DOI: 10.13758/j.cnki.tr.2024.04.021
贾德伟, 周磊, 王宁, 等. 基于 Sentinel-2 数据的豫北平原耕作层土壤有机碳估算. 土壤, 2024, 56(4): 866–872.

基于 Sentinel-2 数据的豫北平原耕作层土壤有机碳估算①

贾德伟, 周 磊*, 王 宁, 刘玉昕

(河南省乡村产业发展服务中心, 郑州 450018)

摘 要: 为精准掌握平原区耕作层土壤有机碳(SOC)含量及其空间分布, 利用滑县实测 SOC 数据, 融合 Sentinel-2 反射率光谱及光谱指数、气象、土壤、地形和人类活动等因子数据, 定量评估了弹性网络(EN)、偏最小二乘(PLS)和随机森林(RF)模型在不同因子组合下的精度,以揭示影响 SOC 估算的最佳因子、关键变量及其空间变异规律。结果显示:在模型精确度和稳定性方面,RF>EN>PLS, 特别是在反射率光谱、光谱指数、气象和土壤等最佳因子组合下, RF 模型达到了最高精度, 决定系数(R^2)为 0.63, 性能与四分位距之比(RPIQ)为 2.75, 平均绝对误差百分比(MAPE)为 5.86。此外, 中心波长 2 190 nm 的第 12 波段(B_{12})、潜在蒸散发、土壤容重和归一化土壤指数(NDSI)累积重要性达到了 69.63%, 成为 RF 模型的关键变量, 在 SOC 估算中起到了决定性作用。RF 模型估算的研究区 SOC 范围在 6.15 ~ 11.08 g/kg, 平均值为 8.16 g/kg, 低于全国平均水平(22.28 g/kg); 从空间分布来看, SOC 含量较高的区域主要集中在东北部, 而含量较低的区域则主要位于西南部。

关键词: 土壤有机碳; 随机森林; Sentinel-2; 耕作层; 估算
中图分类号: S151.9 **文献标志码:** A

Estimation of Soil Organic Carbon in Plow Layer of Northern Henan Plain Based on Sentinel-2 Data

JIA Dewei, ZHOU Lei*, WANG Ning, LIU Yuxin
(*Service Center of Rural Industrial Development in Henan Province*, *Zhengzhou* 450018, *China*)

Abstract: To accurately understand the content and spatial distribution of soil organic carbon (SOC) in the plow layer of Plain, in this paper, actual SOC data were used from Huaxian County, integrated with Sentinel-2 reflectance spectra and spectral indices, and data of meteorological parameters, soil properties, terrain and human activity factors, and then the predictive accuracy of the Elastic Net (EN), Partial Least Squares (PLS), and Random Forest (RF) models were quantitatively evaluated and compared under various factor combinations, the optimal factors and key variables that influence SOC estimation, as well as the spatial variation patterns were revealed. The results showed that in terms of model accuracy and stability, RF was best, followed by EN and PLS. Specially, RF model achieved the highest precision with the best combination of factors including reflectance spectra, spectral indices, meteorological data, and soil properties, with the coefficient of determination (R^2) of 0.63, the ratio of performance to inter-quartile range (RPIQ) of 2.75, and the mean absolute percentage error(MAPE) of 5.86. Moreover, the 12th band with a central wavelength of 2 190 nm (B_{12}), potential evapotranspiration, soil bulk density, and the normalized difference soil index (NDSI) cumulatively accounted for 69.63% of the importance in RF Model, becoming the key variables in RF model and playing a decisive role in SOC estimation. The range of SOC in the study area was between 6.15 and 11.08 g/kg, with an average of 8.16 g/kg, lower than the national average of 22.28 g/kg. Spatially, the area with higher SOC content was mainly located in the northeastern part, while the area with lower content was primarily located in the southwestern part.

Key words: Soil organic carbon; Random forest; Sentinel-2; Plow layer; Estimation

①基金项目: 河南省科技攻关项目(2221021100999, 232102110282)资助。
* 通讯作者(zhoulanr@163.com)
作者简介: 贾德伟(1984—), 男, 河南周口人, 硕士, 高级农艺师, 研究方向为农业遥感。E-mail: jiadewei118@163.com

土壤有机碳(SOC)是衡量土壤肥力的关键指标，估算 SOC 对促进耕地可持续利用和实现作物高产稳产具有重要意义。传统 SOC 估算主要通过地球化学方法[1]，结果准确可靠，但存在投入高、周期长及破坏土壤环境等问题。有学者利用地面高光谱数据构建 SOC 反演模型，得出 SOC 含量与可见光和短波红外光谱指数呈显著相关[2-3]，实现了 SOC 快速估算，既降低了成本，又保证了土壤环境完整性，但上述研究无法反映区域尺度上 SOC 的空间变异特征。GIS 及地统计学的结合解决了区域 SOC 估算问题，使得平原、丘陵等不同地貌的 SOC 空间变异得以估算[4-5]，但该方法对样本数量要求高，无法全面揭示环境因子对 SOC 的影响，在表达 SOC 空间格局细节方面仍有待提高。卫星遥感技术以其高时空分辨率、易于获取等优势，为 SOC 快速估算及变化监测开辟了新的途径。基于 MODIS 数据的研究得出，中国东北部、西南部以及东南部 SOC 含量高，而西北部 SOC 含量低[6]，但其 250 m 的空间分辨率限制了其在中小尺度 SOC 估算中的应用。为了提高估算的空间精度，研究者开始将 Landsat(30 m)、Sentinel-2(10 m)系列卫星遥感数据应用于水稻土区和黑土区 SOC 估算[1,7]。同时，偏最小二乘算法、弹性网络及随机森林[6−11]等机器学习方法的应用，为 SOC 估算提供了坚实的理论基础。

在宏观尺度上，SOC 受气象、土壤、地形以及植被等自然因素的复杂影响，导致其在水平和垂直方向上呈现显著分异[6,10]。然而，在小尺度区域，如以潮土为主的豫北平原耕作区，其地貌特征相对均一，气象和土壤的空间异质性较小，各环境因素对 SOC 的影响机制将发生变化，探究不同环境因子组合下 SOC 估算模型精度变化情况，以及各环境因子对模型的具体贡献度，均是当前亟需深入研究的问题。

河南是全国粮食大省，其 SOC 水平关系到农作物的产量和质量，对保障国家粮食安全具有显著影响。基于此，本文将有“豫北粮仓”之称的滑县作为研究区，基于实测耕作层 SOC 和 Sentinel-2 高分辨率卫星遥感反射率和光谱指数数据，辅以气象、土壤、地形及人类活动等因子，综合分析各因子中不同变量与 SOC 的相关性，定量评估不同机器学习方法在不同因子组合下的表现，旨在构建高精度的 SOC 估算模型，探究其最佳因子和关键变量，揭示 SOC 空间变异规律。

1　材料与方法

1.1　研究区概况

研究区为河南省东北部的滑县(114°25′ E ~ 114°57′ E，35°13′ N ~ 35°39′ N)(图 1)，总面积 1 814 km^2，耕地面积 1 340 km^2。该县海拔高度 29 ~ 71 m，地势西高东低，以平原为主，年均气温 13.7°C，平均降水量 634.3 mm，日照时数 2 365.5 h，无霜期 201 d，温带季风气候。耕地以潮土为主[12]，春季主种小麦，秋季主种玉米。

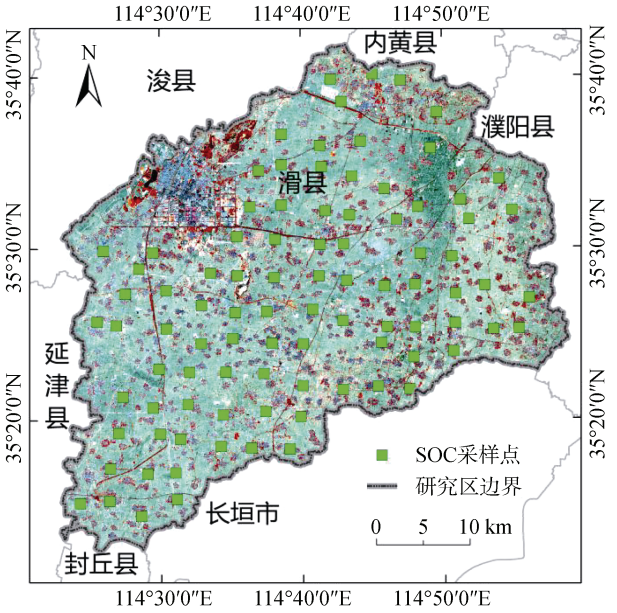

图 1　研究区地理位置及 SOC 样方点分布
Fig. 1　Locations of study area and SOC sampling sites

1.2　土样采集及测定

2023 年 2 月开展土样野外采集工作。在研究区内，构建 3 km × 3 km 网格，基于网格内耕地面积占比，结合代表性和均匀性原则，筛选出 92 个土样采集样方，在每个样方内，随机选择 30 m × 30 m 的样地(图 1)，按照“S”形路线，每隔 10 m 取 1 钻土，共取 10 钻土混合成 1 个土样，取样深度 20 cm。所有土样经自然风干、研磨、过筛处理，利用重铬酸钾氧化滴定法测定 SOC 含量。数据采集过程中，部分田块刚进行翻耕、浇水或施肥等农事活动，导致数据存在离群情况，因此，本文以四分位数和四分位距为基础的箱型图法剔除异常点，最后保留 85 个样本数据用于后续模型构建。

1.3　其他数据来源及预处理

卫星遥感数据选择裸土期(2022 年 10 月 19 日) Sentinel-2 的 L2A 产品，来自哥白尼数据空间生态系统，该产品包含 12 个波段(B_1 ~ B_{12})，中心波长 443 ~

2 190 nm，均已经过大气校正和几何精校正。基于该数据提供的 BOA_QUANTIFICATION_VALUE 和 BOA_ADD_OFFSET 参数，利用 ENVI 5.6 进行增益值计算，获取地表反射率数据。同时，计算生成卫星遥感数据衍生的光谱指数数据，一是使用增强植被信号较好的归一化差值指数(NDVI)和土壤调节植被指数(SAVI)[13-14]；二是使用增强裸土信息较好的裸土指数(BSI)、增强型裸土指数(EBSI)、归一化差值裸地与建筑用地指数(NDBBI)和归一化土壤指数(NDSI)[15-16,9]。

气象数据包括 2022 年平均潜在蒸散发、气温和降水量，来自国家青藏高原科学数据中心，空间分辨率为 1 km。土壤数据包括土壤容重、黏粒、砂粒和粉粒，来自时空三极环境大数据平台，空间分辨率为 250 m。DEM 数据来自 NASA 阿拉斯加卫星设备处的 ALOS ，空间分辨率为 12.5 m，并通过 SAGA 9.3 计算坡度、坡向、地形位置指数(TPI)、汇流动力指数(SPI)和地形粗糙指数(TRI)等变量。人类活动数据是在道路、居民点等第三次全国国土调查数据基础上，生成代表人类活动强弱的最近欧式距离栅格数据，空间分辨率 10 m。

为最大限度发挥 Sentinel-2 的高空间分辨率优势，提升 SOC 估算精度，采用 ArcGIS 10.2 软件的最邻近插值法，将气象、土壤和地形数据重采样至 10 m×10 m 栅格，空间参考系统也统一转换至 UTM50，确保所有数据空间尺度及位置保持一致。

1.4 研究方法

1.4.1 偏最小二乘 偏最小二乘(Partial least square，PLS)是 Herman Word 提出的统计学方法[17]，其对自变量和因变量重新投影，在新坐标系中，将对原自变量贡献率最强的前几位新变量作为主成分，减少原自变量间多重共线性冗余信息，最后通过多元线性回归进行分析。PLS 通过 Python 语言 sklearn 库的 PLSRegression 函数实现。

1.4.2 弹性网络 弹性网络(Elastic net，EN)是 Zou 和 Hastie [18]提出的算法，用于解决多重共线性问题。它通过一个混合惩罚项(α)来“弹性”结合 Lasso 和 Ridge 回归的优点，当 α 为 0 时，等价于 Ridge 回归；当 α 为 1 时，等价于 Lasso 回归；当 α 在(0,1)时，EN 兼具两者特性。EN 通过 Python 语言 sklearn 库的 ElasticNet 函数实现。

1.4.3 随机森林 随机森林(Random forest，RF)由 Breiman[19]提出，通过 bootstrap 抽样法，多次随机有放回地将原始数据分为袋内和袋外随机样本，每次训练袋内样本子集，生成大量相互独立的决策树组成随机森林，并用相应袋外数据的误差评估最佳回归树数量和最优分裂节点数，所有决策树预测平均值作为回归的最终值。RF 通过 Python 语言 sklearn 库的 RandomForestRegressor 函数实现。

1.5 评价指标

采用决定系数(R^2)、性能与四分位距之比(RPIQ)、平均绝对误差百分比(MAPE)对模型进行评估。R^2 越接近 1，预测值与实际值拟合度越好，精度越高。MAPE 是预测值与实际值的误差百分比，值越小，预测结果越准确。RPIQ 兼顾预测值误差和实际值变化，值越大，模型性能越好[20]。

2 结果与分析

2.1 土壤有机碳描述性统计

根据 7∶3 原则，研究区 85 个样本数据中 59 个用于训练集，26 个用于验证集。总样本 SOC 含量介于 5.86 ~ 11.56 g/kg，平均为 7.98 g/kg，变异系数为 11.70%，表明总样本呈现弱中等变异性(表 1)。训练集和验证集的统计特性与总样本相似。

表 1 SOC 含量描述性统计

Table 1 Descriptive statistics of SOC content

样本	最小值(g/kg)	最大值(g/kg)	均值(g/kg)	标准差(g/kg)	方差	变异系数(%)
总样本	5.86	11.56	7.98	0.93	0.87	11.70
训练集	5.86	11.56	7.99	0.97	0.94	12.10
验证集	6.01	9.11	7.95	0.87	0.76	10.96

2.2 土壤有机碳与各类因子的相关性

利用 SPSS 27 软件计算 SOC 含量与反射率光谱(R)、光谱指数(I)、气象(M)、土壤(S)、DEM(D)和人类活动(H)因子中不同变量的 Pearson 相关性(表 2)，结果显示，SOC 与各类因子中不同变量的相关性大小存在显著差异。SOC 与 R 因子的相关性随波长变化呈现“W”形模式，其中与中心波长为 2 190 nm(B_{12})和 440 nm(B_1) 的反射率光谱相关性极显著。SOC 与 I 因子中的裸土指数呈负相关，与植被指数呈正相关，且与多数裸土指数的相关性强度大于其与植被指数的相关性，如 SOC 与 NDSI、BSI 和 NDBBI 的相关性大于 SOC 与 NDVI 和 SAVI 的相关性。SOC

与 M 因子中的潜在蒸散发和气温呈极显著负相关，而与降水量呈显著正相关。SOC 与 S 因子的容重和粉粒呈极显著相关，而与黏粒则相关性不显著。SOC 与 D 因子均呈负相关，但仅与高程的相关性极显著。SOC 与 H 因子呈正相关，且仅与居民点距离的相关性显著。

表 2　SOC 与各类因子中变量的相关性

Table 2　Correlations between SOC contents and different variables in various factors

因子	变量	相关性	因子	变量	相关性
R	B_1	–0.291**		NDVI	0.306**
	B_2	–0.170		SAVI	0.257**
	B_3	–0.160	M	潜在蒸散发	–0.472**
	B_4	–0.070		气温	–0.371**
	B_5	–0.020		降水量	0.218*
	B_6	0.040	S	容重	–0.416**
	B_7	0.050		粉粒	0.334**
	B_8	0.070		砂粒	–0.308**
	B_9	0.100		黏粒	–0.190
	B_{10}	–0.010	D	高程	–0.387**
	B_{11}	–0.160		坡度	–0.120
	B_{12}	–0.515**		坡向	–0.070
I	NDSI	–0.502**		TPI	–0.160
	BSI	–0.432**		TRI	–0.130
	NDBBI	–0.374**		SPI	–0.040
	EBSI	–0.150	H	居民点距离	0.216*
	BI	–0.110		城镇道路距离	0.040

鉴于变量间存在多重共线性问题，并非所有变量都会对 SOC 估算产生显著影响[21]，本文采取分步筛选变量策略：首先，从每类因子中筛选出与 SOC 相关性最强的单一变量，即 B_{12}、NDSI、潜在蒸散发、容重、高程和居民点距离；其次，在 R、I、M 和 S 因子中，若有两个以上变量与 SOC 呈(极)显著相关，则在上步基础上再添加一个相关性最强的变量代表该因子，即 B_1、BSI、气温和粉粒。最终，有 10 个变量进入后续模型构建，在适当控制变量多重共线性问题的同时，又确保了模型估算能力。

2.3　土壤有机碳估算模型及精度检验

2.3.1　环境因子组合　　根据 R、I、M、S、D 和 H 因子与 SOC 相关性强度，依次添加不同因子，采用 10 折交叉验证的格网搜索方法构建 EN、PLS 和 RF 模型(表 3)。EN 模型在 R、I、M、S 和 D 因子依次添加后，R^2 从 0.27 持续增至 0.47，RPIQ 也达到最高值 2.29；随着 H 因子加入后，R^2 和 RPIQ 均出现降低，表明 H 因子降低了 EN 模型精度；然而，EN 模型的 MAPE 却在 R+I+M 因子组合下达到最低值 6.82。PLS 模型在 R、I、M、S 和 D 因子依次添加后，R^2 从 0.32 持续增至 0.54；尽管加入 H 因子后，R^2 略有下降，但 RPIQ 和 MAPE 却有所改善，表明 H 因子在一定程度上可提高 PLS 模型精度；然而，PLS 模型的 RPIQ 和 MAPE 却在 R 和 I 因子组合下达到最佳值 1.46 和 10.37。RF 模型在 R、I、M 和 S 因子依次添加后，R^2 从 0.39 持续增至 0.63，RPIQ 和 MAPE 也分别达到最佳的 2.75 和 5.86，表明此因子组合下 RF 模型精度最佳；然而，当加入 D 和 H 因子后，3 个指标均显示 RF 模型性能下降。

表 3　不同因子组合下模型精度对比

Table 3　Accuracy comparisons of various models under various factor combinations

模型	因子组合	R^2	RPIQ	MAPE
EN	R	0.27	1.95	7.65
	R+I	0.29	1.99	7.48
	R+I+M	0.46	2.29	6.82
	R+I+M+S	0.46	2.29	6.82
	R+I+M+S+D	0.47	2.29	6.89
	R+I+M+S+D+H	0.43	2.22	7.05
PLS	R	0.32	1.44	10.47
	R+I	0.38	1.46	10.37
	R+I+M	0.50	1.37	11.03
	R+I+M+S	0.54	1.30	11.43
	R+I+M+S+D	0.54	1.31	11.40
	R+I+M+S+D+H	0.53	1.31	11.40
RF	R	0.39	2.15	7.46
	R+I	0.49	2.33	6.79
	R+I+M	0.60	2.65	6.19
	R+I+M+S	0.63	2.75	5.86
	R+I+M+S+D	0.60	2.64	6.05
	R+I+M+S+D+H	0.59	2.61	6.15

3 个评价指标中，有 2 个以上显示为最优，则可判定该因子组合为最佳因子组合。综上，EN、PLS 和 RF 模型最佳因子组合分别为 R+I+M+S+D、R+I 和 R+I+M+S。

2.3.2　精度对比　　RF 模型的最大 R^2 为 0.63，较 EN 和 PLS 模型分别提高 34.04% 和 15.81%，说明 RF 模型能更好拟合 SOC 变化；RPIQ 方面，RF 模型得分最高(2.75)，优于 EN 模型(2.29)和 PLS 模型(1.46)，依照 RPIQ 评估标准[20]可知，RF 模型估算最为精准。此外，RF 模型的 MAPE 最低(5.86)，其估算准确率达 94.14%，较 EN 和 PLS 模型分别增加 1.03 和 4.51 个百分点。

综上，3 个评价指标均显示 RF 模型优于 EN 和 PLS 模型，其中 2 个评价指标显示 EN 模型优于 PLS

模型。因此，模型整体精度为：RF>EN>PLS。

2.3.3 变量重要性 计算 RF 模型各变量的重要性(图 2)，结果显示，R 因子对 SOC 估算贡献最大(39.97%)，其次为 M 因子(23.98%)、S 因子(20.14%)和 I 因子(15.91%)。其中，R 因子中 B_{12}(34.17%)> B_1(5.8%)，M 因子中潜在蒸散发(14.64%)>气温(9.34%)，S 因子中容重(11.43%)>粉粒(8.71%)，I 因子中 NDSI(9.39%)>BSI(6.52%)，表明虽然上述变量均与 SOC 显著相关，但它们对估算 SOC 的贡献程度不同。重要性排名前 4 位的变量依次为 B_{12}、潜在蒸散发、容重和 NDSI，累积贡献率达 69.63%，它们作为 4 类因子的代表，是 SOC 估算的关键变量。

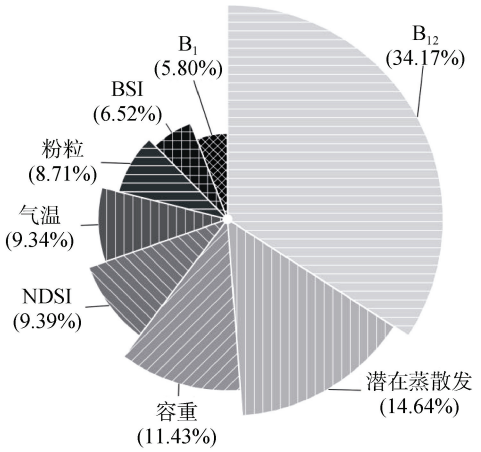

图 2 RF 模型的变量重要性

Fig. 2 Importance of variables in RF model

2.4 土壤有机碳空间分布格局

以耕地矢量数据为边界，分别绘制 EN、PLS 和 RF 模型在其最佳因子组合下的 SOC 空间分布预测图(图 3)。结果显示，3 个模型估算的 SOC 空间分布整体相似，呈东北高、西南低趋势。SOC 值域范围差距明显(表 4)，EN 模型(3.32 ~ 13.73 g/kg)和 PLS 模型(5.14 ~ 14.97 g/kg)的估算结果相较于总样本(5.86 ~ 11.56 g/kg)，波动性大，稳定性差，而 RF 模型估算结果(6.15 ~ 11.08 g/kg)更接近于总样本，说明 RF 模型既能很好描述研究区 SOC 空间差异，又能确保结果稳定。RF 模型估算的 SOC 均值为 8.16 g/kg，低于全国平均值(22.28 g/kg)[6]。其中，中值区占比最高，占耕地总面积 78.82% 以上，主要分布在研究区中部；低值区和高值区占比为 6.37% 和 14.81%，分别分布在研究区西南部和东北部。

3 讨论

3.1 最佳模型

SOC 分布受多重因素影响，往往表现为复杂的非线性关系[22]。相较于 EN 和 PLS 模型，RF 模型通过构建不同决策树，并考虑变量间多种分割和相互作用，更能精确模拟非线性关系[23]。本研究表明，在样本点拟合、估算结果分布范围等方面，RF 模型优于 EN 和 PLS 模型，与先前研究[24,11]结论一致。

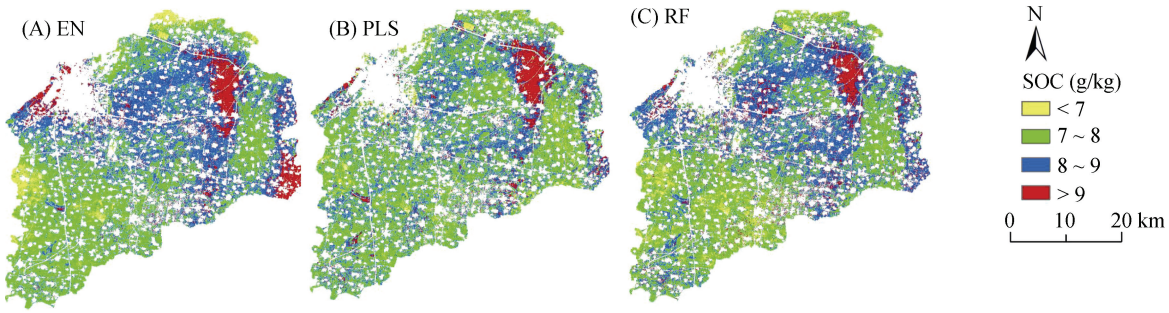

图 3 不同模型估算的 SOC 空间分布

Fig. 3 Spatial distribution of SOC contents estimated by different models

表 4 不同模型估算的 SOC 基本统计特征

Table 4 Basic statistical characteristics of SOC contents estimated by different models

模型	最小值 (g/kg)	最大值 (g/kg)	均值 (g/kg)	标准差 (g/kg)	方差	变异系数(%)
EN	3.32	13.73	8.15	0.82	0.67	10.05
PLS	5.14	14.97	8.05	0.86	0.74	10.68
RF	6.15	11.08	8.16	0.82	0.68	10.07

此外，RF 模型估算结果受 SOC 与各类因子相关性强弱的影响[10]。如，当加入与 SOC 相关性强的 R 和 M 因子时，模型精度显著提升，但加入与 SOC 相关性较弱的高程及居民点距离时，模型精度不升反降。因此，构建 RF 模型时，应慎重选择与 SOC 相关性强的变量，以此提高其估算能力。

3.2 最佳因子及关键变量

在 R、I、M 和 S 因子组合下，RF 模型精度最高，使该因子组合成为研究区 SOC 估算的最佳因子。模型重要性揭示，B_{12}、潜在蒸散发、土壤容重和 NDSI

对 RF 模型的累积贡献度达 69.63%，是模型的关键变量。原因可能是：①B_{12} 中心波长 2 190 nm，该波段能敏感地捕捉裸土信息，表现与 SOC 空间变异直接对应的光谱值，与 Suleymanov 等[23]结论一致；②潜在蒸散发反映干旱程度和水资源状况，值升高说明干旱趋势加剧，孔隙度变大，颗粒更加松散，风蚀风险变大，SOC 积累少、流失多，与 Li 等[25]结论一致；③土壤容重反映土壤的通气透水性和孔隙度[22]，该值增加说明土壤通气性变差，微生物因氧气供给受限而活动性减弱，从而降低了 SOC 含量，与 Wang 等[26]结论一致；④NDSI 能增强裸土反射特性，值越大，土壤暴露程度越严重，水分越少，植被生长和土壤微生物活性减弱，SOC 含量下降，与闫蒙等[27]结论一致。

同时，气温、粉粒、BSI 和 B_1 变量对 RF 模型也有一定的贡献度，累积达 30.37%，部分原因是：①气温升高会促使 K-策略为主的微生物群落提高分解活动和土壤呼吸，以维持其代谢所需能量，降低有机合成碳分配，促使 SOC 含量下降[28]；②粉粒因颗粒细，表面积大，利于提高土壤的保水性和通气性，有更多正电荷与土壤中带负电荷的腐殖质结合[27]，促进有机质吸附，增加 SOC 积累，这与多个研究结论一致[22,27]。

另外，地形不能成为本研究区 SOC 估算模型的重要因子，因高程的加入仅略微提高了 EN 模型的 R^2 和 RPIQ，及 PLS 模型的 R^2，而 RF 模型精度却未得到提升。该结论与部分研究结果不一致[6,10-11]，可能是因为研究区地形起伏不大(29 ~ 71m)，该因子无法精准描述 SOC 的空间异质性。同时，居民点距离也不能提高 RF 模型和 EN 模型精度，仅对 PLS 模型的 RPIQ 和 MAPE 指标有相应的微弱提升，可能与人类活动对土壤的随机性干扰有关。

3.3　误差分析

SOC 估算模型的 R^2 大多小于 0.4[7]。本研究区为小尺度的平原区，各类因子空间异质性较低，模型 R^2 达 0.63，高于现有部分研究结果[7,21]，处于中上等水平，但仍低于个别研究[8,29]，可能误差源为：①本文对气象、土壤及地形因子数据进行了空间尺度的向上重采样，该操作会因尺度变换而引入误差；②SOC 与各类因子关系复杂，本文参与模型构建的仅是各因子中与 SOC 呈(极)显著相关的变量，并不能全部解释 SOC 的空间变异。

本文提出了一种利用 Sentinel-2 卫星影像、气象和土壤因子快速估算 SOC 的技术方法，揭示了平原区耕作层 SOC 估算中的最佳因子及关键变量，不仅为提高 SOC 估算精度提供了理论依据，而且为因地制宜增强农业综合生产力提供了方法支撑。未来应进一步探究不同尺度因子及多时相卫星遥感数据对 SOC 估算精度的影响。

4　结论

以河南省滑县为研究区，基于 SOC 样本数据，融合 R、I、M、S、D 及 H 类因子数据，探究估算 SOC 的最佳模型、最佳因子、关键变量，结果显示：①各类因子中均有变量与 SOC 呈(极)显著相关；②不同因子组合下，RF 模型在精度和空间异质性描述方面均优于 EN 和 PLS 模型；③RF 模型在 R、I、M 和 S 最佳因子的组合下，表现出了最高估算精度，R^2 为 0.63，RPIQ 为 2.75，MAPE 为 5.86，其中，B_{12}、潜在蒸散发、容重和 NDSI 的累积贡献率达 69.63%，成为 SOC 估算的关键变量；④RF 模型估算结果显示，研究区 SOC 含量呈东北高、西南低的空间分布趋势，变动范围为 6.15 ~ 11.08 g/kg，表现出弱中等变异性，平均值为 8.16 g/kg，含量相对较低，低于全国平均值(22.28 g/kg)。

参考文献：

[1] 杨佳佳, 林楠, 于秀秀, 等. 东北典型黑土区有机碳遥感定量反演研究[J]. 地质与资源, 2020, 29(4): 357–362.

[2] Summers D, Lewis M, Ostendorf B, et al. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties[J]. Ecological Indicators, 2011, 11(1): 123–131.

[3] 罗德芳, 彭杰, 冯春晖, 等. 可见光-近红外、中红外光谱的土壤有机质组分反演[J]. 光谱学与光谱分析, 2021, 41(10): 3069–3076.

[4] 赵昕, 张晓光, 宋祥云, 等. 胶莱平原县域表土有机碳空间变异特征研究及自相关分析[J]. 干旱区资源与环境, 2023, 37(4): 127–136.

[5] 朱阳春, 张振华, 赵学勇, 等. 河套灌区土壤有机碳和总碳的空间异质性及相关性分析[J]. 江苏农业学报, 2017, 33(6): 1294–1300.

[6] 罗梅, 郭龙, 张海涛, 等. 基于环境变量的中国土壤有机碳空间分布特征[J]. 土壤学报, 2020, 57(1): 48–59.

[7] 吴启航, 姚园, 李一凡, 等. 福建省漳州市水稻物候特征对稻田土壤有机碳制图的影响[J]. 土壤学报, 2024, 61(2): 385–397.

[8] 赵启东, 葛翔宇, 丁建丽, 等. 结合分数阶微分技术与机器学习算法的土壤有机碳含量光谱估测[J]. 激光与光电子学进展, 2020, 57(15): 253–261.

[9] 牛芳鹏, 李新国, 麦麦提吐尔逊·艾则孜, 等. 基于光谱指数的博斯腾湖西岸湖滨绿洲土壤有机碳含量估算模型[J]. 江苏农业学报, 2022, 38(2): 414–421.

[10] 袁玉琦, 陈瀚阅, 张黎明, 等. 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.

[11] 卢宏亮, 赵明松, 刘斌寅, 等. 基于随机森林模型的安徽省土壤属性空间分布预测[J]. 土壤, 2019, 51(3): 602–608.
[12] 李笑莹, 张学雷, 任圆圆. 河南省土壤及地形与耕地多样性格局的特征[J]. 土壤, 2019, 51(4): 775–785.
[13] 贾德伟, 周磊, 黄灿辉, 等. 基于 MODIS 数据的冬小麦雹灾空间分布信息提取研究——以河南省平顶山市为例[J]. 地域研究与开发, 2018, 37(6): 134–138.
[14] 彭燕, 何国金, 张兆明, 等. 中国区域 Landsat 遥感指数产品[J]. 中国科学数据, 2020, 5(4): 83–90.
[15] 李虎, 钟韵, 冯雅婷, 等. 无人机遥感的多植被指数土壤水分反演模型[J]. 光谱学与光谱分析, 2024, 44(1): 207–214.
[16] 吴志杰, 赵书河. 基于 TM 图像的“增强的指数型建筑用地指数” 研究[J]. 国土资源遥感, 2012, 24(2): 50–55.
[17] 刘潜, 王梦迪, 郭龙, 等. 基于机载高光谱影像的农田尺度土壤有机碳密度制图[J]. 遥感学报, 2024, 28(1): 293–305.
[18] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, 67(2): 301–320.
[19] Breiman L. Random forests[J]. Machine Learning, 2001, 45: 5–32.
[20] 王瑾杰, 丁建丽, 葛翔宇, 等. 分数阶微分技术在机载高光谱数据估算土壤含水量中的应用[J]. 光谱学与光谱分析, 2022, 42(11): 3559–3567.
[21] 周琪清, 赵小敏, 郭熙, 等. 基于物候与极端气候信息的耕地土壤有机碳空间分布预测研究[J]. 土壤学报, 2024, 61(3): 648–661.
[22] 常琳溪, 梁新然, 王磊, 等. 中国稻田土壤有机碳汇特征与影响因素的研究进展[J]. 土壤, 2023, 55(3): 487–493.
[23] Suleymanov A, Abakumov E, Nizamutdinov T, et al. Soil organic carbon stock retrieval from Sentinel-2A using a hybrid approach[J]. Environmental Monitoring and Assessment, 2023, 196(1): 23.
[24] 杨珺婷, 李晓松. 应用哨兵 2 号卫星遥感影像数据和机器学习算法对锡林郭勒草原土壤表层有机碳及全氮的估算[J]. 东北林业大学学报, 2022, 50(1): 64–71.
[25] Li C J, Fu B J, Wang S, et al. Drivers and impacts of changes in China’s drylands[J]. Nature Reviews Earth & Environment, 2021, 2: 858–873.
[26] Wang Y Q, Shao M A, Liu Z P, et al. Prediction of bulk density of soils in the Loess Plateau Region of China[J]. Surveys in Geophysics, 2014, 35(2): 395–413.
[27] 闫蒙, 王旭洋, 周立业, 等. 科尔沁沙地沙漠化过程中土壤有机碳含量变化特征及影响因素[J]. 中国沙漠, 2022, 42(5): 221–231.
[28] Li H, Yang S, Semenov M V, et al. Temperature sensitivity of SOM decomposition is linked with a K-selected microbial community[J]. Global Change Biology, 2021, 27(12): 2763–2779.
[29] 李宏达, 李德成, 曾荣. 基于光谱相似性匹配的土壤有机碳估算[J]. 土壤学报, 2021, 58(5): 1224–1233.