.267\ 6x_7+1.032\ 4x_{93}+1.182\ 6x_{95}$	0.180 4	<0.001

注：sd5代表0~5cm土层SOM含量的算术平方根，其他同上。逐步线性回归方程变量进入和剔除的概率水平均为0.05 Note: sd5 stands for arithmetic square root of SOM contents in the 0~5 cm soil layer and the same with the others. The probabilities of entering and removing variables of the stepwise linear regression equation are both 0.05

被生长茂密程度低于东部相一致。海南岛土壤类型繁多，研究表明^[25]，在热带土壤中，不同的土壤类型，SOM含量差别较大。曾迪等^[26]的研

究也表明SOM含量的空间变异受土壤类型的影响。从表2中还可以看出，土地利用类型变量仅在0~5 cm和0~20 cm土层进入方程，表明该变量

对耕层的影响较大，而对深层影响较小。这是因为该区土地利用类型以水田、旱地和园地为主，受人为影响较大，随着土层加深，施肥、耕作等人为扰动减小，所以对SOM含量变异的影响也减小。

由图3可知，应用RK预测SOM含量的空间变异规律与OK预测所得结果略有不同，呈现出西南、北部、东北高，西部、东南沿海地区低的空间分布趋势。RK对细节刻画地更加清晰，这是引入了辅助变量的缘故。该法对SOM含量划分等级相对于OK增多，各个等级所占的面积有所变化，高低值分布略有不同。0~5 cm土层SOM含量各等级占全岛面积的比例排序如下：15~20 (30.28%)>10~15 (22.40%)>20~25 (21.83%)>25~30 (10.18%)>6~10 (6.75%)>30~40 (6.46%)>高于

40 (1.18%)>低于6 g kg⁻¹ (0.84%)，与OK相比，该法增加了高于40 g kg⁻¹和低于6 g kg⁻¹两个等级，对SOM含量的预测范围更广。高值（25~30、30~40 g kg⁻¹）较OK预测区域有所增加，主要分布在琼山、儋州北部、定安南部、琼中西南部、五指山东北部、保亭西部、三亚东北部、乐东西北部地区，低值（6~10 g kg⁻¹）与OK法所得结果相似，主要分布在岛西部地区。这主要是引入辅助变量土壤类型所致。以下三个土层与0~5 cm分布规律相似，但等级数减少，根据表1统计，随着土层深度的增加，SOM含量减少。0~20 cm土层SOM含量>10 g kg⁻¹等级的土壤面积占全岛的86.26%，与OK该等级占全岛面积的比例相当，而10~15 g kg⁻¹和15~20 g kg⁻¹较OK有所减少，分别减少5.56%和3.02%，两个等级共占全岛面积的

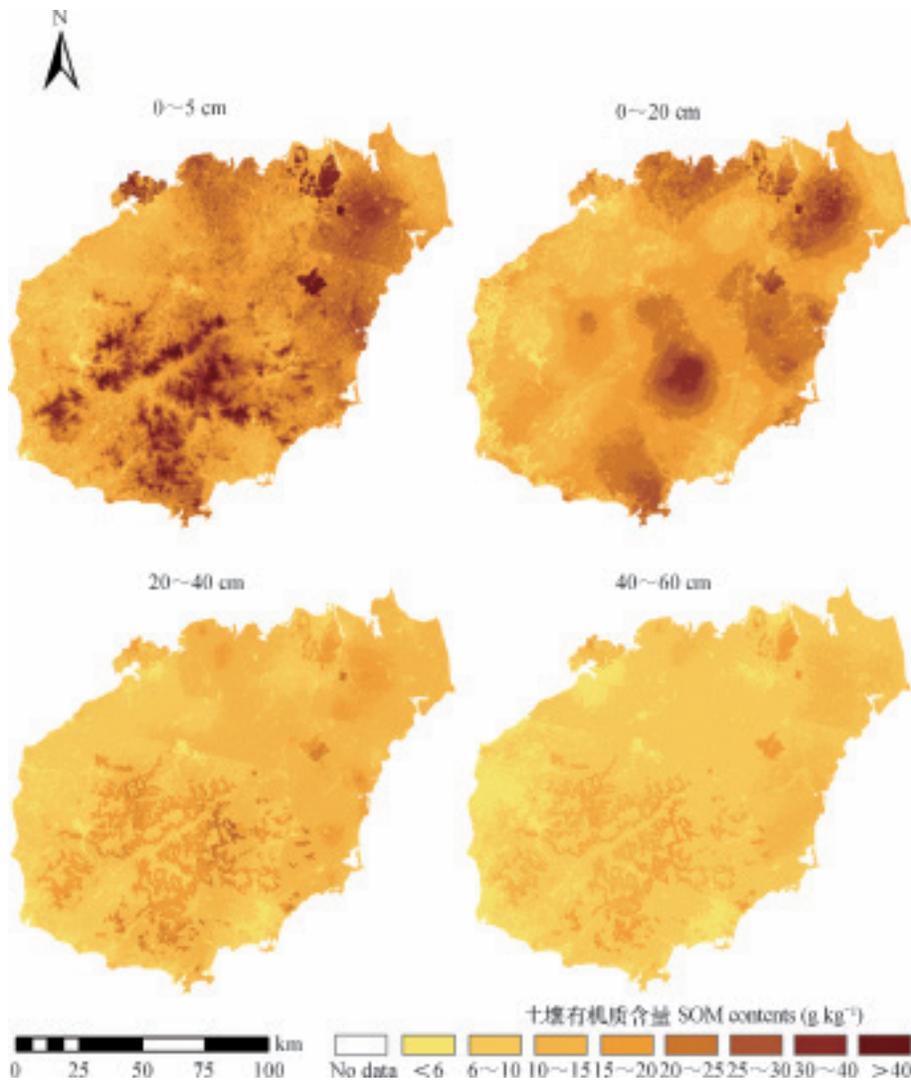


图3 RK预测SOM含量空间分布图

Fig. 3 RK-based SOM content spatial distribution map

62.40%; 20~40 cm土层 $>10 \text{ g kg}^{-1}$ 等级的土壤面积占全岛比例较OK增加10.55%, 与OK相比增加了15~20、20~25、 $>25 \text{ g kg}^{-1}$ 三个等级; 40~60 cm土层主要分布在 $6 \sim 10 \text{ g kg}^{-1}$ (54.35%), 与OK法所得结果相似。

2.4 基于RF的SOM含量空间分布

由图4可知, RF预测的SOM含量空间分布图与RK预测图相似, 呈现出西南、东北高, 西部、东南沿海地区低的空间分布趋势, 因为二者均引入了辅助变量, 能够更好地描述SOM含量空间变异的细节信息。对SOM含量等级的划分, 该法在0~5 cm和0~20 cm土层与OK相同, 40~60 cm土层与RK相同, 20~40 cm土层介于两者之间。RF四个土层 $>10 \text{ g kg}^{-1}$ 等级的土壤面积占全岛比例

在三种方法中均最高, 分别为99.79%、98.08%、60.32%、34.80%。0~5 cm土层SOM含量主要在 $15 \sim 25 \text{ g kg}^{-1}$, 占全岛面积的74.6%; 0~20 cm土层SOM含量主要在 $10 \sim 20 \text{ g kg}^{-1}$ (81.72%), 与前两种方法所得结果一致; 20~40 cm土层SOM含量主要在 $6 \sim 15 \text{ g kg}^{-1}$ (84.2%), 与RK法所得结果相同; 40~60 cm土层SOM含量主要在 $6 \sim 15 \text{ g kg}^{-1}$ (89.15%), 与前两种方法相比, $10 \sim 15 \text{ g kg}^{-1}$ 等级所占比例明显增加。

2.5 不同预测模型的精度评价

由表3可知, 0~5 cm土层, 三种预测方法中, 从 R^2 来看, $\text{RF} > \text{RK} > \text{OK}$, 表明RF对SOM含量空间变异的解释能力最强, 其次为RK法, OK法最弱; 从RMSE来看, $\text{RF} < \text{RK} < \text{OK}$, 表明RF法预测精度最高, RK法

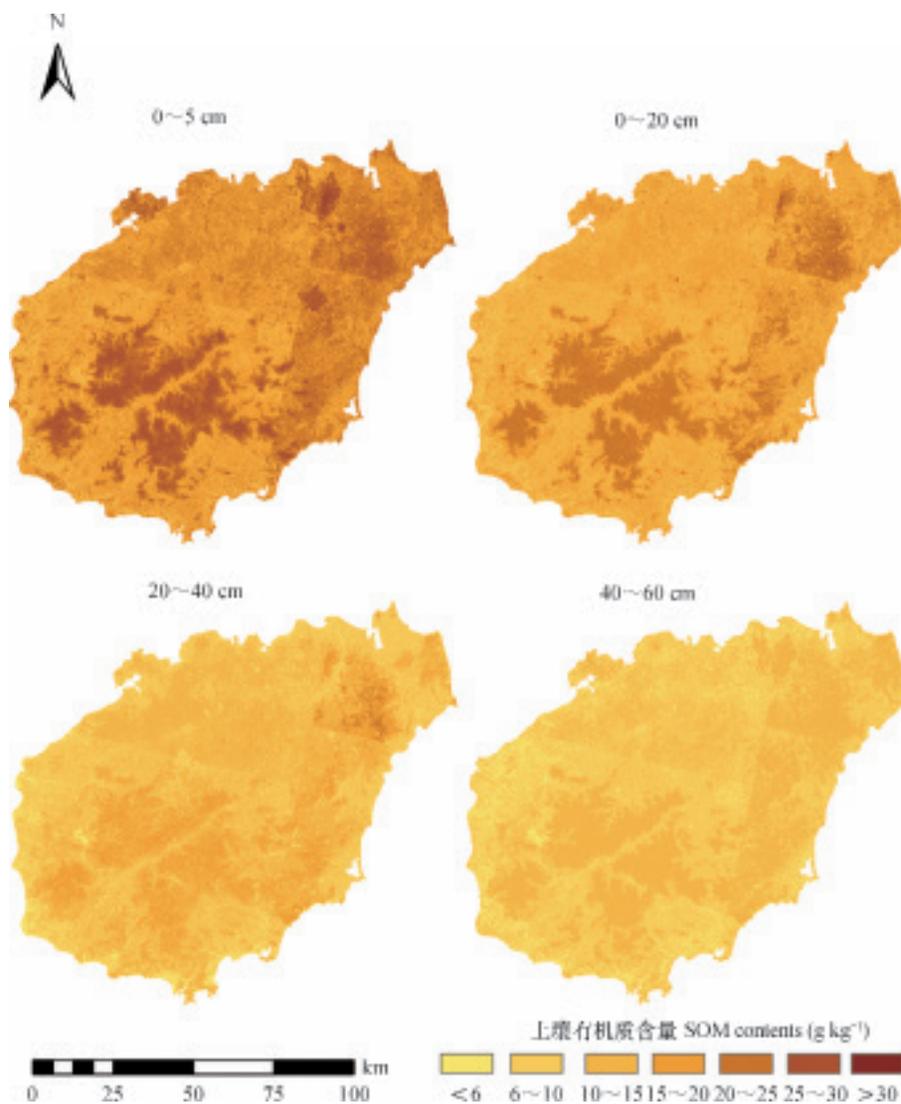


图4 RF预测SOM含量空间分布图

Fig. 4 RF-based SOM content spatial distribution map

居中, 而OK法预测精度最低; OK法的ME更接近于0, 其次为RK法, 而RF法的无偏估计相对较差; 从以上三个参数综合来看, 该土层的最优拟合模型为RF。

以此方法分析其他土层, 确定0~20、20~40和40~60 cm三个土层的最优拟合模型, 分别为RF法、RF法和OK法。

表3 各土层SOM含量预测模型精度

Table 3 Accuracy of the prediction of SOM contents relative to soil layer

土层 Depth (cm)	OK			RK			RF		
	ME	RMSE	R^2	ME	RMSE	R^2	ME	RMSE	R^2
0~5	0.001 5	0.964 1	0.070 8	0.005 0	0.910 4	0.171 5	0.179 8	0.886 7	0.214 1
0~20	-0.029 5	0.745 7	0.222 7	0.036 4	0.820 0	0.060 2	0.166 0	0.691 5	0.331 6
20~40	-0.177 5	0.720 3	0.233 5	-0.084 1	0.729 6	0.213 5	0.052 4	0.652 3	0.371 3
40~60	-0.178 6	0.679 6	0.290 5	-0.106 6	0.709 2	0.227 4	0.034 0	0.724 1	0.194 7

虽然RF对该区SOM含量空间变异的描述在0~5、0~20和20~40 cm土层表现较好。但由表3可知, 验证集中决定系数 (R^2) 在0.19~0.37之间, 较前人的研究结果^[17-18,27] 偏低。这是因为以往的研究采样密度在0.03~2.52个 km^{-2} , 是本研究区采样密度 (0.004个 km^{-2}) 的8倍~610倍, 采样密度小可能是造成本研究区RF预测精度偏低的原因之一。此外, 本研究选取的辅助因子相对较少, 有研究表明^[19,28], 气温、降水和人为因素 (如施肥等) 是影响SOM含量空间变异的重要因素, 如果增加上述因子可能会提高模型的解释率。

OK在本研究中仅适合40~60 cm土层, 且从预测图 (图2) 来看, 该法图斑较大, 仅能够预测出SOM含量的整体空间分布规律, 对SOM含量空间变异的细节信息描述地不够详细。这是因为本研究区地形复杂, 而有研究表明, OK通常适用于土壤属性变化较为均匀的区域, 对地形复杂、土壤属性变化较为强烈的区域, 其制图精度不太理想^[29], 40~60 cm土层受地形因子的影响相对表层要小, 故OK法的制图精度相对表层和其他方法均有所提高; 此外, 本研究采样密度较小, 对OK的制图精度也会有一定的影响。

RK在本研究区的应用效果不是很理想, 决定系数 R^2 与前人^[30] 的研究相比较小, 这可能是由于剔除趋势的模型选择不合适, 在本研究中采用的是逐步线性回归方程来剔除趋势, 而土壤属性与辅助变量间的关系常常是复杂的非线性关系, 且因子之间可能存在交互作用, 逐步线性回归模型难以描述上述信息。以后的研究中可以考虑采用其他回归模

型 (如回归树或RF等) 来剔除土壤属性与辅助变量间的趋势。

从研究结果中还可以看出, 不同土层所选择的最优模型有所不同, 这可能是不同土层的影响因子略有不同, 不同的模型根据自身机理, 对于土壤属性与辅助变量间的关系的解释程度不同, 因而导致最终的预测结果不同。

3 结论

本研究在较大区域尺度地形复杂地区, 对海南岛128个样点数据, 采用OK、RK和RF结合高程、归一化植被指数、土地利用类型和土壤类型等辅助变量对该区0~5、0~20、20~40、40~60 cm四个土层SOM含量的空间分布规律进行预测, 并以32个验证点进行验证。主要结论如下: 0~5、0~20和20~40 cm土层的最优拟合模型均为RF, 而40~60 cm土层的最优拟合模型为OK, RK和RF相对于OK对SOM含量的局部变异信息描述地更加详细。四个土层SOM含量的均值分别为19.67、15.89、10.30、8.07 g kg^{-1} , 随着土层深度的增加, SOM含量逐渐减小。从空间预测图来看, 四个土层SOM含量均呈现出西南、东北高, 西部、东南沿海地区低的空间分布趋势。

参考文献

- [1] 张素梅, 王宗明, 张柏, 等. 利用地形和遥感数据预测土壤养分空间分布. 农业工程学报, 2010, 26 (5): 188—194

- Zhang S M, Wang Z M, Zhang B, et al. Prediction of spatial distribution of soil nutrients using terrain attributes and remote sensing data (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26 (5): 188—194
- [2] 顾成军, 史学正, 于东升, 等. 省域土壤有机碳空间分布的主控因子——土壤类型与土地利用比较. *土壤学报*, 2013, 50 (3): 425—432
- Gu C J, Shi X Z, Yu D S, et al. Main factor controlling SOC spatial distribution at the province scale as affected by soil type and land use (In Chinese). *Acta Pedologica Sinica*, 2013, 50 (3): 425—432
- [3] 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. *农业工程学报*, 2009, 25 (7): 237—242
- Lian G, Guo X D, Fu B J, et al. Prediction of the spatial distribution of soil properties based on environmental correlation and geostatistics (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 2009, 25 (7): 237—242
- [4] 陆访仪, 赵永存, 黄标, 等. 海伦市耕层土壤有机质含量空间预测方法研究. *土壤通报*, 2012, 43 (3): 662—667
- Lu F Y, Zhao Y C, Huang B, et al. Comparison of predicting methods for mapping the spatial distribution of topsoil organic matter content in cropland of Hailun (In Chinese). *Chinese Journal of Soil Science*, 2012, 43 (3): 662—667
- [5] 赵永存, 史学正, 于东升, 等. 不同方法预测河北省土壤有机碳密度空间分布特征的研究. *土壤学报*, 2005, 42 (3): 379—385
- Zhao Y C, Shi X Z, Yu D S, et al. Different methods for prediction of spatial patterns of soil organic carbon density in Hebei Province, China (In Chinese). *Acta Pedologica Sinica*, 2005, 42 (3): 379—385
- [6] 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. *应用生态学报*, 2015, 26 (6): 1649—1656
- Yang S H, Zhang H T, Guo L, et al. Spatial interpolation of soil organic matter using regression kriging and geographically weighted regression kriging (In Chinese). *Chinese Journal of Applied Ecology*, 2015, 26 (6): 1649—1656
- [7] 刘爱利, 王培法, 丁园圆, 等. *地统计学概论*. 北京: 科学出版社, 2012
- Liu A L, Wang P F, Ding Y Y, et al. Introduction of geostatistics (In Chinese). Beijing: Science Press, 2012
- [8] 刘超, 卢玲, 胡晓利. 数字土壤质地制图方法比较——以黑河张掖地区为例. *遥感技术与应用*, 2011, 26 (2): 177—185
- Liu C, Lu L, Hu X L. Comparison analysis on digital soil texture mapping in an area of Zhangye, Heihe River basin (In Chinese). *Remote Sensing Technology and Application*, 2011, 26 (2): 177—185
- [9] 李启权, 王昌全, 岳天祥, 等. 基于神经网络模型的中国表层土壤有机质空间分布模拟方法. *地球科学进展*, 2012, 27 (2): 175—184
- Li Q Q, Wang C Q, Yue T X, et al. Method for spatial simulation of topsoil organic matter in China based on a neural network model (In Chinese). *Advances in Earth Science*, 2012, 27 (2): 175—184
- [10] 周斌, 王繁. 基于决策树模型的土壤性质空间推断. *土壤通报*, 2004, 35 (4): 385—390
- Zhou B, Wang F. Spatial prediction of soil properties based on decision tree modeling (In Chinese). *Chinese Journal of Soil Science*, 2004, 35 (4): 385—390
- [11] 徐剑波, 宋立生, 彭磊, 等. 土壤养分空间估测方法研究综述. *生态环境学报*, 2011, 20 (8/9): 1379—1386
- Xu J B, Song L S, Peng L, et al. Research review on methods of spatial prediction of soil nutrients (In Chinese). *Ecology and Environmental Sciences*, 2011, 20 (8/9): 1379—1386
- [12] 黄文, 王正林. *数据挖掘: R语言实战*. 北京: 电子工业出版社, 2014
- Huang W, Wang Z L. *Data mining: R in action* (In Chinese). Beijing: Electronic Industry Press, 2014
- [13] Breiman L. Bagging predictors. *Machine Learning*, 1996, 24: 123—140
- [14] McKenzie N J, Ryan P J. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 1999, 89: 67—94
- [15] Ito E, Ono K, Ito Y M, et al. A neural network approach to simple prediction of soil nitrification potential: A case study in Japanese temperate forests. *Ecological Modeling*, 2008, 219: 200—211
- [16] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述. *统计与信息论坛*, 2011, 26 (3): 32—38
- Fang K N, Wu J B, Zhu J P, et al. A review of technologies on random forests (In Chinese). *Statistics & Information Forum*, 2011, 26 (3): 32—38

- [17] 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. 农业工程学报, 2015, 31 (5): 194—200
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 (In Chinese). Transactions of the Chinese Society of Agricultural Engineering, 2015, 31 (5): 194—200
- [18] 王茵茵, 齐雁冰, 陈洋, 等. 基于多分辨率遥感数据与随机森林算法的土壤有机质预测研究. 土壤学报, 2016, 53 (2): 342—354
Wang Y Y, Qi Y B, Chen Y, et al. Prediction of soil organic matter based on multi-resolution remote sensing data and random forest algorithm (In Chinese). Acta Pedologica Sinica, 2016, 53 (2): 342—354
- [19] 吕丽平, 王登峰, 魏志远, 等. 近20年海南岛土壤有机碳时空变异. 热带作物学报, 2016, 37 (2): 338—344
Lü L P, Wang D F, Wei Z Y, et al. The spatial and temporal variability of SOC in Hainan island over the past 20 years (In Chinese). Chinese Journal of Tropical Crops, 2016, 37 (2): 338—344
- [20] 方圆. 海南省土壤有机碳时空变异. 海口: 海南大学, 2012
Fang Y. On the spatial distribution patterns of soil organic carbon in the Hainan Province (In Chinese). Haikou: Hainan University, 2012
- [21] 张甘霖, 龚子同. 土壤调查实验室分析方法. 北京: 科学出版社, 2012
Zhang G L, Gong Z T. Soil survey laboratory methods (In Chinese). Beijing: Science Press, 2012
- [22] Breiman L. Random forests. Machine Learning, 2001, 45: 5—32
- [23] 李河, 麦劲壮, 肖敏, 等. 哑变量在Logistic回归模型中的应用. 循证医学, 2008, 8 (1): 42—45
Li H, Mai J Z, Xiao M, et al. Application of dummy variable in logistic regression models (In Chinese). The Journal of Evidence-Based Medicine, 2008, 8 (1): 42—45
- [24] 李燕丽, 潘贤章, 王昌昆, 等. 广西中南部耕地土壤有机质和全氮变化的遥感监测. 生态学报, 2014, 34 (18): 5283—5291
Li Y L, Pan X Z, Wang C K, et al. Monitoring changes of soil organic matter and total nitrogen in cultivated land in Guangxi by remote sensing (In Chinese). Acta Ecologica Sinica, 2014, 34 (18): 5283—5291
- [25] 王景华, 卫林, 景可. 我国海南岛热带土壤的养分特征. 生态学杂志, 1984 (4): 15—18
Wang J H, Wei L, Jing K. Nutrient characteristics of tropic soils in Hainan Island of China (In Chinese). Chinese Journal of Ecology, 1984 (4): 15—18
- [26] 曾迪, 漆智平, 黄海杰, 等. 海南儋州耕地土壤有机质空间变异. 热带作物学报, 2015, 36 (1): 199—204
Zeng D, Qi Z P, Huang H J, et al. Spatial variability of soil organic matter content in Danzhou City, Hainan Province (In Chinese). Chinese Journal of Tropical Crops, 2015, 36 (1): 199—204
- [27] Wiesmeier M, Barthold F, Blank B, et al. Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. Plant and Soil, 2011, 340: 7—24
- [28] 郭乃嘉, 史学正, 赵永存, 等. 人为与环境因子对农田土壤有机质影响的比较研究——以典型黑土区和水稻土区为例. 土壤学报, 2016, 53 (5): 1097—1106
Guo N J, Shi X Z, Zhao Y C, et al. Comparative study on impacts of anthropogenic and environment factors on soil organic matter: A case study of typical black soil region and paddy soil region (In Chinese). Acta Pedologica Sinica, 2016, 53 (5): 1097—1106
- [29] 龙军, 张黎明, 沈金泉, 等. 复杂地貌类型区耕地土壤有机质空间插值方法研究. 土壤学报, 2014, 51 (6): 1270—1281
Long J, Zhang L M, Shen J Q, et al. Spatial interpolation of soil organic matter in farmlands in areas complex in landform (In Chinese). Acta Pedologica Sinica, 2014, 51 (6): 1270—1281
- [30] 刘艳芳, 宋玉玲, 郭龙, 等. 结合高光谱信息的土壤有机碳密度地统计模型. 农业工程学报, 2017, 33 (2): 183—191
Liu Y F, Song Y L, Guo L, et al. Geostatistical models of soil organic carbon density prediction based on soil hyperspectral reflectance (In Chinese). Transactions of the Chinese Society of Agricultural Engineering, 2017, 33 (2): 183—191

Research on Spatial Distribution of Soil Organic Matter in Hainan Island Based on Three Spatial Prediction Models

JIANG Saiping¹ ZHANG Huaizhi¹ ZHANG Renlian¹ LI Zhaojun¹ XIE Liangshang² XU Aiguo^{1†}

(¹ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

(² Agricultural Environment and Soil Research Institute of Hainan Academy of Agricultural Sciences, Haikou 571100, China)

Abstract The knowledge about spatial distribution characteristics of soil organic matter (SOM) contents is of important significance to development of the agriculture. Models different in simulation mechanism are used to predict distribution of SOM contents, because different regions have soils of different soil properties and different environmental factors affecting soil properties. It is, therefore, essential to choose an optimal suitable model site-specifically for the prediction, this paper is oriented to explore for a model that is capable of predicting spatial distribution of SOM contents in a tropical region complicated in relief, like Hainan Island. Ordinary kriging (OK), regression-kriging (RK) and random forest (RF) were the three models used to predict spatial distribution of SOM contents in the Hainan Island based on the training set of 128 samples, combined with topographic factors, normalized vegetation index, soil types and landuse types, and then the predictions were validated with a validating set of 32 validation points in the study area for comparison of prediction accuracy. Results show: (1) The optimal fitting model for prediction of distribution of organic matter contents in the 0 ~ 5, 0 ~ 20, 20 ~ 40 and 40 ~ 60 cm soil layer is RF, RF, RF and OK, respectively. RK and RF can better describe local variation of SOM contents; (2) The mean value of SOM contents in the four soil layers was 19.67, 15.89, 10.30 and 8.07 g kg⁻¹, respectively, showing a spatial distribution trend that SOM contents are higher in the southwest and northeast parts and lower in the west and southeast parts of the island.

Key words Hainan Island; Soil organic matter; Random forest model; Spatial distribution

(责任编辑：檀满枝)