

DOI: 10.11766/trxb202001140623

袁玉琦, 陈瀚阅, 张黎明, 任必武, 邢世和, 童珺玥. 基于多变量与 RF 算法的耕地土壤有机碳空间预测研究——以福建亚热带复杂地貌区为例[J]. 土壤学报, 2021, 58(4): 887–899.

YUAN Yuqi, CHEN Hanyue, ZHANG Liming, REN Biwu, XING Shihe, TONG Junyue. Prediction of Spatial Distribution of Soil Organic Carbon in Farmland Based on Multi-Variables and Random Forest Algorithm—A Case Study of A Subtropical Complex Geomorphic Region in Fujian as an Example[J]. Acta Pedologica Sinica, 2021, 58(4): 887–899.

基于多变量与 RF 算法的耕地土壤有机碳空间预测研究* ——以福建亚热带复杂地貌区为例

袁玉琦, 陈瀚阅[†], 张黎明, 任必武, 邢世和, 童珺玥

(福建农林大学资源与环境学院, 土壤生态系统健康与调控福建省高校重点实验室, 福州 350002)

摘要: 耕地土壤有机碳(Soil Organic Carbon, SOC)含量既是土壤质量的重要表征, 也是农业温室气体重要源库, 而环境变量与随机森林算法(Random Forest, RF)是提高土壤有机碳空间预测精度的重要方法, 但不同组合环境变量对 RF 模型预测精度的影响仍需深入研究。本文以福建闽东南复杂地貌区为例, 以两种环境变量组合(遥感变量+气候因子和遥感变量+气候因子+土壤属性)为输入数据集, 利用 RF 算法对耕地表层 SOC 含量进行模拟预测和精度对比, 并与普通克里格(Ordinary Kriging, OK)插值模型进行对比。结果表明, 基于全部环境变量构建的 RF 模型表现最佳, 其模型拟合度和预测精度相较于未加入土壤属性的模型有显著提高(r 提高 7.95%, 为 0.95, RMSE 下降 45.13%), 且对 SOC 空间分异信息的捕获更精确, OK 模型总体预测精度最弱。利用最优模型反演得到的研究区耕地 SOC 含量为 $14.70 \pm 2.95 \text{ g} \cdot \text{kg}^{-1}$, 东部沿海低于西部内陆。变量贡献率分析显示, 除了与土壤碳紧密相关的水解性氮(N), 遥感变量中数字高程模型(DEM)也是影响闽东南地区 SOC 预测精度的重要变量, 因此, 遥感变量、气候因子和土壤属性共同驱动的随机森林模型可作为闽东南复杂地貌区耕地有机碳含量空间预测的有效方法。

关键词: 土壤有机碳; 随机森林; 变量组合; 空间分布; 精度评价

中图分类号: S15 文献标志码: A

Prediction of Spatial Distribution of Soil Organic Carbon in Farmland Based on Multi-Variables and Random Forest Algorithm—A Case Study of a Subtropical Complex Geomorphic Region in Fujian as an Example

YUAN Yuqi, CHEN Hanyue[†], ZHANG Liming, REN Biwu, XING Shihe, TONG Junyue

(University Key Lab of Soil Ecosystem Health and Regulation in Fujian, College of Resource and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

* 国家自然科学基金项目(41971050)、福建农林大学科技创新专项基金项目(KFA17616A)、福建省科技计划项目(2017N5006)和国家级大学生创新训练计划项目(201910389026)共同资助 Supported by the National Natural Science Foundation of China(No. 41971050), Special Fund for Science and Technology Innovation of Fujian Agriculture and Forestry University (No. KFA17616A), Science and Technology Planning Project of Fujian Province (No. 2017N5006) and National Undergraduate Innovation Program (No.201910389026)

[†] 通讯作者 Corresponding author, E-mail: chenhanyue.420@163.com

作者简介: 袁玉琦(1995—), 女, 江苏南通人, 硕士研究生, 主要从事土壤属性制图研究。E-mail: yuanyuqi918@126.com

收稿日期: 2020-01-14; 收到修改该日期: 2020-05-21; 网络首发日期(www.cnki.net): 2020-08-24

Abstract: 【 Objective 】 Soil organic carbon (SOC) plays an important role in soil fertility and the terrestrial ecosystem carbon cycle. A detailed understanding of the spatial distribution of SOC is vital to management of the soil resources and mitigation of the global climate change. With the development of the 3S technology, the models for predicting soil properties based on environmental variables are getting increasingly popular. The purpose of our study is to try to simulate the complex and nonlinear relationship between SOC and environmental variables, and evaluate the importance of soil attributes to accuracy in SOC mapping. 【 Method 】 For this purpose, machine learning methods and a random forest (RF) model was applied to map the spatial distribution of topsoil organic carbon contents for farmlands in the high-yield agricultural areas in Southeast Fujian. A set of environmental variables (including 5 hard-to-obtain quantitative soil attributes such as hydrolysable nitrogen, available phosphorus, pH, etc) and 11 easy-to-obtain variables (i.e. topography factors, vegetation indexes and climate factors) were acquired through analysis of a large number of soil samples collected from that region, and then processed with the RF algorithm to predict spatial distribution of SOC content in the topsoil layers of the farmlands of that region. Two different combinations of the above variables were entered as input to RF-S model and RF-A model separately. The RF-S model functioned only on the basis of easy-to-obtain variables and the RF-A model did on the basis of all the variables, both easy-to- or hard-to-obtain ones, for predicting SOC. Root mean square errors (RMSE), mean absolute errors (MAE), Pearson correlation coefficients (r), coefficients of variation (CV), relative errors (RE) and relative root mean square errors (RRMSE) of the two models were worked out for evaluation of accuracy of their predictions, and screening-out of an optimal RF model for mapping SOC in the study area based the raster datasets of all variables. Then cross-validation was performed to compare the optimal RF model with the Ordinary Kriging (OK) interpolation model. 【 Result 】 Results show that of the two models, different in input of environmental variables, the RF-A model that functioned based on remote sensing variables, climate factors and soil attributes was much better than the other in performance and could explain the most of the spatial heterogeneity of SOC. Compared with the RF-S model, the RF-A model significantly improved in fitting and prediction (r increased by 7.95% and RMSE decreased by 45.13%). The SOC contents of the farmlands of the region predicted with the RF-A model varied in the range of $14.70 \pm 2.95 \text{ g}\cdot\text{kg}^{-1}$ and were quite similar to what was obtained with the OK model in spatial distribution, i.e. an ascending trend from the east coastal area to the western inland of the study area. And despite sampling percentage, the RF-A model was generally higher than the OK model in prediction accuracy, and in capability of capturing spatial heterogeneity, and preferred especially in the case of relatively fewer sampling sites. Among the variables, hydrolysable nitrogen (N) was the most important one for the RF-A model, and followed by elevation (DEM). Both of them significantly affected spatial heterogeneity of the SOC, exhibiting positive relationships with SOC. 【 Conclusion 】 It is therefore concluded that the random forest model that functions based on remote sensing variables, climate factors as well as soil attributes is a promising approach to predicting spatial distribution of SOC in Southeast Fujian. In addition, soil attributes variables, such as N and P, should be taken into account for improving prediction accuracy for mapping of SOC in regions with complex geomorphology.

Key words: Soil organic carbon; Random forest; Combination of variables; Spatial distribution; Accuracy evaluation

耕地土壤有机碳 (Soil Organic Carbon, SOC) 是衡量土壤质量和表征土壤肥力的重要指标,也是农业温室气体减排潜力的主要来源。掌握土壤有机碳的含量及空间分布,对提升耕地质量、评估土壤健康和缓解全球气候变化具有重要意义。实地采样和分析是获取土壤属性空间分布格局的传统方法,但因费时费力、成本高、区域通达性差限制了其在地区、国家乃至全球尺度的应用^[1]。随着 3S 技术的发展,越来越多学者通过建立辅助环境变量与土壤属性

之间的预测模型来实现区域土壤属性空间制图^[2]。

目前,用于数字土壤制图的预测模型主要包括逐步线性回归模型^[3]、递归偏最小二乘算法^[4]及随机森林算法 (Random Forest, RF)^[5]等。其中 RF 算法是一种机器学习算法,它由分类和回归树 (Classification and Regression Tree, CART) 模型构成,因其在处理多元非线性数据方面的优势,越来越多地被用于土壤属性空间预测^[6]。相继有学者将 RF 模型用于大区域有机碳含量及储量^[7-9]的空间预

测，并取得了较好的预测精度（ r 值在 0.7~0.95 范围内）。但 RF 算法在土壤有机碳空间制图的应用起步较晚，其不同区域的预测精度和可行性方面仍需进行深入研究，其中需要关注的一个研究重点即是环境变量的筛选^[10-11]。

在基于 RF 算法的土壤有机碳空间预测中，最为常用的辅助环境因子包括数字高程模型（Digital Elevation Models, DEM）和遥感植被指数^[7, 11]等遥感变量。土壤有机碳含量高低与地表植被覆盖状况和地形地貌密切相关^[12]，而遥感变量因在地形和植被空间信息表达上具备独特优势，且较易获取，常被用于数字土壤制图。但土壤理化性质是多方面成土因素综合作用的结果，单纯利用遥感变量预测土壤有机碳有所欠缺。部分学者^[8, 13]在此基础上增加了相对较易获取的气候因子（如温度、降水）和土壤类型、成土母质等土壤定性属性。然而，与土壤有机碳机理上互为相关的土壤定量属性数据因获取难度较大，以往较少被用于土壤有机碳空间预测，但随着测定技术的发展，该类变量获取难度有所降低，日渐被部分学者关注。Liu 等^[14]在植被要素和气候因子等较易获取的环境变量基础上增加了容重、土层厚度等土壤属性，基于 RF 算法模拟中国北方草地土壤有机碳密度（ R^2 达到 0.73）。Were 等^[9]在对比三种机器学习算法（包括 RF 算法）用于南非地区 SOC 储量预测的研究中，创新性地选择 Mg、Ca、P、全氮（Total Nitrogen, TN）和 pH 等土壤定量属性联合遥感数据和地形及衍生因子作为模型驱动因子，结果显示土壤属性中的 TN 变量在所有三种模型的变量贡献率中占绝对主导地位。

由此可见，基于 RF 模型预测 SOC 所用的环境变量随科学测定技术的发展有所创新，但国内外研究少有基于不同组合的环境变量对预测精度影响的评估，尤其是获取难度较大的土壤定量属性，作为变量加入对模型精度的贡献如何尚不十分明确。此外，针对国内基于 RF 算法的 SOC 预测多应用于地貌类型较单一的小尺度区域^[13, 15]，而大范围复杂地貌区因地形、气候、植被、土壤肥力状况等因素在水平和垂直方向上存在显著分异^[16]，导致 SOC 的空间分布规律及其与环境变量间的关系更为复杂，不同变量对 SOC 预测精度的贡献率仍需进一步评价。RF 模型正因为模拟变量间复杂非线性关系和高阶相互作用方面的突出优势^[17]，更适用于本文亚热

带复杂地貌区土壤有机碳含量的空间预测。

目前，我国已经完成两次全国性的土壤普查、农业农村部测土配方施肥重大国家计划，也正在进行农业长期定位实验和第三次土地调查等，积累了大量的土壤样点及属性数据，这些较难获取的土壤定量属性数据可为提高模型预测精度提供极为便利的研究条件。基于此，本文以典型的亚热带复杂地貌区——闽东南地区为研究区域，采用随机森林模型，选取易获取的遥感变量和气候因子及不易获取的土壤属性作为模型输入，基于大量土壤有机碳实测样点分别训练两种不同变量组合数据集驱动的 RF 模型，并进行有机碳含量空间分布格局的预测和对比，以期为区域地形地貌复杂区有机碳储量的准确估算提供理论基础。

1 材料与方 法

1.1 研究区概况

研究区位于福建省东南部，简称闽东南地区（ $23^{\circ}32' \sim 26^{\circ}04' \text{ N}$ ， $116^{\circ}53' \sim 119^{\circ}91' \text{ E}$ ），由漳州市、厦门市、莆田市、泉州市（除永春县、德化县和金门县）和福州市（仅长乐区、福清市和平潭综合试验区）组成（图 1），区域总面积为 29 373 km²，其中耕地面积约 4 384 km²。该区位于闽中大山带戴云山-博平岭段东南侧，地貌复杂、地形起伏，海拔最高 1 152 m，最低 0 m，从内陆向沿海递减，其中山地丘陵（海拔 > 200 m）面积高达 44.33%。受亚热带季风气候与地形因素综合作用，年平均气温为 20.0 ℃，年平均降雨量为 1 526 mm，主要集中于 3—8 月，占全年降雨量 80% 以上。从土地利用方式来看，可以用作农业用途的土地占比小（耕地仅为 14.93%），且受海拔、地形等要素影响，80.15% 的耕地位于福建省最主要的农业高产区——沿海平原台地（海拔 ≤ 200 m）^[18]。该区土层深厚、肥力高，农业生态气候条件好，耕地利用强度大，土壤属性空间变异强烈。全区耕作田块分布较为零散，耕地面积小而破碎，土壤类型以水稻土为主，占耕地总面积的 70.97%，其次为赤红壤，占耕地总面积的 19.91%，而红壤、滨海盐土、潮土、风砂土、黄壤、石灰土和紫色土的分布面积较小，合计面积仅占总面积 10% 左右。总之，地貌的特殊性与复杂性导致气候、植被、土壤等空间分异显著，造成不同地区、不同海拔高程的耕地自然条件差异明显^[16]。

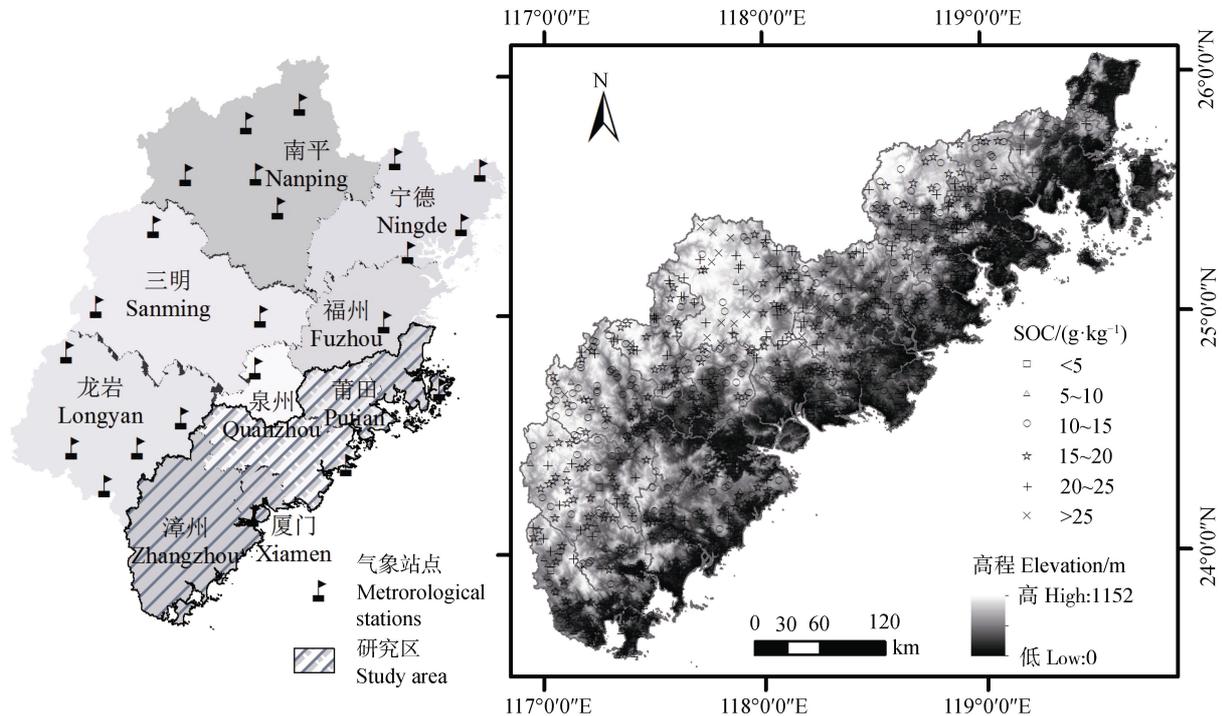


图1 研究区地理位置及采样点、气象站点分布

Fig. 1 Location of the study area and the distribution of soil sampling sites and meteorological stations

1.2 数据来源

本研究采用的土壤有机碳实测数据来源于国家农业农村部2017年末测土配方施肥调查样点数据,共计1257个,包括调查样点地理坐标及土壤有机质(Soil Organic Matter, SOM)、有效磷(AP)、速效钾(AK)、水解性氮(HN)、交换性镁(Ex-Mg)和pH等土壤理化属性(图1)。每个样点均按照密度控制、代表性、均匀性、优先性及适当性原则设置,采样深度为0~20 cm。在所选地块均匀随机采集10~15个耕层土样,充分混合后采用四分法留取1.5 kg土样装袋,自然风干过筛后备用。土壤有机质含量采用重铬酸钾氧化—外加加热测定,有效磷采用碳酸氢钠浸提—钼锑抗比色法测定,速效钾采用乙酸铵提取—火焰光度计法测定,水解性氮采用碱解扩散法测定,交换性镁采用乙酸铵浸提—原子吸收分光光度法测定,pH采用酸度计法测定。

选取样点中测定的土壤理化属性,联合遥感变量和气候变量作为SOC空间预测的环境变量,具体构成见表1。遥感变量中的植被指数和地形因子分别基于Landsat8 OLI影像和ASTER GDEM高程影像提取得到,数据来源于美国地质勘探局(United

States Geological Survey, <https://www.usgs.gov/>)和地理空间数据云网站(Geospatial Data Cloud, <http://www.gscloud.cn/>)。影像空间分辨率均为30 m,投影坐标系为WGS_1984_UTM_Zone_50N。为使植被指数能反映采样时地表真实状况,影像选取月份与实际调查样点获取时间一致,且天气晴朗,基本无云层覆盖。气象因子基于2017年福建全省22个标准气象站点的地面气候资料日值数据集插值而来(具体站点位置见图1),数据来源于国家气象中心网站(National Meteorological Information Center, <http://data.cma.cn/>)。

1.3 环境变量的获取、组合与筛选

为了探索多源协同变量解释SOC空间变异的可能性及土壤属性的加入对提高模型预测精度的贡献程度,本研究将表1中16个环境变量根据获取的难易程度分成两种不同的组合:(1)仅基于易获取的遥感变量(Remote Sensing Variables, RS)和气象因子(Climatic Factor, CF)训练的模型 Training model based on simple-to-obtain variables (RF-S);(2)基于遥感变量、气象因子和土壤属性(Soil Attribute, SA)所有变量训练的预测模型 Training model based on all variables (RF-A)。

表 1 环境变量的构成

Table 1 The composition of environmental variables

| 变量类别 Categories of variables | 指标 Indexes | |
|-----------------------------------------|-----------------------------------|-------------------------------------------------------|
| 遥感变量 Remote Sensing Variable (RS) | 植被指数 Vegetation Index (VI) | 归一化植被指数 Normalized Difference Vegetation Index (NDVI) |
| | | 转换植被指数 Transformed Vegetation Index (TVI) |
| | 地形因子 Topographical Factor (TF) | 数字高程模型 Digital Elevation Model (DEM) |
| | | 坡度 Slope、坡向 Aspect、曲率 Curvature |
| 气候因子 Climate Factor (CF) | | 地形湿度指数 Topographic Wetness Index (TWI) |
| | | 年最高气温 Max Annual Temperature (Maxt) |
| | | 年最低气温 Min Annual Temperature (Mint) |
| | | 年均气温 Mean Annual Temperature (Meant) |
| 土壤属性 Soil Attribute (SA) | | 年平均降水量 Mean Annual Rainfall (Rainfall) |
| | | 水解性氮 HN、有效磷 AP、速效钾 AK、交换性镁 Ex-Mg、pH |

影响土壤有机碳含量的环境变量众多，模型训练前需利用 RF 算法预测所产生的袋外误差的大小对部分变量进行剔除^[10]，即依据逐次剔除某一变量后 RF 模型袋外得分 (Out-of-bag Score, OOB Score) 的增减判断该变量是否保留，OOB Score 值增加则变量剔除，反之保留^[11]。

1.4 RF 模型构建和验证

以各模型筛选后的变量数据集为输入，利用 RF 模型进行有机碳含量的回归预测。RF 模型是建立在决策树基础上的一种集成学习方法，通过多次 bootstrap 抽样获取多个随机样本，并通过这些样本子集分别构建相应决策树，从而构建随机森林^[10]。当模型用于回归预测时，取所有决策树预测结果的均值作为最终的预测结果^[9]。模型的运算过程中需设定两个关键参数：n_estimators 和 max_depth，其中 n_estimators 为决策树的数量，即使用 bootstrap 重抽样的次数，max_depth 为决策树的最大深度。依据预测过程中产生的 OOB Score 的大小，RF-S 和 RF-A 模型设定的 (n_estimators, max_depth) 分别为 (1 400, 9) 和 (1 400, 12)。

为衡量两种不同变量组合下模型的表现，将所有训练样本点作为模型验证点进行验证。此外，使用不同抽样百分比 (20%、30%、40%、50%、60% 和 70%) 的验证数据集独立评估模型性能 (例如，80% 样点训练，余下 20% 样点进行验证，以此类推)，并与常用的地统计插值方法——普通克里格插值模型 (Ordinary Kriging, OK) 结果进行对比。使用平均绝对误差 (MAE)、均方根误差 (RMSE)、相关

系数 (r) 和变异系数 (CV) 评估模型预测的绝对误差表现，使用相对误差 (Relative Error, RE) 和相对均方根误差 (Relative RMSE, RRMSE) 定量化模型的准确性程度，数值越小，模型准确性越佳。

1.5 数据处理方法

数据处理主要为环境变量提取、RF 模型构建和验证，以及耕地土壤有机碳空间分布图的生成 3 个方面，具体如下：

环境变量提取。土壤定量属性采用 ArcGIS10.2 地统计插值模块中普通克里格法分别进行插值以获取 30 m × 30 m 栅格数据。遥感变量中的 NDVI 和 TVI 植被指数通过 Landsat8 OLI 影像波段运算得到，计算公式分别如下：

$$\frac{NIR-red}{NIR+red} \quad (1)$$

$$\left(\frac{NIR-red}{NIR+red} + 0.5 \right)^{\frac{1}{2}} \times 100 \quad (2)$$

式中，近红外 (Near Infrared, NIR) 和红光 (Red) 波段反射率为 Landsat8 OLI 影像利用 ENVI5.3 软件进行大气校正、镶嵌、裁剪等预处理后获取的研究区地表反射率。地形因子中的 DEM 由福建省 23 幅 ASTER GDEM 影像数据经坐标转换、镶嵌及研究区掩膜运算提取。而其他 4 个地形因子 (坡度、坡向等) 是利用 ArcGIS10.2 软件 Spatial Analyst 模块，基于 DEM 计算得到。气象因子栅格数据由全省气候资料数据集使用 ArcGIS10.2 地统计插值模块中反

距离权重法 (Inverse Distance Weighted, IDW) 插值后, 再通过研究区掩膜提取得到, 空间分辨率与遥感变量一致, 为 30 m。将上述环境变量形成的栅格数据集, 利用 ArcMap 提取到与土壤有机碳实测样点空间匹配的训练数据集中 (如图 1 所示), 用于 RF 模型的构建和验证。

RF 模型构建和预测的实现均通过 Python scikit-learn 库中 RandomForestRegressor 包实现。变量相对重要性排序可直接调用工具包中 feature_importances 属性实现。

基于 RF 模型和 OK 模型生成耕地土壤有机碳空间分布图, 用于评价不同模型在 SOC 空间异质性表达上的优劣。针对 RF 模型, 将空间分辨率为 30 m 的遥感变量、气候因子栅格和土壤属性栅格依据对应的变量组合分别输入 RF-S 与 RF-A 模型, 得到

SOC 空间分布格局。OK 模型则是基于所有 SOC 样本点, 使用普通克里格方法插值后重采样为 30 m 获取。利用 ArcGIS10.2 制图模块完成 SOC 空间分布专题制图。

2 结果

2.1 同变量组合下 RF 模型预测精度对比

如表 2 所示, 基于全部样本点, RF-A 和 RF-S 模型精度较好, 均呈现高度相关 ($r > 0.8$) 和中等变异水平 ($CV > 23\%$), 相对误差 RE 小于 10%。但与 RF-S 模型相比, 加入了土壤属性变量的 RF-A 模型预测误差有显著下降, RMSE 和 MAE 分别下降 45.13% 和 42.68%, 表明加入土壤属性变量 (N、P) 有利于提升模型拟合度及预测精度。

表 2 两种不同变量组合下 RF 模型的预测精度

Table 2 Prediction accuracies of RF models under two different combinations of variables

| 模型 Models | 环境变量组合 Combination of environmental variables | 训练集百分比 Percentage of training dataset /% | 验证集百分比 Percentage of validation dataset /% | RMSE / (g·kg ⁻¹) | MAE / (g·kg ⁻¹) | r | CV /% | RE /% | RRMSE /% |
|--------------|-----------------------------------------------------|------------------------------------------------|--------------------------------------------------|---------------------------------|--------------------------------|------|----------|----------|-------------|
| RF-S | RS+CF | 100 | 100 | 2.83 | 2.24 | 0.88 | 23.06 | -9.18 | 20.15 |
| RF-A | RS+CF+SA | 100 | 100 | 1.95 | 1.57 | 0.95 | 27.09 | -5.91 | 13.89 |

注: 模型输入环境变量详细信息见表 4。Note: Detailed information of environmental variables for model input is shown in table 4.

如表 3 所示, 两种 RF 模型预测值的均值与 SOC 实测值的均值非常接近, 约等于 14 g·kg⁻¹, 但预测结果范围明显被压缩, 变异系数和标准差值均变小。由图 2 也可发现, 两种模型在 SOC 低值区 (累积百分比 40% 以下) 略高估实测值, 而在 SOC 高值区 (累积百分比 75% 以上) 明显低估实测值, 但高值样点数 (SOC > 26 g·kg⁻¹) 所占比重较小, 所以两种 RF 模型的预测结果能够解释 SOC 大部分空间变异。总体而言, 加入了土壤属性变量的 RF-A 模型预测值

的累积分布图更接近实际结果, 能更好地表征区域 SOC 的动态变化范围。

2.2 RF 模型环境变量重要性

各模型最终筛选出的用于土壤有机碳预测的变量如表 4 所示, 各类别均有环境变量被保留参与模型构建。由表 4 可知, 在 RF-S 模型中, 重要性最高的环境变量为 DEM (23.22%), 气候因子分列重要性排序第二到五位, 虽然单个变量重要性稍弱, 但累积贡献率达 58.20%, 依旧占据主导地位。在加入土壤属性

表 3 两种不同变量组合下 RF 模型的预测结果与 SOC 实测值对比

Table 3 Comparison SOC measured value with prediction results of RF models under two different combinations of variables

| 模型 Models | 最小值 Min / (g·kg ⁻¹) | 平均值 Mean / (g·kg ⁻¹) | 最大值 Max / (g·kg ⁻¹) | 标准差 SD / (g·kg ⁻¹) | 变异系数 CV / % |
|--------------|------------------------------------|-------------------------------------|------------------------------------|-----------------------------------|----------------|
| SOC | 1.51 | 14.04 | 31.09 | 5.22 | 37.18 |
| RF-S | 3.82 | 14.05 | 26.20 | 3.24 | 23.06 |
| RF-A | 3.28 | 14.06 | 27.39 | 3.81 | 27.10 |

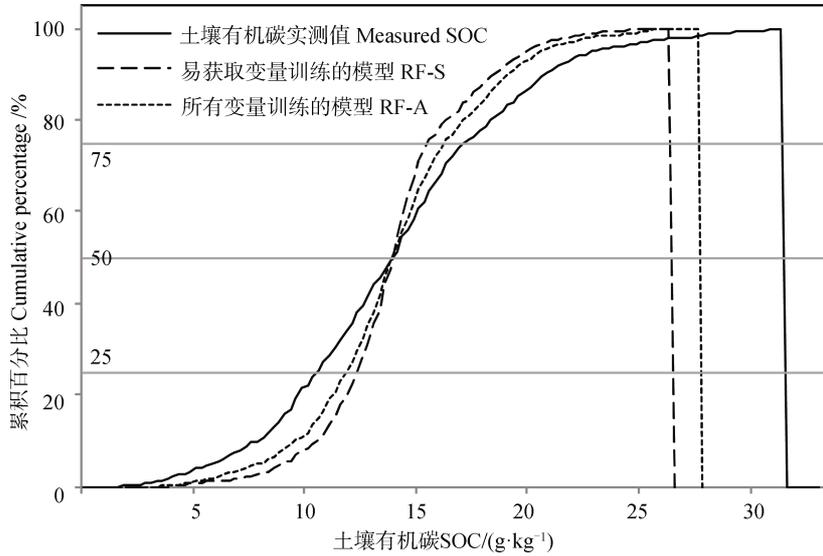


图 2 土壤有机碳实测值与两种不同变量组合模型预测值的累积分布图

Fig. 2 Cumulative distribution map of SOC measured value and predicted value of two different combinations of variables

表 4 RF 模型特征变量重要性排序

Table 4 Ranking of relative importance of environmental variables of RF model

| 模型 Models | 变量类别 Category of variables | 变量相对重要性 Relative importance of environmental variables | 模型 Models | 变量类别 Category of variables | 变量相对重要性 Relative importance of environmental variables |
|--------------|-------------------------------|-----------------------------------------------------------|--------------|-------------------------------|-----------------------------------------------------------|
| RF-S | RS-TF | 高程 DEM (23.22) | RF-A | SA | 水解性氮 N (22.83) |
| | CF | 年最低气温 Mint (16.64) | | RS-TF | 高程 DEM (16.47) |
| | CF | 年最高气温 Maxt (16.24) | | CF | 年最低气温 Mint (14.12) |
| | CF | 年均气温 Meant (13.45) | | CF | 年均气温 Meant (11.00) |
| | CF | 年均降水量 Rainfall (11.87) | | CF | 有效磷 P (10.56) |
| | RS-TF | 曲率 Curvature (7.16) | | CF | 年均降水量 Rainfall (8.49) |
| | RS-VI | 归一化植被指数 NDVI (5.73) | | CF | 年最高气温 Maxt (7.77) |
| | RS-VI | 转换植被指数 TVI (5.69) | | RS-VI | 归一化植被指数 NDVI (4.41) |
| | | | RS-VI | 转换植被指数 TVI (4.35) | |

注：括号中的数值为各模型中环境变量的相对重要性，标准化为 100%，按由高到低顺序排列。RS-TF，遥感变量中的地形因子；RS-VI，遥感变量中的植被指数；CF，气象因子；SA，土壤属性。√表示变量被保留参与构建 RF 模型，×表示变量被剔除。Note: The values in parentheses mean relative importance of environmental variables to each model, and are normalized to 100% and arranged in a descending order. RS-TF stands for the topographical factor included in remote sensing variables; RS-VI, for the vegetation index included in remote sensing variables; CF, for climate factors; and SA, for soil attributes. √ means variables retained and involved in establishment of the model, and × means variables excluded.

变量的 RF-A 模型中，N 的重要性超过 DEM，位列第一，说明这两个变量是影响闽东南地区耕地 SOC 空间变异的主要协同因子，且 N 和 SOC 的关系更为密切。值得注意的是，遥感变量中的植被指数 NDVI 和 TVI，全部通过变量筛选被保留参与模型预测，尽管单因素贡献较弱，对空间结果的预测也是不可或缺的。各类

环境变量中贡献率最高的因子分别为地形因子 DEM、植被指数 NDVI、气象因子 Mint 和土壤属性 N。

2.3 基于不同抽样百分比的精度检验

在不同梯度抽样百分比下，对两种 RF 模型以及 OK 模型的训练和预测精度进行检验，结果如表 5 所示。在训练数据集中，不同抽样百分比下，RF-S

表 5 基于 RF 模型与 OK 模型土壤有机碳模拟在训练数据集和验证数据集的精度对比

| 预测模型 Prediction models | Table 5 Comparison of prediction accuracies based on RF model and OK model in training and validation dataset | | | | | | | | | | | | | |
|------------------------------|---------------------------------------------------------------------------------------------------------------|------------------------------|-------------------------------|------|----------|--------------------------|-------------|---------------------------------|------------------------------|-------------------------------|------|----------|----------|-------------|
| | 训练数据集 Training dataset | | | | | 验证数据集 Validation dataset | | | | | | | | |
| | 梯度抽样 Gradient Sampling /% | MAE (g·kg ⁻¹) | RMSE (g·kg ⁻¹) | r | CV /% | RE /% | RRMSE /% | 梯度抽样 Gradient Sampling /% | MAE (g·kg ⁻¹) | RMSE (g·kg ⁻¹) | r | CV /% | RE /% | RRMSE /% |
| 随机森林 | 80 | 2.41 | 3.06 | 0.85 | 22.26 | -10.07 | 21.80 | 20 | 3.70 | 4.62 | 0.45 | 15.83 | -17.63 | 32.88 |
| | 70 | 2.81 | 3.57 | 0.75 | 19.27 | -11.77 | 25.55 | 30 | 3.90 | 4.90 | 0.43 | 16.26 | -17.30 | 34.50 |
| | 60 | 2.73 | 3.47 | 0.77 | 20.28 | -11.59 | 24.96 | 40 | 3.89 | 4.92 | 0.39 | 17.85 | -16.64 | 34.53 |
| | 50 | 2.42 | 3.03 | 0.84 | 21.99 | -10.13 | 21.82 | 50 | 3.83 | 4.90 | 0.40 | 18.02 | -16.37 | 34.51 |
| | 40 | 2.56 | 3.19 | 0.81 | 20.32 | -11.01 | 22.91 | 60 | 3.93 | 5.00 | 0.37 | 18.42 | -17.64 | 35.40 |
| | 30 | 2.78 | 3.49 | 0.77 | 19.20 | -12.59 | 25.09 | 70 | 3.84 | 4.89 | 0.39 | 18.00 | -18.07 | 34.68 |
| 随机森林 RF-A | 80 | 2.21 | 2.74 | 0.89 | 23.88 | -8.53 | 19.52 | 20 | 3.66 | 4.59 | 0.46 | 18.27 | -15.71 | 32.66 |
| | 70 | 1.94 | 2.40 | 0.92 | 24.47 | -7.34 | 17.18 | 30 | 4.62 | 4.62 | 0.53 | 18.81 | -14.27 | 32.53 |
| | 60 | 1.88 | 2.33 | 0.92 | 25.28 | -7.22 | 16.76 | 40 | 3.66 | 4.68 | 0.48 | 19.93 | -13.78 | 32.84 |
| | 50 | 2.07 | 2.57 | 0.90 | 24.19 | -7.95 | 18.51 | 50 | 3.62 | 4.64 | 0.49 | 19.44 | -14.28 | 32.68 |
| | 40 | 1.93 | 2.40 | 0.91 | 24.26 | -7.61 | 17.24 | 60 | 3.75 | 4.75 | 0.47 | 20.82 | -16.47 | 33.63 |
| | 30 | 2.31 | 1.88 | 0.92 | 24.78 | -7.68 | 13.52 | 70 | 3.68 | 4.66 | 0.48 | 20.74 | -16.36 | 33.05 |
| 普通克里格 OK | 80 | 3.06 | 3.99 | 0.65 | 21.51 | -11.87 | 28.42 | 20 | 3.65 | 4.56 | 0.47 | 17.34 | -16.88 | 32.45 |
| | 70 | 3.18 | 4.15 | 0.59 | 20.52 | -12.52 | 29.70 | 30 | 3.68 | 4.66 | 0.51 | 18.66 | -14.31 | 32.81 |
| | 60 | 3.14 | 4.09 | 0.61 | 20.93 | -12.34 | 29.41 | 40 | 3.66 | 4.67 | 0.49 | 19.93 | -14.02 | 32.77 |
| | 50 | 3.08 | 3.98 | 0.63 | 20.60 | -12.13 | 28.66 | 50 | 3.76 | 4.76 | 0.45 | 19.58 | -15.27 | 33.52 |
| | 40 | 3.22 | 4.14 | 0.57 | 19.24 | -13.00 | 29.74 | 60 | 3.76 | 4.76 | 0.46 | 18.90 | -16.32 | 33.70 |
| | 30 | 3.32 | 4.26 | 0.55 | 17.59 | -14.42 | 30.63 | 70 | 3.73 | 4.73 | 0.45 | 17.82 | -17.07 | 33.55 |

与 RF-A 模型预测的 SOC 值与实测值的相关系数 r 分别在 0.75~0.85 和 0.89~0.92 之间，而 OK 模型 r 处于 0.60 左右，个别情况下出现低值 0.55，整体精度明显低于 RF 模型。由误差 MAE、RMSE、RE 和变异系数 CV 值亦可发现，RF-A 模型在各梯度下计算误差最小，变异程度最大，OK 模型表现最弱。显然，RF-A 模型拟合精度最高，RF-S 次之，而 OK 模型最低。三种模型的拟合能力随训练样本量呈现不同的变化趋势：RF-A 模型受样本数量影响小，整体表现稳定；RF-S 模型小幅波动，规律性不强，在训练样本为 70% 时精度最低，其余情况较稳定；而 OK 模型波动较大，在训练样本为 80% 时表现最佳 ($r=0.65$)，30% 时表现最差 ($r=0.55$)。由于多种环境变量的协同作用，RF 模型鲁棒性较好，而 OK 模型对采样点要求较高，且插值结果随取样空间尺度增大会产生明显的平滑效应，这与陈飞香等^[20]使用克里格法对土壤原始样点在不同采样密度下的插值结果相一致。所以，当采样点较少时，应选择 RF 模型，并优先选含有土壤属性变量的 RF-A 模型。在验证数据集中，相较于其他两个模型，RF-S 整体精度不高，RF-A 模型计算的 r 除在 20% 和 40% 验证数据集较 OK 模型低 0.1，其他情况下均略高于 OK 模型。总体而言，RF-A 模型预测精度普遍优于 OK 模型，表现最好。

2.4 耕地土壤有机碳含量空间分布

如图 3a)、图 3b)、图 3c) 所示，RF-S、RF-A 和 OK 模型预测结果的空间分布均呈现为东部沿海较低、西部内陆较高。虽然三种模型的总体趋势比较相似，但在研究区南部和中部偏北地区，OK 模型与 RF-S 模型预测的 SOC 高值区明显小于 RF-A 模型。在上述预测差异较大区域随机选取子区域(图 3a~图 3c) a、b 框) 放大显示(图 3a~图 3c) 右边的子图 a、b)，可发现 RF-A 模型的 SOC 含量分级区间数明显多于其他两个模型，且空间变异更强。总体来看，RF-A 模型无论在模型精度或空间异质性表达上均为最优模型，以下仅对最优模型预测的 SOC 空间分布格局进行分析。

RF-A 模型反演得到闽东南区 SOC 均值为 $14.70 \pm 2.95 \text{ g}\cdot\text{kg}^{-1}$ ，范围为 $3.63\sim 25.51 \text{ g}\cdot\text{kg}^{-1}$ ，其中 $13\sim 19 \text{ g}\cdot\text{kg}^{-1}$ 区间的面积占比最高，超过研究区耕地总面积的 65%，主要分布在西部内陆闽中大山带戴云山-博平岭段东南侧；小于 $10 \text{ g}\cdot\text{kg}^{-1}$ 和大于 $19 \text{ g}\cdot\text{kg}^{-1}$

的面积占比较低，不足 10%，分别分布在闽东南地区三大平原(漳州平原、泉州平原、莆仙平原)和西部海拔最高地； $10\sim 13 \text{ g}\cdot\text{kg}^{-1}$ 区间所占面积在 19% 左右，位于高低值过渡带。

通过数值分析发现，SOC 空间特征与各类环境变量中贡献率最高的四个因子(DEM、NDVI、Mint、N) 呈现明显的相关性(如图 4 所示)。针对 DEM 和 Mint 两个因子，一般海拔越高，相应温度越低，而 SOC 含量正是与海拔高度呈正相关(图 4a)，而与年最低温度 Mint 呈负相关(图 4c)。这与杨顺华等^[21]的观点一致，认为平原丘陵过渡带土壤有机碳与高程等稳定因素呈极显著正相关。NDVI 值越大，植被覆盖度越高，SOC 含量越高(图 4b)，这是由于生物量也是土壤有机碳最重要的来源。值得注意的是，SOC 含量在一定范围内随着 N 含量的升高而增加，但当 SOC 含量均值 $>17 \text{ g}\cdot\text{kg}^{-1}$ 时，对应 N 含量呈下降趋势(图 4d)，可能原因是土壤碳氮比(C/N) 是一个相对固定的数值，土壤 N 素含量极大地影响 SOC 含量，然而在高海拔地区，虽氮素平均含量有所减少，但鉴于高海拔地区低温降水少的条件下有机质分解速率较低，SOC 含量并无下降。

3 讨论

3.1 闽东南地区土壤有机碳空间预测及主要环境变量影响

基于 RF-A 模型预测得到闽东南地区耕地 SOC 东部沿海较低、西部内陆较高，这与刘素真^[22]利用全国第二次土壤普查数据和近期野外采样数据开展的福建省有机碳含量模拟结果基本一致。西部地区因海拔高度较高伴随温度较低，用作农业生产的可能性降低，植被逐渐转化为枯落物较多的自然植被，同时土壤微生物分解有机质的速度减慢，矿化作用减弱，从而导致有机质的积累量逐渐增加^[10]，而东部低海拔区域，温度较高，农业生产较便利，频繁的耕作加剧土壤扰动，促进有机碳分解，使土壤有机碳周转速率加快，积累量减少，导致 SOC 含量较低，这与上述模型计算结果相一致。由此可见，RF-A 模型预测的研究区 SOC 含量空间分布格局是合理的，从这方面证明了该制图方法的可行性。

特征变量贡献率分析显示，水解性氮(N) 是预测土壤有机碳含量最重要的环境变量，这与预期

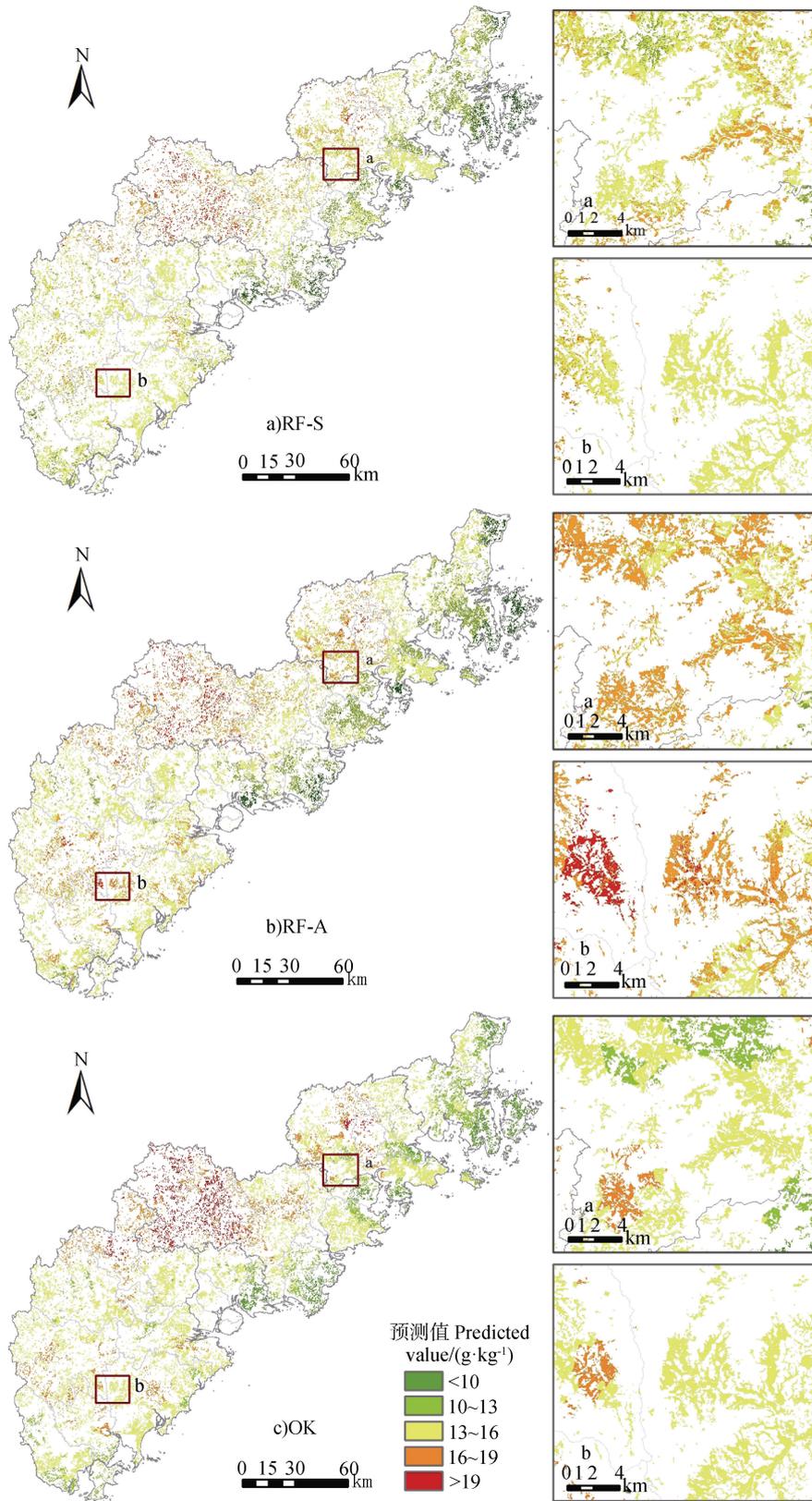


图3 基于 RF-S (a)、RF-A (b) 和 OK (c) 模型的闽东南地区耕地 SOC 空间分布
Fig. 3 SOC spatial distribution in Southeast Fujian estimated by RF-S model (a) RF-A model (b) and OK model (c)

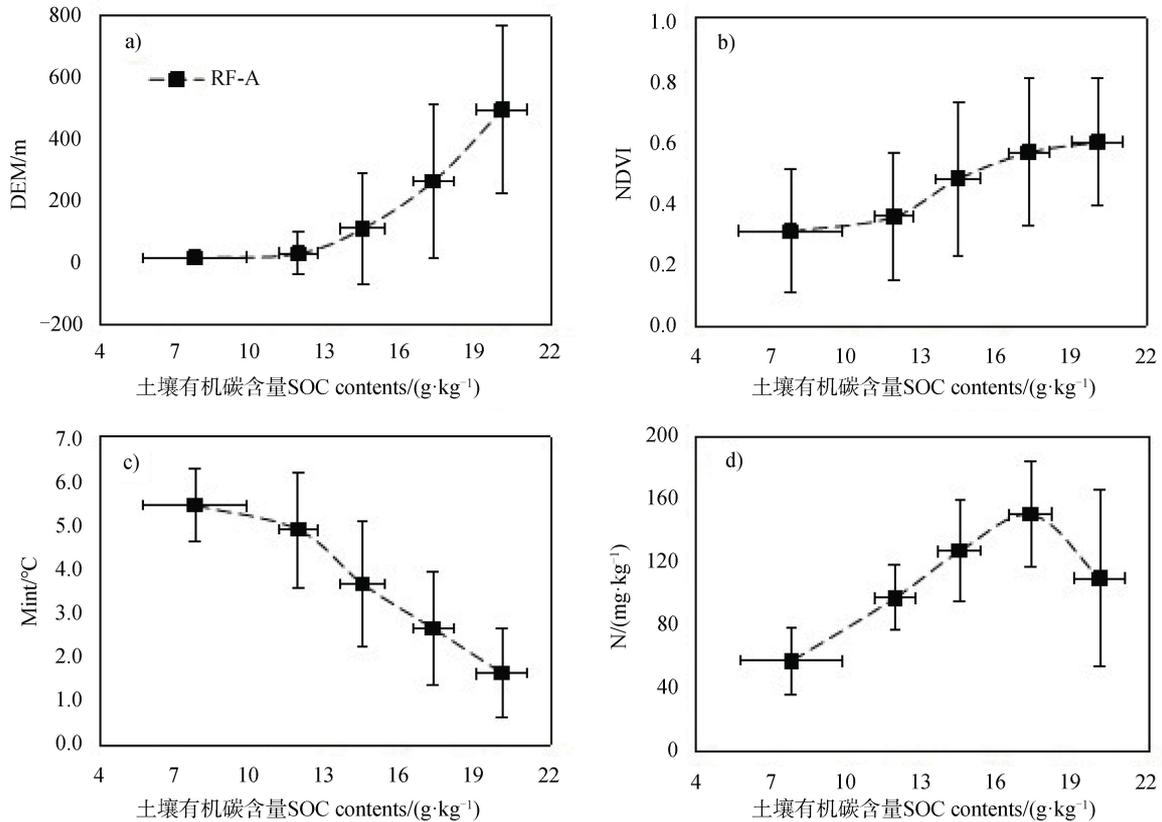


图 4 基于 RF-A 模型的土壤有机碳含量与代表性因子关系

Fig. 4 Comparison of soil organic carbon contents based on RF-A model and representative factors

一致。大量研究已证实陆地生态系统碳氮循环存在紧密耦合^[23]，碱解氮与土壤有机碳存在显著正相关关系^[24]，但因数据获取不易，以往较少用于土壤碳空间预测。Were 等^[9]和谢恩泽等^[25]在 RF 模型研究中也证明全氮 TN 是解释 SOC 空间变异最重要的变量，这与本文研究结论基本一致。

遥感变量中的 DEM 的重要性仅次于 N。但与 N 不同，DEM 是通过地形地貌差异间接影响 SOC 的空间分布，属于外部因素。与一般区域耕地分布于相对平缓且交通便利地带不同，本研究区位于亚热带复杂地貌区，耕地分布的海拔差异大，不同高度主要分布的土壤类型也有所不同，这就强调了 DEM 对 SOC 预测结果的影响。齐雁冰等^[10]在利用 RF 模型反演陕西省土壤有机质的研究中得出了相似的结果，即在地形地貌复杂且耕地面积较小的区域内，DEM 对 SOC 空间预测的贡献率相对较高。此外，高程和地形可通过影响降水情况间接影响 SOC 含量的空间分布。当来自太平洋的暖湿气流进入福建沿海并向西北运行过程中首先遇到闽中大山带，由于地形对气流的抬升作用，在东南坡产生较多的降

水，这与 SOC 的空间分布高度一致。由此可见，高程和降水、温度对 SOC 空间分布的影响存在重叠，并且高程的影响更大，所以降低了气候因子在模型中的重要性。

在基于遥感影像获取土壤有机碳的研究中，植被指数是最常用的变量之一，而本研究中所选取的两个植被指数在 RF 模型中的重要性并不突出，但二者均通过 OOB Score 的筛选被保留参与模型预测。从数值分析上看，SOC 含量大致与 NDVI 呈正相关关系（图 4b），这与以往研究相一致^[26]。

3.2 RF-A 模型精度

研究结果可得，RF-A 模型无论在模型精度或空间异质性表达均为闽东南区 SOC 空间预测的最佳选择。基于全部 1 257 个样点训练的 RF-A 模型的 $r=0.95$ ，RRMSE 值与 RE 值分别为 13.89% 和 -5.91%（表 2）。Hengl 等^[27]提出，RRMSE $\leq 40\%$ 以内为可以接受的模型准确性，当 RE $\leq \pm 10\%$ 时，模型模拟结果处于可接受范围。可见，RF-A 模型精确度高，且模型的准确性处于可接受范围。相较于 RF-S 模型，加入土壤属性因子的 RF-A 模型 RMSE 和 MAE

分别下降 45.13%和 42.68%，可见，RF 模型的预测精度受目标变量与辅助环境变量之间的相关性强弱控制，当加入与 SOC 紧密耦合的内部影响因子（如 N），RF 模型的精度显著提升。谢恩泽等^[25]在对比保留或移除辅助因子 TN 对苏南农田土壤有机质空间分布预测精度的影响时，得到了一致的结论，进一步证实了土壤因子用于亚热带复杂地貌区有机碳空间预测的可行性。

但 RF-A 模型在验证数据集中的 r 值（表 5）与以往文献^[10-11]相比，没有明显的优势。可能原因在于研究区为复杂地貌区，耕地面积小而破碎，SOC 预测难度较一般地势平坦、地形地貌单一的小区域大，且本文采用的大样本数据也会显著增加验证结果的误差。总体而言，RF-A 模型在全部样点和不同抽样百分比的预测精度均处于准确度范围内（不同梯度下 RRMSE 值均小于 34%，见表 5），且预测结果与实际相符，可用于研究区 SOC 空间分布格局的获取。

4 结 论

本文采用随机森林模型，选择易获取的遥感变量和气候因子及不易获取的土壤属性作为模型输入，基于大量土壤有机碳实测样点分别训练两种不同变量组合数据集驱动的 RF 模型，对比验证加入与未加入土壤属性变量构建的 RF 模型精度，并与普通克里格插值模型进行对比。结果显示联合遥感变量、土壤属性和气候因子共同构建 RF-A 模型精度最高，可作为预测该研究区 SOC 含量的高效方法，在模型驱动因子中加入 N、P 等土壤属性变量，能显著改进模型拟合度与精度，更好地提升区域 SOC 空间异质性的预测能力。相较于普通克里格插值模型，RF-A 综合模型在不同抽样百分比的预测精度普遍更优，且稳定性更好，因此在采样点较少的情况下，应优先选择 RF-A 模型。变量贡献率结果表明，影响研究区耕地 SOC 预测结果空间分布最重要的环境变量是与 SOC 直接相关的土壤属性——水解性氮，其次是 DEM。

参考文献 (References)

[1] Forkuor G, Hounkpatin O K L, Welp G, et al. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A

comparison of machine learning and multiple linear regression models[J]. *PLoS One*, 2017, 12(1): e0170478. <https://doi.org/10.1371/journal.pone.0170478>.

- [2] Zhang C T, Yang Y, He L Y, et al. Prediction of spatial distribution of soil organic matter based on environmental factors and a joint probability method[J]. *Acta Pedologica Sinica*, 2014, 51(3): 666–673. [张楚天, 杨勇, 贺立源, 等. 基于环境因子和联合概率方法的土壤有机质空间预测[J]. *土壤学报*, 2014, 51(3): 666–673.]
- [3] Sun X L, Zhao Y G, Zhao L, et al. Prediction and mapping of spatial distribution of soil attributes by using soil-landscape models[J]. *Soils*, 2008, 40(5): 837–842. [孙孝林, 赵玉国, 赵量, 等. 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, 40(5): 837–842.]
- [4] Jia S Y, Yang X L, Li G, et al. Quantitatively determination of available phosphorus and available potassium in soil by near infrared spectroscopy combining with recursive partial least squares[J]. *Spectroscopy and Spectral Analysis*, 2015, 35(9): 2516–2520. [贾生尧, 杨祥龙, 李光, 等. 近红外光谱技术结合递归偏最小二乘算法对土壤速效磷与速效钾含量测定研究[J]. *光谱学与光谱分析*, 2015, 35(9): 2516–2520.]
- [5] Zhou Z Y, Huang W, Xu W, et al. Updating traditional soil maps based on random forest algorithm[J]. *Journal of Huazhong Agricultural University*, 2019, 38(3): 53–59. [周紫燕, 黄魏, 许伟, 等. 基于随机森林算法的原始土壤图更新研究[J]. *华中农业大学学报*, 2019, 38(3): 53–59.]
- [6] Fang K N, Wu J B, Zhu J P, et al. A review of technologies on random forests[J]. *Statistics & Information Forum*, 2011, 26(3): 32–38. [方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. *统计与信息论坛*, 2011, 26(3): 32–38.]
- [7] Grimm R, Behrens T, Märker M, et al. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis[J]. *Geoderma*, 2008, 146(1/2): 102–113.
- [8] Grinand C, Maire G L, Vieilledent G, et al. Estimating temporal changes in soil carbon stocks at ecoregional scale in Madagascar using remote-sensing[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 54: 1–14.
- [9] Were K, Bui D T, Dick Ø B, et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape[J]. *Ecological Indicators*, 2015, 52: 394–403.
- [10] Qi Y B, Wang Y Y, Chen Y, et al. Soil organic matter prediction based on remote sensing data and random

- forest model in Shaanxi Province[J]. *Journal of Natural Resources*, 2017, 32 (6): 1074–1086. [齐雁冰, 王茵茵, 陈洋, 等. 基于遥感与随机森林算法的陕西省土壤有机质空间预测[J]. *自然资源学报*, 2017, 32 (6): 1074–1086.]
- [11] 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[J]. *Acta Pedologica Sinica*, 2016, 53 (2): 342–354. [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[J]. *土壤学报*, 2016, 53 (2): 342–354.]
- [12] Xu X W, Pan G X, Cao Z H, et al. A study on the influence of soil organic carbon density and its spatial distribution in Anhui Province of China[J]. *Geographical Research*, 2007, 26 (6): 1077–1086. [许信旺, 潘根兴, 曹志红, 等. 安徽省土壤有机碳空间差异及影响因素[J]. *地理研究*, 2007, 26 (6): 1077–1086.]
- [13] Ren L, Yang L A, Wang H, et al. Spatial prediction of soil organic matter in apple region based on random forest[J]. *Journal of Arid Land Resources and Environment*, 2018, 32 (8): 141–146. [任丽, 杨联安, 王辉, 等. 基于随机森林的苹果区土壤有机质空间预测[J]. *干旱区资源与环境*, 2018, 32 (8): 141–146.]
- [14] Liu S S, Yang Y H, Shen H H, et al. No significant changes in topsoil carbon in the grasslands of Northern China between the 1980s and 2000s[J]. *Science of the Total Environment*, 2018, 624: 1478–1487.
- [15] Guo P T, Li M F, Luo W, et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach[J]. *Geoderma*, 2015, 237/238: 49–59.
- [16] Xu W M, Luo X, Chen W F. Spatial distribution characteristics of arable land grade in Fujian Province[J]. *Journal of Fuzhou University (Natural Science Edition)*, 2018, 46 (3): 355–359. [徐伟铭, 罗星, 陈伟锋. 福建省耕地等别空间分布特征研究[J]. *福州大学学报(自然科学版)*, 2018, 46 (3): 355–359.]
- [17] 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 Chinese Society of Agricultural Engineering*, 2015, 31 (5): 194–202. [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): 194–202.]
- [18] Zhang W K. Study on the quality utilization of the cultivated land resource in Fujian Province[D]. Fuzhou: Fujian Normal University, 2002. [张文开. 福建省耕地资源优化利用[D]. 福州: 福建师范大学, 2002.]
- [19] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45 (1): 5–32.
- [20] Chen F X, Cheng J C, Hu Y M, et al. Spatial prediction of soil properties by RBF neural network[J]. *Scientia Geographica Sinica*, 2013, 33 (1): 69–74. [陈飞香, 程家昌, 胡月明, 等. 基于 RBF 神经网络的土壤铬含量空间预测[J]. *地理科学*, 2013, 33 (1): 69–74.]
- [21] Yang S H, Zhang H T, Chen J Y, et al. The spatial variability of soil organic carbon in plain-hills transition belt and its environmental impact[J]. *China Environmental Science*, 2015, 35 (12): 3728–3736. [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.]
- [22] Liu S Z. Estimation of soil organic carbon storage and its spatial distribution - A case of Fujian Province[D]. Beijing: Beijing Forestry University, 2016. [刘素真. 土壤有机碳储量估算及其空间分布——以福建省为例[D]. 北京: 北京林业大学, 2016.]
- [23] Lu M. The effects of nitrogen additions on ecosystem carbon and nitrogen cycles: A meta-analysis[D]. Shanghai: Fudan University, 2009. [卢蒙. 氮输入对生态系统碳、氮循环的影响: 整合分析[D]. 上海: 复旦大学, 2009.]
- [24] Zhang H D, You W Z, Wei W J, et al. Soil physical and chemical properties and correlation with organic carbon in original Korean pine forest in Eastern Liaoning mountainous area[J]. *Journal of Northwest A&F University (Natural Science Edition)*, 2017, 45 (1): 76–82. [张慧东, 尤文忠, 魏文俊, 等. 辽东山区原始红松林土壤理化性质及其与土壤有机碳的相关性分析[J]. *西北农林科技大学学报(自然科学版)*, 2017, 45 (1): 76–82.]
- [25] Xie E Z, Zhao Y C, Lu F Y, et al. Comparison analysis of methods for prediction of spatial distribution of soil organic matter contents in farmlands south Jiangsu, China[J]. *Acta Pedologica Sinica*, 2018, 55 (5): 1051–1061. [谢恩泽, 赵永存, 陆访议, 等. 不同方法预测苏南农田土壤有机质空间分布对比研究[J]. *土壤学报*, 2018, 55 (5): 1051–1061.]
- [26] Zhang W, Wang K L, Chen H S, et al. Use of satellite information and GIS to predict distribution of soil organic carbon in depressions amid clusters of Karst peaks[J]. *Acta Pedologica Sinica*, 2012, 49 (3): 601–606. [张伟, 王克林, 陈洪松, 等. 典型喀斯特峰丛洼地土壤有机碳含量空间预测研究[J]. *土壤学报*, 2012, 49 (3): 601–606.]
- [27] Hengl T, Heuvelink G B M, Stein A. A generic framework for spatial prediction of soil variables based on regression-kriging[J]. *Geoderma*, 2004, 120 (1/2): 75–93.

(责任编辑: 檀满枝)