

DOI: 10.11766/trxb202111020591

宋洁, 王思维, 赵艳贺, 于东升, 陈洋, 王鑫, 冯凯月, 马利霞. 消减残差自相关性的县域土壤有机质整合模型预测研究[J]. 土壤学报, 2023, 60 (6): 1569–1581.

SONG Jie, WANG Siwei, ZHAO Yanhe, YU Dongsheng, CHEN Yang, WANG Xin, FENG Kaiyue, MA Lixia. Soil Organic Matter Prediction Research on the Integrating Models with Reduction of Residual Autocorrelation[J]. Acta Pedologica Sinica, 2023, 60 (6): 1569–1581.

消减残差自相关性的县域土壤有机质整合模型预测研究*

宋洁^{1, 3}, 王思维², 赵艳贺², 于东升^{1, 3†}, 陈洋^{1, 3}, 王鑫^{1, 3},
冯凯月^{1, 3}, 马利霞¹

(1. 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008; 2. 河北省承德市滦平县农业农村局, 河北滦平 068250; 3. 中国科学院大学, 北京 100049)

摘 要: 提升土壤属性空间预测精度对实现农田精准施肥和保护生态环境具有重要意义。利用河北省滦平县采集的 1 773 个样点耕地表层(0~20 cm)土壤有机质(SOM)及其地理环境数据, 通过逐步回归分析方法筛选出最优环境变量; 基于其中 1 426 个农田样点分别建立多元线性回归(Multiple Linear Regression, MLR)、随机森林(Random Forest, RF)、贝叶斯正则化神经网络(Bayesian regularization neural network, BRNNBP)以及与普通克里格(OK)整合模型(MLR-OK、RF-OK 和 BRNNBP-OK)预测 SOM 空间分布, 以其余 347 个样点数据为测试集检验分析不同模型预测精度, 并对模型残差进行半方差函数和空间自相关分析以评价模型拟合效果。结果表明, 研究区耕地表层土壤 SOM 处在 8.62~35.64 g·kg⁻¹ 变化区间, 变异系数为 20.26%, 属中等程度空间变异; SOM 高值区主要分布在东北及东南海拔较高地区, 低值区多分布在西南及中部河谷地区; 海拔、坡度和年均温度与 SOM 关系密切($P<0.001$); 整合模型 BRNNBP-OK 的平均绝对误差 MAE 和均方根误差 RMSE 最低分别为 2.162 g·kg⁻¹ 和 2.801 g·kg⁻¹, 相较于 OK、MLR、RF、BRNNBP、MLR-OK 和 RF-OK 预测模型, R^2 提升 1.84%~43.72%, 成为 SOM 空间预测优选模型。与单一模型相比, 整合模型残差块金系数大于 0.75, Moran's I 指数均小于 0 且数值更趋近于 0, 表明整合模型残差空间自相关性减弱且空间分布呈离散状态。同时, 各模型精度与模型残差 Moran's I 指数呈显著相关。因此, 整合模型可以拟合更多的趋势项, 模型残差空间聚集性降低甚至趋于离散时, 模型总体精度提升, 揭示了模型精度提升的内在原因。

关键词: 土壤有机质; 机器学习; 普通克里格; 残差; 数字化土壤制图

中图分类号: X153.62 **文献标志码:** A

Soil Organic Matter Prediction Research on the Integrating Models with Reduction of Residual Autocorrelation

SONG Jie^{1, 3}, WANG Siwei², ZHAO Yanhe², YU Dongsheng^{1, 3†}, CHEN Yang^{1, 3}, WANG Xin^{1, 3}, FENG Kaiyue^{1, 3},

* 国家重点研发计划专项(2018YFC1800104) 和国家自然科学基金项目(42001302, 41571206)资助 Supported by the Special Project of the National Key Research and Development Program (No. 2018YFC1800104), the National Natural Science Foundation of China (Nos. 42001302, 41571206)

† 通信作者 Corresponding author, E-mail: dshyu@issas.ac.cn

作者简介: 宋洁, 女, 博士研究生, 主要从事数字土壤制图研究。E-mail: 2536646749@qq.com

收稿日期: 2021-11-02; 收到修改稿日期: 2022-06-06; 网络首发日期(www.cnki.net): 2022-09-09

MA Lixia¹

(1. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 2. Agricultural and Rural Bureau of Luanping County, Luanping, Hebei 068250, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: 【Objective】Improving the spatial prediction accuracy of soil attributes is of great significance for achieving accurate fertilization of farmland and protecting the ecological environment. 【Method】Soil organic matter (SOM) data was collected from 1 773 samples from soil surface layer (0–20 cm) of cultivated land in Luanping County, Hebei Province. The optimal environmental variables were screened through a stepwise regression analysis method. Multiple linear regression (MLR), ordinary kriging (OK), random forest (RF), Bayesian regularized neural network (BRNNBP), and the corresponding three integrated models combined with a geostatistical model (MLR-OK, RF-OK and BRNNBP-OK) were utilized to predict SOM content via the training set including 1 426 sampling points. Also, the prediction accuracy of each method was compared with 347 sampling points of the testing set. Autocorrelation analysis was carried out based on the residual of the integrated model to evaluate the fitting effect of the model. 【Result】Results showed that the range of SOM content in the study area was 8.62–35.64 g·kg⁻¹, and the coefficient of variation was 20.26%, which showed a moderate spatial variation. High concentrations of SOM were mainly distributed in the northeast and southeast areas with higher altitudes, while relative low concentrations of SOM were mostly observed in the southwest and central valleys of the study area. Elevation, slope and temperature selected by stepwise regression were closely related to SOM content ($P < 0.001$). The lowest average absolute error and the root mean square error of the BRNNBP-OK model were 2.162 g·kg⁻¹ and 2.801 g·kg⁻¹, respectively. Compared with the OK, MLR, RF, BRNNBP, MLR-OK and RF-OK models, the goodness of fit of the BRNNBP-OK model increased by 1.84%–43.72%, making it an optimal model for SOM spatial prediction. Compared with the single model, the nugget coefficient of the integrated model residual was greater than 0.75, and the Moran's I was less than 0 and numerically closer to 0, indicating that the spatial autocorrelation of the integrated model residual was weakened and the residual presented a more discrete spatial distribution. At the same time, the accuracy of all models was significantly correlated with Moran's index of model residuals. 【Conclusion】In this study, the integrated model fitted more trends and the spatial aggregation of model residuals decreased and even tended to be discrete. Thus, the overall prediction accuracy of the integrated models was improved.

Key words: Soil organic matter; Machine Learning; Ordinary Kriging; Residual; Digital soil mapping

土壤有机质 (SOM) 是衡量土壤肥力质量的重要指标^[1], 其空间分布特征影响土壤养分供应、土壤结构及土壤生态功能^[2]。由于成土过程受自然及人为活动等因素影响, 土壤 SOM 空间分布具有高度异质性^[3]。探索土壤 SOM 空间变异规律、提高其空间制图精度可为土壤质量的精细化管理提供科学依据。

土壤属性空间预测经典方法为地统计学的普通克里格 (Ordinary Kriging, OK) 方法^[4]。尽管 OK 方法^[5]充分考虑了土壤属性的空间自相关特征, 但忽略了地形和植被等环境因素对土壤属性空间分布的影响; 多元线性回归模型 (Multiple Linear Regression, MLR) 可定量评价环境因子对土壤属性的影响程度^[6], 但其基于土壤属性与环境因子之间普遍存在线性关系假设, 未能揭示内在非线性关系。因此, 传统方

法对提升土壤属性预测精度极为有限。

目前广泛流行的机器学习 (Machine Learning, ML) 方法, 如随机森林 (Random Forest, RF) 和反向传播 (Back Propagation, BP) 神经网络等, 被证明能有效刻画土壤属性与环境因子之间的非线性关系, 并提升土壤属性预测精度^[7-10]。其中, RF 具有对多元共线性不敏感及不易出现过拟合等优点, Chagas 等^[8]利用 RF 预测半干旱地区土壤砂粒含量, 模型精度 R^2 较多元线性回归模型提高了 10%。BP 神经网络可以通过梯度下降算法充分逼近复杂的非线性关系, 追求预测误差最小化或低于某个阈值^[9-10], 但易陷入局部最小值而造成过拟合^[11]。而贝叶斯正则化算法 (Bayesian regularization) 则通过修正网络的训练性能函数, 有效改善网络结构, 提升神经网络的泛化能力, 从而避免过拟合。余世鹏等^[12]分别

利用多元线性回归模型和 BP 神经网络模型预测土壤盐分含量, 结果表明 BP 神经网络模型较多元线性回归模型具有更高的预测精度。

然而, ML 仅能模拟环境因子与土壤属性空间分布之间的关系, 无法纳入土壤属性的空间自相关性信息。空间自相关是指同一变量在不同空间位置上的相关性, 是空间单元属性值聚集程度的一种度量, 可以揭示区域化变量取值的空间分布特征^[13]。因此有学者提出通过整合 ML 拟合的非线性模型与地统计学对 ML 模型产生残差的空间估计, 进而补充了土壤属性本身空间自相关信息, 更加真实反映复杂环境下土壤属性的空间变异^[14], 有助于提升土壤属性的空间预测精度。例如, Liu 等^[15]在土壤 As 预测中显示, RFOK 的预测能力优于单一 RF 或 OK 模型。李启权等^[16]在川中地区研究表明, 径向基神经网络克里格 (RBFNN-OK) 模型较 RBFNN 模型对土壤有机质的预测精度 RMSE 降低了 22.25%。

然而, 上述研究仅关注某一种机器学习方法与地统计的整合模型, 关于不同整合模型在 SOM 预测方面的性能差异仍不明确, 并且整合模型预测性能与其残差空间自相关性的关联尚不清楚。同时, 模型受环境因子、地域条件及采样数量等因素影响^[17], 对 MLR、RF 和 BRNNBP 模型与 OK 的整合模型区域适用性仍需充分研究和论证。

因此, 本研究以河北省承德市滦平县为研究区, 提取地形因子、气象因子、遥感光谱和微波数据等环境变量, 经逐步回归方法筛选出最优变量后采用 OK、MLR、RF、人工神经网络 BRNNBP 以及整合模型预测土壤 SOM, 探讨不同整合模型预测性能差异及其残差空间自相关性, 为提升土壤属性空间预测制图精度、实现精准施肥和保护生态环境提供技术支撑。

1 材料与方法

1.1 研究区概况

滦平县 (40°39'21"—41°12'53"N, 116°40'15"—117°46'03"E) 位于河北省东北部、承德市西部, 县域面积达 3 213 km²。全县群山环抱, 中部沟谷纵横, 海拔为 203~1 730 m。地貌单元主要包括中山、低山、丘陵、谷地等四种类型。属半干旱半湿润、大陆性季风型山地气候, 四季分明, 年平均气温 7.7℃, 年均降水量 351.1 mm, 全年无霜期 151 d。土地利

用类型以林地、耕地和园地为主, 其中耕地面积为 222.2 km²; 土壤类型主要为棕壤、褐土和草甸土, 成土母质包括残积-坡积物、黄土、洪积物、冲积物和人工堆垫等。农作物以玉米、谷类、大豆为主^[18]。

1.2 数据来源

滦平县 2007 年耕地土壤调查点位 1 773 个 (图 1), 采集和分析样点耕层 (0~20 cm) 土壤 SOM; 样点坐标采用手持 GPS 定位, SOM 采用重铬酸钾容量法测定^[19]。

研究区环境数据主要包括地形因子、气象因子、遥感光谱和微波数据。分辨率 30 m 数字高程模型 (DEM) 从地理空间数据云 (<http://www.gscloud.cn>) 获取, 利用 SAGA GIS 提取坡度 (Slope)、坡向 (Aspect)、地形因子 (LS)、高程 (Elevation)、剖面曲率 (Curvepro)、多尺度山谷平坦指数 (MrVBF)、地形位置指数 (TPI)、地表纹理 (Texture)、地表粗糙指数 (TRI)、谷深 (VD)、汇集指数 (CI) 和地形湿度指数 (TWI) 等地形因子; 从中国科学院资源环境科学数据中心 (<http://www.resdc.cn>) 获取 2007 年 1 km 气象数据, 通过重采样和裁剪得到年均温 (Tem) 和年降水量 (Pre) 数据; Landsat 5 TM 遥感影像数据从美国地质勘探局 (<https://earthexplorer.usgs.gov>) 获取, 利用 ENVI5.3 软件对影像进行辐射和大气校正后进行主成分变换, 获取前三个主成分分量 PCA1、PCA2、PCA3, 同时提取差值植被指数 (DVI)、比值植被指数 (RVI)、归一化植被指数 (NDVI)、归一化水体指数 (NDWI)、土壤调整植被指数 (SAVI) 和黏土矿物植被指数 (CMR)。此外, SAR 微波数据为 4 景 ALOS/PALSAR 影像数据, 实验选择 FBD 模式的 1.5 级产品, 利用 SNAP 软件完成辐射定标、滤波处理、地理编码、图像镶嵌和裁剪等预处理后提取后向散射系数 (HH 和 HV)。所有环境变量均利用 ArcGIS10.8 软件统一空间范围、投影坐标以及 30 m 分辨率。

1.3 特征变量优化选择

利用 SPSS 24.0 软件逐步回归方法筛选变量, 即每引入一个变量进行 F 检验, 并对已经进入模型的变量逐个进行 *t* 检验。当原引入变量由于后引入变量而变得不再显著时将其删除, 以确保每次引入新变量之前回归方程中只包含显著性变量, 从而有效解决变量间冗余和共线问题。使用显著性 F 检验的概率时, 设置 F 值大于 0.05 时变量进入方程, F 值小于 0.1 时则剔除变量。

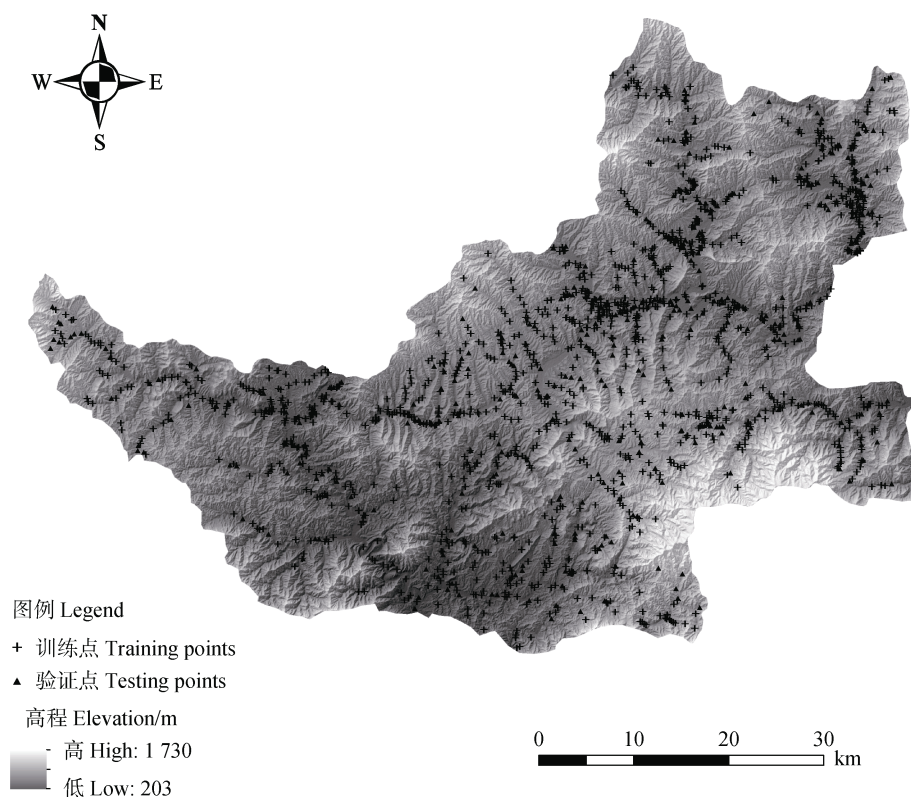


图 1 研究区高程图与土壤样点分布图

Fig. 1 Digital elevation map and distribution of soil sampling sites in the study area

1.4 SOM 预测模型

(1) 多元线性回归模型。多元线性回归 (MLR) 模型通过两个及两个以上的解释变量解释因变量的一种模型, 被广泛应用于土壤养分预测及影响因素等研究^[20], 其数学模型如下:

$$Y = a_0 + \sum_{i=1}^n a_i x_i = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n \quad (1)$$

式中, Y 为土壤 SOM 含量, x_i ($i=1, 2, 3, \dots, n$) 为各样点的环境变量, a_1, a_2, \dots, a_n 表示回归拟合系数, a_0 代表回归残差。MLR 建模借助 R 语言中的 `lm` 函数实现。

(2) 随机森林模型。随机森林 (RF) 模型是 Breiman 提出的由多棵 CART 决策树集成的机器学习算法, 可用于解决分类与回归问题^[7]。RF 模型回归分析原理为: (1) 采用 bootstrap 方法, 从原始样本中有放回地反复抽取 n 个样本训练集以构建 n 棵回归树 (ntree), 每次未被抽到的样本组成了袋外数

据 (OOB), 作为随机森林测试样本; (2) 从解释变量中抽取最能有效分割数据的 m 个解释变量 (mtry), 按照袋外预测误差最小原则确定 mtry 值; (3) 将生成的多棵回归树集成森林, 以所有 CART 预测值的平均值作为最终结果。借助 R 语言 `caret` 包和 `randomForest` 包构建模型。

(3) 贝叶斯正则化神经网络模型。贝叶斯正则化神经网络 (BRNNBP) 模型特点是修改神经网络误差函数, 即通过在样本数据外设置一定约束并以正则向的形式加入误差函数^[21]。一般神经网络的训练性能函数采用平方误差函数 E , 即

$$E = \frac{1}{N} \sum_{i=1}^n (m_i - o_i)^2 \quad (2)$$

式中, N 为样本总数, m_i 和 o_i 分别为第 i 次训练的网络期望输出值和实际输出值。

贝叶斯正则化算法中增加一个惩罚项, 则训练性能函数被调整为:

$$E_w = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (3)$$

$$F = \alpha E_w + \beta E \quad (\alpha + \beta = 1) \quad (4)$$

式中, $E_w = \frac{1}{n} \sum_{j=1}^n w_j^2$ 为网络权值, α 和 β 为正则化系数。

借助 R 语言 caret 包和 neuralnet 建立 BRNNBP, 设置初始值为 1, 步长为 1, 以 R^2 最大为约束条件确定最优神经元数。

(4) 克里格模型。克里格法 (OK) 是建立在变异函数理论及结构分析基础之上, 在有限区域内对变量进行的一种无偏最优估计^[22]。半方差函数又称变异函数, 用于定量描述土壤属性的空间变异结构, 其公式如下:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (5)$$

式中, h 为矢量距离; $N(h)$ 为相隔间距为 h 的点个数; $Z(x_i)$ 和 $Z(x_i + h)$ 分别为点 x_i 与 $x_i + h$ 的样点属性值。根据决定系数 (R^2) 最大、残差 (RSS) 最小确定最优半方差函数拟合模型。

(5) 整合模型。MLR-OK、RF-OK 和 BRNNBP-OK 是 MLR、RF、BRNNBP 分别与 OK 的整合模型 (式 (6) ~ 式 (8)), 即神经网络模型的不确定性趋势项与其局部回归残差 OK 预测项之和。

$$Y_{\text{MLR-OK}}^*(i) = Y_{\text{MLR}}^*(i) + \varepsilon_{\text{OK}}^*(i) \quad (6)$$

$$Y_{\text{RF-OK}}^*(i) = Y_{\text{RF}}^*(i) + \varepsilon_{\text{OK}}^*(i) \quad (7)$$

$$Y_{\text{BRNNBP-OK}}^*(i) = Y_{\text{BRNNBP}}^*(i) + \varepsilon_{\text{OK}}^*(i) \quad (8)$$

式中, $Y_{\text{MLR-OK}}(i)$ 、 $Y_{\text{RF-OK}}(i)$ 和 $Y_{\text{BRNNBP-OK}}(i)$ 为栅格 i 处 MLR-OK、RF-OK 和 BRNNBP-OK 模型 SOM 预测值; $Y_{\text{MLR}}(i)$ 、 $Y_{\text{RF}}(i)$ 和 $Y_{\text{BRNNBP}}(i)$ 为栅格 i 处 MLR、RF 和 BRNNBP 模型 SOM 预测值; $\varepsilon_{\text{OK}}(i)$ 为栅格 i 处 SOM 回归残差项的 OK 预测值。

1.5 模型残差空间自相关分析

空间自相关分析的目的在于确定某个变量在空间上是否相关, 通常采用 Moran's I 表示空间要素自相关程度^[23]。 I 的取值范围为 $-1 \sim 1$ 。 $I > 0$ 表示变量在

空间上呈正相关, 表现出空间集聚特征; 反之为负相关, 表现出离散分布特征; $I=0$ 表示无空间自相关性。通常用 Z 值检验自相关显著性, 当 $|Z| > 1.96$ 时, 变量的空间自相关显著, 否则不显著, 变量呈随机分布。借助 ArcGIS10.8 软件进行残差空间自相关分析。

$$\text{Moran's } I = \frac{n}{\sum_{ij} W_{ij}} \frac{\sum_{ij} W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (9)$$

式中, n 为预测栅格单元数量; W_{ij} 为表示空间关系的空间权重矩阵; x_i 和 x_j 分别为两个栅格单元模型预测残差的数值; \bar{x} 为模型预测残差的均值。

1.6 模型验证及精度评价

利用 R 软件从原始数据集里随机抽取 80% 样点数据作为建模集 (1426 个), 剩余 20% 为验证集 (347 个)。根据不同模型基于优化变量的 SOM 预测结果, 计算验证点处实测值与预测值之间的平均绝对误差 (MAE)、均方根误差 (RMSE) 及决定系数 (R^2)。其中, MAE 和 RMSE 值越小, R^2 越接近 1, 表明模型预测精度越高, 预测效果越好^[24-25]。

2 结 果

2.1 土壤有机质统计特征及优化变量

研究区耕地样点表层土壤 SOM 含量变化范围为 $8.62 \sim 35.64 \text{ g} \cdot \text{kg}^{-1}$, 平均值为 $17.17 \text{ g} \cdot \text{kg}^{-1}$, 标准差为 $3.48 \text{ g} \cdot \text{kg}^{-1}$, 空间变异系数为 20.26%, 属于中等程度变异, 偏度值相对较小 (表 1)。利用随机抽样方法划分的建模集与验证集的最小值、均值、标准差、变异系数等均较为统一, 表明样本划分合理。经 K-S 检验建模集和验证集均不符合正态分布 ($P < 0.05$), 对原始数据进行对数变换处理后符合正态分布。

土壤 SOM 与环境因子相关和方差分析 (图 2) 表明, 除 Aspect、CI、PCA3、HH 和 HV 外, 其余环境因子与土壤 SOM 含量均显著相关 ($P < 0.05$)。逐步回归方法筛选出 Elevation、Slope、VD、Tem、PCA2、PCA1、TRI 和 CMR 等 8 个变量为最佳变量组合, 判定系数 R^2 为 0.19。每个变量方差膨胀因子 VIF 值均小于 7.5, 表明变量间不存在多重共线性^[26]。

表 1 土壤有机质基本统计特征

Table 1 Basic statistical characteristics of soil organic matter

	样本数	最小值	最大值	均值	标准差	变异系数/%	偏度	峰度
	Number	Min	Max	Mean	Standard deviation	Coefficient of variation	Skewness	Kurtosis
总样本 Whole sets	1773	8.62	35.64	17.17	3.48	20.26	0.63	1.19
建模集 Calibration sets	1426	8.92	35.64	17.21	3.49	20.28	0.62	1.19
验证集 Validation sets	347	8.62	29.34	17.00	3.42	20.17	0.68	1.26

注：* $P < 0.05$ 。Note: * $P < 0.05$. SOM, Soil organic matter 土壤有机质; Elevation 高程; MrVBF, multiresolution index of valley bottom flatness 多尺度山谷平坦指数; Slope 坡度; TPI, Topographic position index 地形位置指数; Curvepro, Profile curvature 剖面曲率; Texture, Terrain surface texture 地表纹理; TWI, Terrain wetness index 地形湿度指数; VD, Valley depth 谷深; Aspect 坡向; LS, LS factor 地形因子; TRI, Terrain ruggedness index 地表粗糙指数; CI, Convergence index 汇集指数; Tem, Temperature 年均温; Pre, Precipitation 年降水量; NDVI, Normalized difference vegetation index 归一化植被指数; RVI, Ratio vegetation index 比值植被指数; DVI, Difference vegetation index 差值植被指数; SAVI, Soil-adjusted vegetation index 土壤调整植被指数; NDWI, Normalized difference water index 归一化水体指数; CMR, Clay minerals ratio 黏土矿物植被指数; PCA1、PCA2、PCA3 第一主成分、第二主成分、第三主成分; HH 和 HV, backscattering coefficient HH and HV 后向散射系数 HH 和 HV.

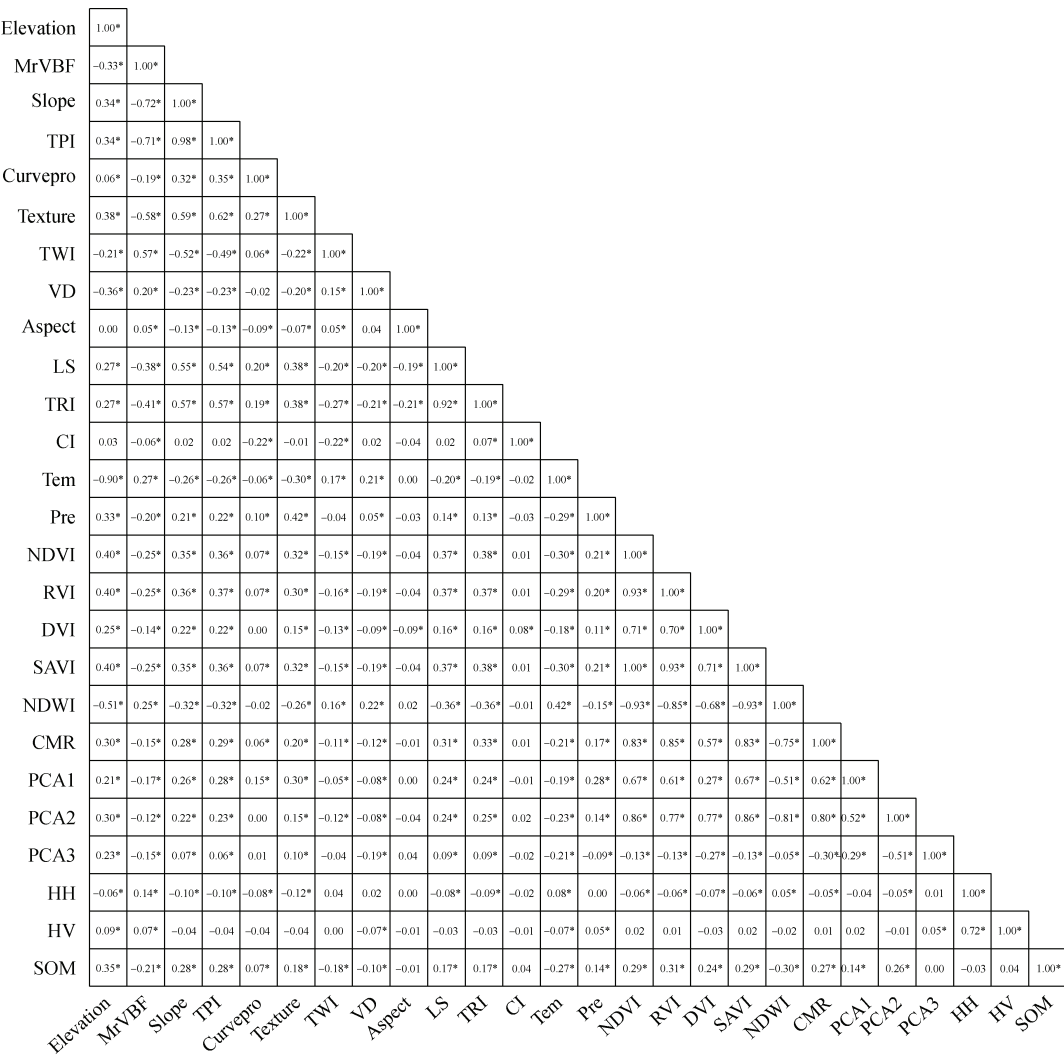


图 2 土壤有机质与环境因子的相关性

Fig. 2 Correlation between SOM and environmental factors

其中, Elevation、Slope、PCA1、PCA2、TRI 和 CMR 与土壤 SOM 呈显著正相关, 而 VD 和 Tem 与土壤 SOM 呈显著负相关(图 2)。

2.2 单一模型的土壤有机质空间预测精度

研究区 SOM 的块金系数为 0.66(表 2), 表明其具有中等程度空间自相关性^[27], 因此, SOM 空间分布差异由结构性因素(气候、母质、地形等)和随机性因素(耕作措施、施肥等)共同引起。单一模型中以 OK 模型预测精度最高(表 3), 且其残差的块金系数达 0.778(大于 0.75)(表 2), 表明 OK 模型充分利用土壤 SOM 空间自相关信息, 因而显示出弱的残差空间相关性。

RF 和 BRNNBP 模型用于拟合土壤 SOM 含量与环境变量之间非线性映射关系。与 MLR 模型相比, RF 和 BRNNBP 模型的 MAE 分别降低了 4.39%和

1.07%, RMSE 分别降低了 1.73%和 0.50%, R^2 分别增加了 11.69%和 3.90%(表 3)。模型残差的半方差函数分析结果表明, MLR、RF 和 BRNNBP 模型残差的块金系数分别为 0.716、0.730 和 0.718(表 2), 表明这些单一模型虽在一定程度上分离了 SOM 趋势项, 但其残差仍然具有空间自相关性^[27], 因此可以对模型残差进行最优 OK 空间估计。此外, 相较于 MLR 和 BRNNBP 模型, RF 模型残差空间自相关性更弱, 模型拟合精度更高。

2.3 整合模型的土壤有机质空间预测精度

在 MLR、RF 和 BRNNBP 建模基础上对其残差项进行 OK 空间插值, 并分别利用式(6)、式(7)和式(8)得到 SOM 预测值。与单一模型 MLR、RF 和 BRNNBP 相比, MLR-OK、RF-OK 和 BRNNBP-OK 整合模型的 MAE 分别降低了 10.76%、4.85%和 10.29%,

表 2 土壤有机质与拟合残差值的半方差函数
Table 2 Semivariance parameters of SOM and residuals

数据项 Data	模型 Model	变程 Range/km	块金值 C_0	基台值 C_0+C	块金效应 $C_0/(C_0+C) \%$
对数变换 LOGSOM	指数	22.59	0.029	0.044	65.91
普通克里格残差 OK residual	高斯	0.54	6.354	8.171	77.76
多元线性回归残差 MLR residual	指数	11.05	7.453	10.409	71.60
随机森林残差 RF residual	指数	12.55	1.571	2.151	73.04
贝叶斯正则化神经网络残差 BRNNBP residual	指数	11.25	7.435	10.354	71.81
多元线性回归克里格残差 MLR-OK residual	高斯	0.38	6.268	7.492	83.66
随机森林回归克里格残差 RF-OK residual	高斯	0.97	2.782	3.384	82.21
贝叶斯正则化神经网络克里格残差 BRNNBP-OK residual	指数	0.38	6.339	7.528	84.21

表 3 不同模型土壤有机质空间预测精度
Table 3 Evaluation of model accuracy for SOM prediction

模型 Model	平均绝对误差 MAE $/(\text{g}\cdot\text{kg}^{-1})$	均方根误差 RMSE $/(\text{g}\cdot\text{kg}^{-1})$	R^2
普通克里格 OK	2.265	2.932	0.265
多元线性回归 MLR	2.436	3.009	0.231
随机森林 RF	2.329	2.957	0.258
贝叶斯正则化神经网络 BRNNBP	2.410	2.994	0.240
多元线性回归克里格 MLR-OK	2.174	2.813	0.326
随机森林回归克里格 RF-OK	2.216	2.829	0.324
贝叶斯正则化神经网络克里格 BRNNBP-OK	2.162	2.801	0.332

RMSE 分别降低了 6.51%、4.33%和 6.45%，模型 R^2 分别提高了 41.13%、25.58%和 38.33%（表 3）。显然，MLR、RF 和 BRNNBP 模型在纳入残差信息后可有效提高 SOM 预测精度，其中 BRNNBP-OK 表现最优。

半方差函数分析表明，MLR-OK、RF-OK 和 BRNNBP-OK 整合模型残差块金系数均大于 0.75（表 2），均显示出弱的残差空间自相关性^[27]，表明整合模型可通过有效降低模型残差的空间自相关性提升拟合优度。

2.4 模型残差空间自相关性特征

模型残差空间自相关分析结果表明，各模型残差的 P 值均小于 0.05，且 Z 值均大于 1.96 或小于 -1.96，表明模型残差存在显著聚集或离散分布具有 95%置信度。OK 模型残差 Moran's I 指数为 -0.030，残差呈现空间离散分布，空间自相关性较弱，而 MLR、RF 和 BRNNBP 模型残差的 Moran's I 指数分

别为 0.118、0.036 和 0.116，表明模型残差具有空间正相关性，呈现聚集分布。整合模型（MLR-OK、RF-OK 和 BRNNBP-OK）残差的 Moran's I 指数分别为 -0.045、-0.039 和 -0.045，模型残差具有空间负相关性，呈现离散分布。同时，整合模型残差 Moran's I 指数更加趋近于 0，残差的空间自相关性减弱，说明整合模型拟合了更多结构性规律，与 2.2 中半方差函数分析结果一致（表 2）。

此外，各模型残差 Moran's I 指数与三种模型精度评价指标（ R^2 、MAE、RMSE）呈显著相关（图 3）。其中，模型残差 Moran's I 指数与 R^2 呈负相关，与 MAE 和 RMSE 呈正相关，表明模型拟合度越高，MAE 和 RMSE 越低，模型残差越趋于离散。显然，无论单一模型还是整合模型，提高土壤 SOM 空间预测精度，需降低模型残差的空间聚集性与自相关性。而这正是整合模型优于单一模型，有效提升土壤 SOM 预测精度的重要原因。

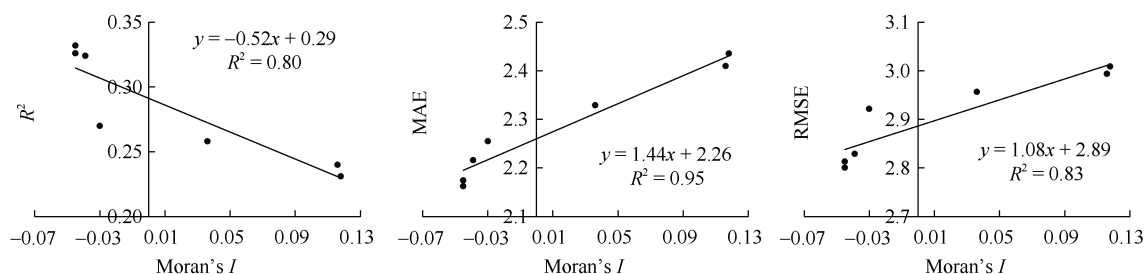


图 3 模型残差 Moran's I 与模型精度（ R^2 、RMSE、MAE）相关性

Fig. 3 Correlation between Moran's I of model residuals and model accuracy (R^2 , RMSE, MAE)

2.5 不同模型预测的土壤有机质空间分布特征

不同模型预测的 SOM 具有相似空间分布格局（图 4），SOM 高值区主要分布在东南部及东北角山地，低值区主要分布在地势较低的西南及中部沟谷地区，与研究区 Elevation、RVI、Slope 等环境变量的空间分布格局吻合。

不同模型预测土壤有机质高值区内均包含明显的低值部分，体现了土壤 SOM 含量随环境变量变化的微局域细节，但也具有明显差异。OK 模型预测的土壤 SOM 含量变化区间最小为 13.33~24.82 g·kg⁻¹，平滑效应最明显。与 MLR、RF 和 BRNNBP 单一模型相比，MLR-OK、RF-OK 和 BRNNBP-OK 整合模型预测的高低值分布区间更宽，其中 BRNNBP-OK 模型预测 SOM 含量变化区间为 12.45~32.71 g·kg⁻¹，与实测值的空间变化区间 8.62~35.64 g·kg⁻¹ 最为接

近，表明 BRNNBP-OK 模型预测效果最好。

3 讨 论

3.1 环境变量及其优选

土壤 SOM 含量空间分布受气候、母质、地形及人为等多种因素影响，准确预测 SOM 空间分布，应充分考虑不同环境因子对 SOM 含量的影响程度^[16]。因此主要选取地形因子、气象因子、遥感光谱数据和雷达数据。地形因子在地表物质能量循环过程中起重要作用，并影响土壤发生和演变过程^[28]。文中选取海拔、坡度、坡向和剖面曲率等因子，其中海拔和坡度是影响研究区土壤 SOM 空间分布的主要因素，SOM 含量随海拔升高、坡度增加而增大，可能是由于随着海拔升高或坡度增加，人为活动减少，有利于

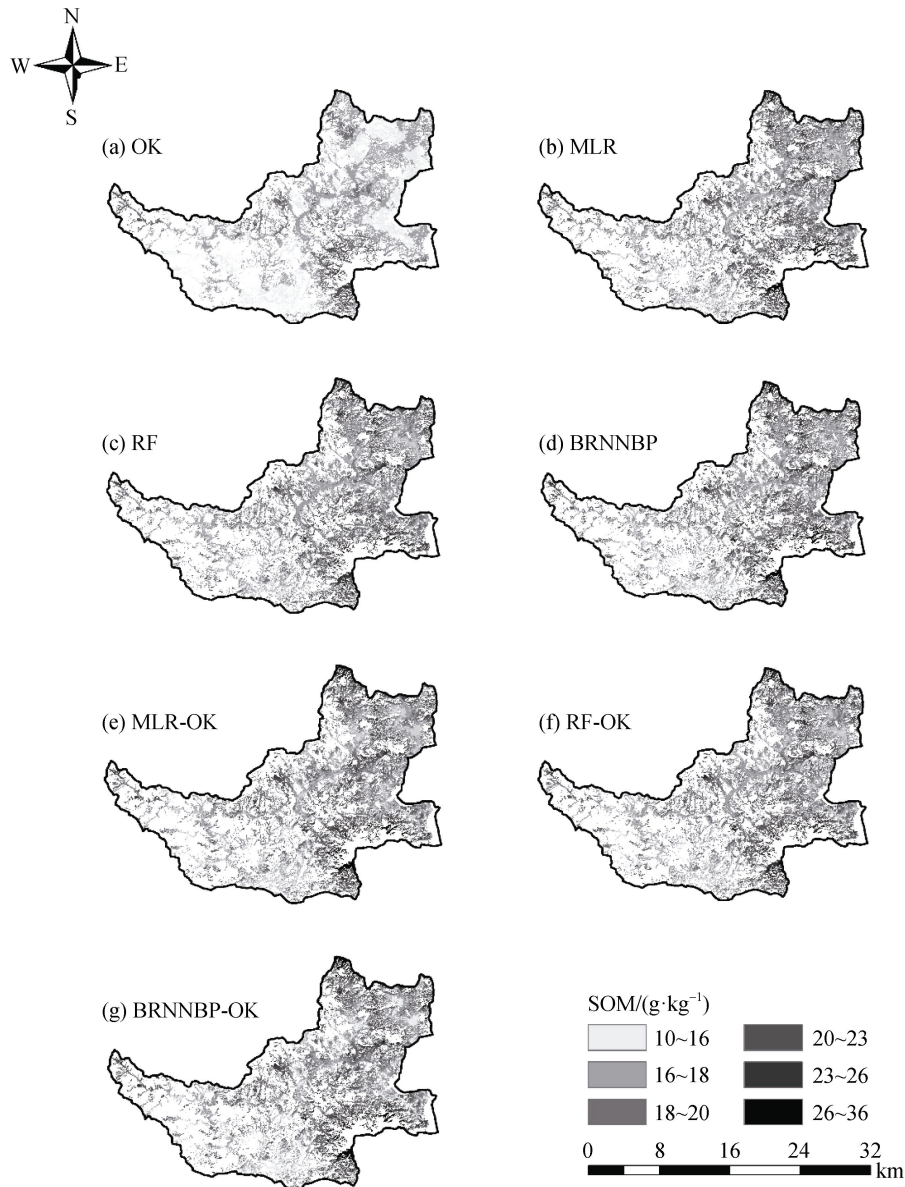


图 4 基于不同模型的土壤有机质空间分布图

Fig. 4 Spatial distribution of soil organic matter based on different models

SOM 的积累^[29]。气象因子选取年均温和年均降水量，主要是因为温度和降水对植物的光合作用、植被类型和植物残体分解过程等产生重要影响，进而影响土壤 SOM 含量^[30]。遥感光谱数据中的波段信息可以较好地反演土壤 SOM 含量，而 SAR 微波数据能够弥补多光谱数据易受云层、天气及地表覆被等影响^[31]，同时其特有的相位信息是其他传感器无法获取的。

随着越来越多变量引入模型，多维变量间的信息冗余和相关性会导致模型训练时间加长、预测精度出现偏差^[32]。因此，减少弱相关因子的干扰、确定合理的变量类型及数量有利于简化模型结构，提

高模型精度和稳定性^[33]。目前多数研究通过 Pearson 相关性分析或者主成分分析建立环境变量与土壤属性之间的关系，筛选出与土壤属性密切相关的特征变量^[34]。主成分分析主要目的在于降维，将多个指标归为少数的几个指标，这些指标一般是原来指标体系中某几个指标的线性组合，可以最大程度反映原始数据中的信息，但未能较好去除弱相关变量的干扰^[35]。本研究则利用逐步回归分析方法筛选出与 SOM 显著相关的 8 个特征变量 ($P < 0.05$)，较好去除弱相关变量的干扰，同时筛选出变量 $VIF < 7.5$ ，表明变量间不存在多重共线性^[36]。

3.2 整合模型预测精度提升及适用区域扩展

OK 模型充分考虑 SOM 空间自相关特征^[37], 因此残差空间自相关性较弱; 单一 MLR、RF 和 BRNNBP 模型虽然考虑了环境要素, 却不足以反映土壤 SOM 在空间位置上的自相关性^[15, 38], 因此单一模型残差的块金系数均小于 0.75, 且 Moran's *I* 指数分别为 0.118、0.036 和 0.116, 表明模型残差仍具有空间自相关性且呈现空间聚集状态; 而 MLR-OK、RF-OK 和 BRNNBP-OK 整合模型不仅考虑 SOM 与多元环境因子的非线性关系, 还纳入单一模型残差的空间自相关性信息, 从而有效捕捉复杂环境下 SOM 的空间变异^[16]。整合模型残差块金系数均大于 0.75 且 Moran's *I* 指数分别为 -0.045、-0.039 和 -0.045, 与单一模型 MLR 和 BRNNBP 相比, 整合模型残差 Moran's *I* 指数更趋近于 0, 表明整合模型残差空间自相关性降低且空间分布上趋于离散状态。类似地, Shahriari 等^[39]对伊朗东南部锡斯坦洪泛区土壤粉粒的预测研究显示, 回归克里格 (RK) 和回归克里格-神经网络残差克里格 (RKNNRK) 模型残差的块金系数为 21% 和 0.09%, 仍然存在较强的空间相关性, 而神经网络残差克里格 (NNRK) 模型残差空间自相关性相对较弱, 同时具有最低 RMSE。

同时, 各模型精度和其残差空间自相关 Moran's *I* 指数显著相关, 即当模型拟合出更多趋势性规律时, 其残差空间自相关性减弱, 空间分布趋于离散状态, 模型拟合精度更高。各模型 Moran's *I* 指数排序为 $MLR > BRNNBP > RF > 0 > RF-OK > BRNNBP-OK = MLR-OK$, 与 MLR、RF、BRNNBP 和 RF-OK 模型相比, MLR-OK 和 BRNNBP-OK 模型残差空间分布更加离散, 同时两种模型 MAE 分别降低 1.90%~10.76% 和 2.44%~11.25%, RMSE 分别降低 0.57%~6.51% 和 0.99%~6.91%, R^2 分别提升 0.62%~41.13% 和 2.47%~43.72%。该结果揭示了模型间精度差异与其残差空间分布关系特征, 为有效提高土壤 SOM 预测精度提供依据。

结合制图研究区域发现, 纳入空间自相关信息的整合模型在土壤属性预测方面的优势未局限于洪泛^[39]、平原^[40]、高原^[41]、山地丘陵^[15]等景观环境。研究区为典型北方土石山区, 农田分布破碎、人为扰动强烈, SOM 具有较强空间异质性和低空间相关性, 整合模型有效弥补了空间自相关性较弱引起的

机器学习模型的局部欠拟合、不稳健的缺陷。整合模型提升 SOM 预测精度的机制分析表明, 适用区域可拓展至中国北方土石山区、青藏高原地区^[41]、川中丘陵区^[15]等以外地区。

3.3 整合模型土壤样点数量需求

样点数量对地统计 OK 模型土壤 SOM 预测结果有影响^[42]。苏晓燕^[43]研究表明普通克里格 OK 模型随样点数量有限增多, 局部变异更详细。但一味增加样本数量, 不仅导致人力、物力及时间成本等增加, 而且导致类似样点数据信息增加或雷同, 空间预测精度并不能得到明显提升, 甚至影响土壤信息全局输出结果的可靠性^[44]。因此, 一味增加样品数量并不能成为 OK 模型显著提升 SOM 空间预测精度的有效途径, 必须寻求各种预测模型方法的优势及其整合^[45-46]。

Khaledian 和 Miller^[47]研究发现, 不同的机器学习模型预测精度对样本数量大小有着不同的响应, Cubist 模型树和 RF 模型对样本数量不敏感, 但小样本数量会限制人工神经网络模型预测准确性。本文耕地样点数量为 1 773, 样点密度约为 $1.50 \text{ 个} \cdot \text{km}^{-2}$, 样本数量不仅适应普通克里格 OK 预测方法特点, 也满足 BRNNBP 神经网络学习方法的大样本要求。

因此, 在未来整合模型研究中, 应在研究区尺度基础上确定合理的样点数量, 充分发挥 OK 和机器学习整合模型优势, 提升工作效率的同时最大程度降低成本。

4 结 论

本文以我国北方土石山区滦平县为研究区, 采用单一模型 (OK、MLR、RF、BRNNBP) 和整合模型 (MLR-OK、RF-OK 和 BRNNBP-OK) 预测土壤 SOM 空间分布, 比较分析各模型精度差异及其残差空间分布特征。主要结论如下: (1) 海拔、坡度、谷深、地表粗糙指数、年均温、黏土矿物比值指数、光谱因子第二主成分和第一主成分为预测 SOM 空间分布的最优变量组合。(2) BRNNBP-OK 整合模型为最优模型, R^2 提高 1.84%~43.72%, MAE 及 RMSE 降低 0.55%~11.25%、0.43%~6.91%。(3) 与单一模型相比, 纳入残差空间自相关信息的整合模型空间预测精度提升, 残差空间自相关性降低, 空间分布趋于离散。

参考文献 (References)

- [1] Zhang L, Hao M J, Lu X X, et al. Study on estimation model of soil organic carbon content of farmland in southern Xinjiang based on spectral data[J]. Journal of Tarim University, 2020, 32 (4): 49—58. [张丽, 郝梦洁, 鲁新新, 等. 基于光谱数据的南疆农田土壤有机碳含量估算模型研究[J]. 塔里木大学学报, 2020, 32 (4): 49—58.]
- [2] Lü Y Z, Li B G. Soil science[M]. Beijing: China Agriculture Press, 2006. [吕贻忠, 李保国. 土壤学[M]. 北京: 中国农业出版社, 2006.]
- [3] Zhang N, Zhang H L, Qu Z Y, et al. Fractal study on organic matter spatial heterogeneity of different soil layers in Inner Mongolia Hetao Irrigation District[J]. Agricultural Research in the Arid Areas, 2017, 35 (5): 157—163. [张娜, 张红玲, 屈忠义, 等. 内蒙古河套灌区不同土层有机质空间变异的分形[J]. 干旱地区农业研究, 2017, 35 (5): 157—163.]
- [4] Jiang S P, Zhang H Z, Zhang R L, et al. Research on spatial distribution of soil organic matter in Hainan Island based on three spatial prediction models[J]. Acta Pedologica Sinica, 2018, 55 (4): 1007—1017. [姜赛平, 张怀志, 张认连, 等. 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[J]. 土壤学报, 2018, 55 (4): 1007—1017.]
- [5] Xia Z S, Bai Y R, Bao W B, et al. Spatial distribution of soil organic carbon in Shizuishan based on multispectral and geographically weighted regression model[J]. Arid Land Geography, 2020, 43 (5): 1348—1357. [夏子书, 白一茹, 包维斌, 等. 基于多光谱和地理加权回归模型的石嘴山城市土壤有机碳空间分布研究[J]. 干旱区地理, 2020, 43 (5): 1348—1357.]
- [6] Han D, Cheng X F, Xie J H, et al. Spatial variability of soil organic matter in jiangzihe watershed of dabie mountainous area and its influencing factors[J]. Acta Pedologica Sinica, 2012, 49 (2): 403—408. [韩丹, 程先富, 谢金红, 等. 大别山区江子河流域土壤有机质的空间变异及其影响因素[J]. 土壤学报, 2012, 49 (2): 403—408.]
- [7] Breiman L. Random forests[J]. Machine Learning, 2001, 45 (1): 5—32.
- [8] da Silva Chagas C, de Carvalho W Jr, Bhering S B, et al. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions[J]. Catena, 2016, 139: 232—240.
- [9] Khosravi V, Doulati Ardejani F, Yousefi S, et al. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods[J]. Geoderma, 2018, 318: 29—41.
- [10] Bao H Y, Wang J F, Li J, et al. Effects of corn straw on dissipation of polycyclic aromatic hydrocarbons and potential application of backpropagation artificial neural network prediction model for PAHs bioremediation[J]. Ecotoxicology and Environmental Safety, 2019, 186: 109745.
- [11] Xu Q F, Yu R Y, Gou Y X, et al. Prediction precision analysis of soil organic matter based on cloud genetic BP neural network in Huang-Huai-Hai dry farming area[J]. Journal of China Agricultural University, 2021, 26 (4): 167—173. [徐清风, 于茹月, 勾宇轩, 等. 基于云遗传 BP 神经网络的黄淮海旱作区土壤有机质预测精度分析[J]. 中国农业大学学报, 2021, 26 (4): 167—173.]
- [12] Yu S P, Yang J S, Liu G M, et al. Simulation and prediction of soil salt dynamics in the Yangtze River Estuary with BP artificial neural network[J]. Soils, 2008, 40 (6): 976—979. [余世鹏, 杨劲松, 刘广明, 等. 基于 BP 人工神经网络的长江河口地区土壤盐分动态模拟及预测[J]. 土壤, 2008, 40 (6): 976—979.]
- [13] Yang L P, Cheng J C, Yang X W, et al. Research on the effects spatial autocorrelation exerts on the interpolation of soil properties[J]. Guangdong Agricultural Sciences, 2012, 39 (20): 40—43, 2. [杨丽萍, 程家昌, 杨小威, 等. 空间自相关性对土壤属性插值的影响[J]. 广东农业科学, 2012, 39 (20): 40—43, 2.]
- [14] Chen S M, Wang N, Qin Y F, et al. Study on spatial variability of SOM in estuary wetland, southeast China based on characteristic variables and SVRK[J]. Soils, 2020, 52 (6): 1298—1305. [陈思明, 王宁, 秦艳芳, 等. 基于特征变量与支持向量机回归克里格 (SVRK) 法的湿地土壤有机质空间变异特征分析[J]. 土壤, 2020, 52 (6): 1298—1305.]
- [15] Liu G, Zhou X, Li Q, et al. Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data[J]. Environmental Pollution, 2020, 267: 115631.
- [16] 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[J]. Advances in Earth Science, 2012, 27 (2): 175—184. [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[J]. 地球科学进展, 2012, 27 (2): 175—184.]
- [17] Lai Y Q, Sun X L, Wang H L. Mapping of soil organic carbon using neural network and its mixed model with geostatistics in a small area of typical hilly region[J]. Chinese Journal of Soil Science, 2020, 51 (6): 1313—1322. [赖雨晴, 孙孝林, 王会利. 人工神经网络及其与地统计的混合模型在小面积丘陵区土壤有机碳预测制图上的应用研究[J]. 土壤通报, 2020, 51 (6): 1313—1322.]
- [18] Ma X T. Evaluation of soil erosion sensitivity in Luanping County[D]. Beijing: Beijing Forestry University, 2020. [马

- 晓彤. 滦平县土壤侵蚀敏感性评价[D]. 北京: 北京林业大学, 2020]
- [19] Zhang G L, Gong Z T. Soil survey laboratory methods[M]. Beijing: Science Press, 2012. [张甘霖, 龚子同. 土壤调查实验室分析方法[M]. 北京: 科学出版社, 2012.]
- [20] 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.]
- [21] Li H X, Xu S G, Fan C R. Long-term prediction of runoff based on Bayesian regulation neural network[J]. Journal of Dalian University of Technology, 2006, 46 (S1): 174—177. [李红霞, 许士国, 范垂仁. 基于贝叶斯正则化神经网络的径流长期预报[J]. 大连理工大学学报, 2006, 46 (S1): 174—177.]
- [22] Yu W X, Zhao M S, Wang M, et al. Effects of sampling sizes and spatial interpolation methods on prediction accuracy of soil properties[J]. Science Technology and Engineering, 2017, 17 (25): 186—191. [于伟宣, 赵明松, 王萌, 等. 采样数量与空间插值方法对土壤属性预测精度的影响[J]. 科学技术与工程, 2017, 17 (25): 186—191.]
- [23] Zuo X H, Lai J X, Liu F, et al. Spatial heterogeneity and zoning of soil organic matter based on geostatistics and spatial autocorrelation: A perspective from land consolidation[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2022, 43 (3): 240—252. [左昕弘, 赖佳鑫, 刘峰, 等. 基于地统计学和空间自相关的土壤有机质空间异质性分析及分区: 土地整治视角[J]. 中国农业资源与区划, 2022, 43 (3): 240—252.]
- [24] Chai T, Draxler R R. Root mean square error (RMSE) or mean absolute error (MAE) ? —Arguments against avoiding RMSE in the literature[J]. Geoscientific Model Development, 2014, 7 (3): 1247—1250.
- [25] Miller F P, Vandome A F, Mcbrewhster J. Coefficient of determination[J]. Alphascript Publishing, 2006, 31 (1): 63—64.
- [26] Sun Y S, Wang W F, Li G C. Spatial distribution of forest carbon storage in Maoershan region, Northeast China based on geographically weighted regression kriging model[J]. Chinese Journal of Applied Ecology, 2019, 30 (5): 1642—1650. [孙钰森, 王维芳, 李国春. 基于地理加权回归克里格模型的帽儿山地区森林碳储量空间分布[J]. 应用生态学报, 2019, 30 (5): 1642—1650.]
- [27] Wang D C, Wu D W, Zhao M S, et al. Prediction and mapping of soil texture of a plain area using reflectance spectra and geo-statistics[J]. Chinese Journal of Soil Science, 2012, 43 (2): 257—262. [王德彩, 郭登巍, 赵明松, 等. 平原区土壤质地的反射光谱预测与地统计制图[J]. 土壤通报, 2012, 43 (2): 257—262.]
- [28] Zhou Y, Hartemink A E, Shi Z, et al. Land use and climate change effects on soil organic carbon in North and Northeast China[J]. Science of the Total Environment, 2019, 647: 1230—1238.
- [29] Zhang W T, Ji J Y, Li B B, et al. Spatial prediction of soil organic matter of farmlands under different landforms in the Loess Plateau, China[J]. Journal of Plant Nutrition and Fertilizers, 2021, 27 (4): 583—594. [张万涛, 吉静怡, 李彬彬, 等. 黄土高原不同地貌区农田土壤有机质预测方法研究[J]. 植物营养与肥料学报, 2021, 27 (4): 583—594.]
- [30] Zhou Y P, Zhang Y C, Luo X Y, et al. Review on spatial variability of soil organic matter and its driving factors[J]. Chinese Journal of Soil Science, 2019, 50 (6): 1492—1499. [周一鹏, 张雨辰, 罗鑫叶, 等. 土壤有机质空间变异性及其驱动因素研究进展[J]. 土壤通报, 2019, 50 (6): 1492—1499.]
- [31] Wang X, Zhang Y H, Atkinson P M, et al. Predicting soil organic carbon content in Spain by combining Landsat TM and ALOS PALSAR images[J]. International Journal of Applied Earth Observation and Geoinformation, 2020, 92: 102182.
- [32] Hu G G, Yang F L, Yang L A, et al. Spatial prediction modeling of soil organic matter content based on principal components and machine learning[J]. Arid Land Geography, 2021, 44 (4): 1114—1124. [胡贵贵, 杨粉莉, 杨联安, 等. 基于主成分和机器学习的土壤有机质含量空间预测建模[J]. 干旱区地理, 2021, 44 (4): 1114—1124.]
- [33] Wang F, Yang S T, Ding J L, et al. Environmental sensitive variable optimization and machine learning algorithm using in soil salt prediction at oasis[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34 (22): 102—110. [王飞, 杨胜天, 丁建丽, 等. 环境敏感变量优选及机器学习算法预测绿洲土壤盐分[J]. 农业工程学报, 2018, 34 (22): 102—110.]
- [34] Jiang Y F, Guo X. Prediction of soil organic matter distribution based on auxiliary variables and regression-radial basis function neural network (R-RBFNN) model[J]. Acta Agriculturae Zhejiangensis, 2018, 30 (4): 640—648. [江叶枫, 郭熙. 基于辅助变量和回归径向基函数神经网络 (R-RBFNN) 的土壤有机质空间分布模拟[J]. 浙江农业学报, 2018, 30 (4): 640—648.]
- [35] Ye Q, Jiang X Q, Li X C, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48 (3): 164—172. [叶勤, 姜雪芹, 李西灿, 等. 基于高光谱数据的土壤有机质含量反演模型比较[J]. 农业机械学报, 2017, 48 (3):

- 164—172.]
- [36] Xu E Q. Spatial variation in drivers of Karst rocky desertification based on geographically weighted regression model[J]. *Resources Science*, 2017, 39 (10): 1975—1988. [许尔琪. 基于地理加权回归的石漠化影响因子分布研究[J]. *资源科学*, 2017, 39 (10): 1975—1988.]
- [37] Li K, Yang Y, Liu Y J, et al. Characteristics of spatiotemporal variation of soil organic matter in Hebei Province based on regression Kriging[J]. *Soil and Fertilizer Sciences in China*, 2020 (3): 1—7. [李可, 杨勇, 刘亚军, 等. 基于回归克里格的河北省土壤有机质时空变化特征研究[J]. *中国土壤与肥料*, 2020 (3): 1—7.]
- [38] 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.
- [39] Shahriari M, Delbari M, Afrasiab P, et al. Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: A case of southeastern Iran[J]. *Catena*, 2019, 182: 104149.
- [40] Song Y Q, Yang L A, Li B, et al. Spatial prediction of soil organic matter using a hybrid geostatistical model of an extreme learning machine and ordinary kriging[J]. *Sustainability*, 2017, 9 (5): 754.
- [41] Dai F Q, Zhou Q G, Lü Z Q, et al. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau[J]. *Ecological Indicators*, 2014, 45: 184—194.
- [42] 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.
- [43] Su X Y. Research on the influence of sampling design on the spatial prediction precision of soil organic matter content[D]. Nanjing: Nanjing Normal University, 2012. 苏晓燕. 采样设计对土壤有机质含量空间预测精度的影响研究[D]. 南京: 南京师范大学, 2012.
- [44] Gascuel-Oudoux C, Boivin P. Variability of variograms and spatial estimates due to soil sampling: A case study[J]. *Geoderma*, 1994, 62 (1/2/3): 165—182.
- [45] Zhang Z Q, Yu D S, Shi X Z, et al. Priority selection rating of sampling density and interpolation method for detecting the spatial variability of soil organic carbon in China[J]. *Environmental Earth Sciences*, 2015, 73 (5): 2287—2297.
- [46] Zhang Z Q, Yu D S, Shi X Z, et al. Effect of sampling classification patterns on SOC variability in the red soil region, China[J]. *Soil and Tillage Research*, 2010, 110 (1): 2—7.
- [47] Khaledian Y, Miller B A. Selecting appropriate machine learning methods for digital soil mapping[J]. *Applied Mathematical Modelling*, 2020, 81: 401—418.

(责任编辑: 檀满枝)