

基于双时相卫星遥感光谱指数估算土壤有机质含量^①

王欣怡^{1,2}, 王昌昆^{1,2*}, 马海艺^{1,2}, 刘杰¹, 袁自然^{1,2}, 姚成硕^{1,2}, 潘贤章^{1,2}

(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 中国科学院大学, 北京 100049)

摘要: 以黄淮海平原典型县——封丘县为研究区, 探讨了在一年两熟、裸土时间窗口较短的区域中, 基于两景影像波段组合构建的双时相光谱指数在有机质含量预测中的表现。研究共计采集 117 个代表性土样, 以分析筛选出的裸土期(10 月)内双时相(获取时间: 2014 年 10 月 6 日和 2017 年 10 月 30 日)高质量 Landsat 8 卫星影像作为数据源, 构建了 4 种类型的光谱指数: 比值光谱指数、差值光谱指数、归一化光谱指数以及优化光谱指数, 并结合最小绝对收缩和选择算子变量筛选方法和支持向量机算法建立了有机质预测模型。留一交叉验证结果表明, 与直接使用影像波段反射率或者基于单景影像构建的光谱指数(单时相光谱指数)相比, 利用双时相光谱指数可以更好地利用时相信息优势, 其有机质预测精度更高($R^2=0.53$, RMSE=2.01 g/kg)。而且, 基于双时相光谱指数所构建的预测模型得到的有机质空间分布格局与真实值较为吻合。可见, 本文提出的在黄淮海平原典型县域利用双时相光谱指数预测土壤有机质的方法, 可以促进具有短裸土期特点区域的高分辨率土壤属性遥感预测与制图研究。

关键词: 土壤有机质; 土壤遥感; 双时相光谱指数; 黄淮海平原

中图分类号: S127 文献标志码: A

Estimation of Soil Organic Matter Content Based on Dual-temporal Satellite Remote-sensing Spectral Index

WANG Xinyi^{1,2}, WANG Changkun^{1,2*}, MA Haiyi^{1,2}, LIU Jie¹, YUAN Ziran^{1,2}, YAO Chengshuo^{1,2}, PAN Xianzhang^{1,2}

(1 State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The performance of the dual-temporal spectral index (the band combination of two images) in predicting soil organic matter (SOM) was investigated over a double-cropping agricultural region (Fengqiu County) in the Huang-Huai-Hai Plain, where the bare soil period is often short for remote sensing of soils. In the study, a total of 117 soil samples were collected and dual-temporal Landsat 8 satellite images during the bare soil period (Acquisition date: October 6, 2014 and October 30, 2017) were selected for establishing four types of spectral indices: ratio spectral index, difference spectral index, normalized spectral index and optimized spectral index. Then, these indices were used as the input in SVM (Support Vector Machine) models of SOM after being selected by the variable selection method of LASSO (Least Absolute Shrinkage and Selection Operator). The results of leave-one-out cross-validation showed that, compared with image bands or spectral indices built by single images (single-temporal spectral index), the dual-temporal spectral index could make better use of temporal information of images and its prediction accuracy was higher for SOM ($R^2=0.53$, RMSE=2.01g/kg). Moreover, the spatial distribution pattern of SOM predicted by the dual-temporal spectral index was consistent with the real condition. Thus, the proposed method of using the dual-temporal spectral index for SOM prediction in the study could promote prediction and mapping of soil properties in areas with short bare soil periods.

Key words: Soil organic matter; Remote sensing of soils; Dual-temporal spectral index; Huang-Huai-Hai Plain

土壤有机质(Soil Organic Matter, SOM)是土壤的重要组成部分, 在土壤多种功能中均扮演着核心角

色, 是土壤质量评估、土壤功能评价等诸多应用中的关键参数, 同时也是土壤资源利用与管理的关键

①基金项目: 国家重点研发计划项目(2021YFD1500102)资助。

* 通讯作者(ckwang@issas.ac.cn)

作者简介: 王欣怡(1998—), 女, 江苏泰州人, 硕士研究生, 主要从事土壤遥感研究。E-mail: wangxinyi@issas.ac.cn

目标^[1], 因此有效获取 SOM 的空间分布信息至关重要。近年来, 遥感技术的快速发展和有效应用为区域 SOM 预测提供了新的途径和有利条件。SOM 是土壤的重要染色物质^[2], 其在可见光波段乃至短波红外波段都能够显著吸收入射光, SOM 含量与土壤光谱反射率之间通常呈显著负相关, 因此利用土壤光谱信息能够有效反映 SOM 含量的变化^[3]。

通常利用室内测量的土壤光谱能够高精度预测 SOM 含量。但利用卫星影像反演预测土壤属性时, 土壤自身条件(如土壤湿度和粗糙度)和土壤表面条件(如作物和作物残茬覆盖)等会显著影响土壤反射率, 进而严重降低土壤属性的预测精度^[4-5]。有研究表明, 利用土壤光谱指数可以一定程度上抑制土壤水分、粗糙度和大气等因素的影响^[6]。与直接利用卫星影像波段反射率相比, 基于线性或非线性波段组合形成的光谱指数所构建的土壤属性预测模型精度通常较好。例如, 杨邵文^[7]在建立北京市密云区耕地 SOM 预测模型时, 选用了 Landsat 8 卫星数据计算的差值光谱指数和比值光谱指数。赵瑞等^[8]基于 GF-5 号卫星高光谱影像, 通过构建多种光谱指数(包括土壤沙化指数、土壤退化指数、归一化亮度指数和土壤盐分指数)反演 SOM, 实现了对内蒙古与黑龙江交界区域的 SOM 预测。上述 SOM 遥感反演研究中, 涉及多种类型的光谱指数, 表明光谱指数方法具有提升区域 SOM 预测精度的潜力, 但这些研究构建的光谱指数没有考虑时相信息, 均是利用单一影像不同波段间组合构建的光谱指数。

为进一步提高 SOM 反演精度, 许多研究者引入了时相信息, 通过将两景不同时期遥感影像波段进行组合, 构建双时相光谱指数^[9-10]。其中, 刘焕军等^[10]认为, 不同 SOM 含量的土壤在相同外部条件(诸如土壤水分)影响下其反射光谱敏感程度不同, 通过选择不同时期影像构建的相同波段组合的双时相光谱指数, 具有提高有机质遥感反演的潜力; 其利用同一年内多景裸土期 MODIS 影像, 构建了包括差值指数等的 3 种双时相光谱指数进行 SOM 反演, 获得了较高的 SOM 预测精度。Dou 等^[6]基于相似的理论假设, 认为选择更长时期的遥感影像组合构建双时相光谱指数同样具有提高 SOM 预测的潜力; 研究中, 其选择与土壤采集时间相距超过 10 年的多年 MODIS 影像, 构建了包括差值指数、比值指数等 4 种双时相光谱指数反演东北黑土区 SOM, 取得了较为理想的 SOM 预测结果。上述研究利用双时相光谱指数方法, 针对东北黑土区 SOM 反演均取得了较好的预测结

果。但与东北黑土区(一年一熟)相比, 我国黄淮海地区(一年两熟)裸土期时间窗口更短, 同时还需要考虑云等因素对影像质量的影响, 因此在黄淮海地区进行土壤 SOM 反演可能需要更长时期的遥感影像。另外, 在遥感影像获取日期跨度较大的情形下(比如 5 年间隔两景不同影像), 影响不同影像间土壤光谱差异的因素可能更为复杂, 除采用相同波段外, 还可以考虑采用不同影像不同波段构建双时相光谱指数。因此, 本研究拟针对黄淮海典型县域, 进一步探索双时相光谱指数在农田土壤 SOM 遥感反演中的潜力。

当运用双时相光谱指数建立 SOM 预测模型时, 可能会出现光谱指数和土壤样品数量接近的情形。例如, 在刘焕军等^[10]的研究中, 共计使用了 61 个土壤样品, 双时相光谱指数高达 49 个; 在 Dou 等^[6]的研究中, 共有 56 个双时相光谱指数, 土壤样品为 147 个, 其中仅一半样品用于建模。这些研究中的模型均属于小样本高维模型, 在相关研究中其通过相关分析和逐步多元回归法来筛选变量, 降低变量间共线性。此外, 有研究表明, 筛选有效的建模变量和选择更为先进的机器学习方法可以提高 SOM 的预测精度^[11-12]。Wang 等^[13]提出可以将最小绝对收缩和选择算子(Least Absolute Shrinkage and Selection Operator, LASSO)算法用于变量选择, 然后结合支持向量机(Support Vector Machine, SVM)算法建立目标对象的预测模型。该方法已被成功用于 SOM 含量估算、森林生物量估算和数字土壤制图等多个领域^[11,14-15], 并取得较为理想的结果。

因此, 本研究以具有短裸土期特点的黄淮海平原典型县域封丘县为研究区, 利用 Landsat 8 影像, 通过构建多种类型的双时相光谱指数, 并结合 LASSO 筛选变量方法与 SVM 算法建立 SOM 预测模型, 探讨了基于两景影像波段组合构建的双时相光谱指数在 SOM 预测中的表现, 以期提高 SOM 预测精度, 并实现区域 SOM 空间分布预测。

1 材料与方法

1.1 研究区概况

研究区选择在河南省封丘县, 该县地处黄淮海平原中部, 其东、南两面毗邻黄河, 地理坐标为 $114^{\circ}14'E \sim 114^{\circ}43'E$, $34^{\circ}53'N \sim 35^{\circ}14'N$, 面积约 $1220 km^2$, 区域内主要土壤类型为潮土^[16]。该地区属于典型的季风气候, 年平均气温为 $15^{\circ}C$, 年平均降水量在 $500 \sim 600 mm$ 。封丘县是重要的粮食生产县, 区域内普遍采用小麦-玉米一年两熟的种植制度^[17], 夏玉米一

般在 6 月小麦收获后种植, 冬小麦在同年 10 月玉米收获后种植。

通常, 研究区内小麦收获后采用秸秆还田的保护性耕作, 此时土壤表面留有较多的小麦秸秆和残茬, 不利于土壤属性信息遥感监测, 而在玉米收获之后通常会进行土壤翻耕。因此, 研究区内 10—11 月内有一段土壤相对裸露、受作物和作物残茬影响相对较小的时期, 是土壤属性遥感反演的潜在时间窗口。

1.2 土壤样品采集与处理

土壤样品于 2014 年 6 月小麦收获后采集。采样前, 在室内按照 $3 \text{ km} \times 3 \text{ km}$ 的规则网格在研究区布设土样采集点位。野外土壤样品采集时, 根据交通通达性等因素进行一定的点位调整, 研究共计采集表层 ($0 \sim 20 \text{ cm}$) 土壤样品 117 件。采样同时使用亚米级高精度 GPS 设备记录采样点坐标, 其点位分布如图 1 所示。

采样时, 应尽可能地远离周边道路、防护林等非耕地区域, 以避免采样点周边道路和田间防护林所导致的混合像元问题。土壤样品采集采用五点采样法, 点位间距离设定为 $5 \sim 10 \text{ m}$, 5 个不同位置采集的土样混合后作为该采样点的代表性土样。采集的土样运至实验室, 经风干、研磨、过筛处理后, 采用重铬酸钾容量法测定 SOM 含量。

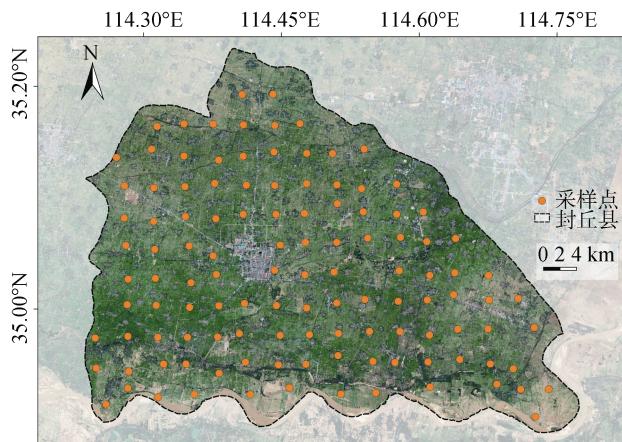


图 1 研究区采样点分布图

Fig. 1 Distribution map of sampling points in the study area

1.3 遥感影像数据获取

选择 Landsat 8 卫星平台 OLI 传感器影像作为影像数据源, 其空间分辨率为 30 m 。选择前 7 个光学范围内的波段用于 SOM 含量反演, 波段 1 至波段 7 的光谱范围分别为: $0.44 \sim 0.45$ 、 $0.45 \sim 0.51$ 、 $0.53 \sim 0.59$ 、 $0.64 \sim 0.67$ 、 $0.85 \sim 0.88$ 、 $1.57 \sim 1.65$ 、 $2.11 \sim 2.29 \mu\text{m}$ 。获取的影像为 Landsat 8 地表反射率产品,

已经经过辐射定标、大气处理和几何校正处理。

研究区的潜在裸土窗口期为 10—11 月, 该时间窗口内受土壤表面作物和作物残茬的影响相对较小, 能够较好地反映土壤光谱信息。此外, 由于 SOM 含量随时间变化相对较小, 并且为了获得更多有效的影像, 首选初选了采样时间近 5 年内(2013—2017 年间)10—11 月之间云量小于 20% 的影像, 经过筛选能够覆盖全部土壤采样点的 Landsat 8 影像共有 7 景(条带号: 123, 行编号: 036), 其时间和云量信息如表 1 所示。对所有影像人工检查后发现, 部分影像仍然受云的较大影响, 因此本研究进一步选取云量小于 1% 的高质量影像来预测 SOM, 其获取日期分别为 2014 年 10 月 6 日和 2017 年 10 月 30 日。为了尽可能去除云的影响, 研究中还针对选取的两景影像利用 Landsat 8 影像数据中的 QA_PIXEL 波段(质量评价波段), 对云和云阴影进行掩膜处理, 其中 2014 年 10 月 6 日获取的影像有 1 个采样点被云覆盖, 2017 年 10 月 30 日没有采样点被云覆盖, 因此, 在去除该采样点后共计 116 个采样点用于数据分析。

表 1 Landsat 8 影像采集时间与云量
Table 1 Dates and cloud covers of collected Landsat 8 images

影像采集时间	云量(%)
2013 年 11 月 4 日	18.32
2013 年 11 月 20 日	11.84
2014 年 10 月 6 日	0.51
2014 年 10 月 22 日	5.44
2015 年 10 月 9 日	2.42
2015 年 11 月 26 日	10.13
2017 年 10 月 30 日	0.61

注: 图中字体加粗的影像为选择的两景高质量影像。

1.4 光谱指数构建

已有研究表明, 与直接使用影像波段反射率相比, 光谱指数能够提高土壤遥感反演精度^[6-7]。本研究中, 采用比值光谱指数(Ratio Spectral Index, RI)、差值光谱指数(Difference Spectral Index, DI)、归一化光谱指数(Normalized Spectral Index, NDI)以及优化光谱指数(Optimized Spectral Index, OSI)^[18]进行 SOM 预测。构建的光谱指数包括单时相光谱指数和双时相光谱指数。单时相光谱指数指的是同一景影像不同波段之间构建的指数, 双时相光谱指数是指两景不同影像波段之间构建的指数。计算公式如表 2 所示。

研究中共有 2014 年 10 月 6 日和 2017 年 10 月 30 日获取的两景有效遥感影像, 每一景影像可以得到单时相光谱指数 168 个, 其中 RI、DI、NDI 以及

OSI 指数各 42 个。两景遥感影像共计得到双时相光谱指数 196 个, 各类光谱指数均为 49 个。

表 2 光谱指数计算公式
Table 2 Calculation formula of spectral indices

光谱指数	计算公式
比值光谱指数 RI_{ij}	ρ_i/ρ_j
差值光谱指数 DI_{ij}	$\rho_i-\rho_j$
归一化光谱指数 NDI_{ij}	$(\rho_i-\rho_j)/(\rho_i+\rho_j)$
优化光谱指数 OSI_{ij}	$(\rho_i+\rho_j)/\rho_j$

注: ρ_i 表示第 i 波段反射率, ρ_j 表示第 j 波段反射率。

1.5 变量筛选

变量选择是模型构建的一个重要方面, 特别是当自变量个数较多时, 通过变量筛选能够有效降低变量个数, 提高模型构建效率及稳定性。本研究选择 LASSO 算法进行变量筛选, 并与未经变量筛选的模型结果进行对比分析。LASSO 算法在化学计量学相关研究中已有应用, 其在最小化模型残差平方和的基础上增加了惩罚函数, 将模型回归系数的绝对值约束在一定的阈值, 能够有效改善模型精度和稳定性^[19]。其通过优化目标函数将相关性较小的变量系数压缩至 0, 通过删除相关系数为 0 的变量, 达到筛选变量的目的, 可以有效防止过拟合问题。LASSO 变量筛选通过使用 R 语言中的“glmnet”包实现。

1.6 模型建立与验证

为探究双时相光谱指数在 SOM 反演中的表现, 本研究分别针对影像波段、单时相光谱指数以及双时相光谱指数构建模型。针对影像波段建模以及针对单时相光谱指数建模时, 分别使用单景影像以及两景影像相关波段和指数构建模型; 针对双时相光谱指数构建模型时, 直接使用基于两景影像构建的指数建立模型。同时为了对比分析变量筛选在模型构建中的作用, 除直接使用影像波段或光谱指数构建模型外, 还构建了经 LASSO 算法变量筛选后的模型。

本研究中采用的建模算法为 SVM, SVM 算法是一种经典的机器学习算法, 在许多应用场景中均有良好的表现^[20]。该算法使用核函数将数据投影到高维空间, 以找到用于对不同组进行分类或建立最佳回归模型的超平面^[21]。SVM 模型构建采用 R 语言中的“caret”包实现, 选用线性核函数, 并采用留一交叉验证对模型参数进行优化。模型精度评价采用留一交叉验证的结果, 根据决定系数(R^2)和均方根误差(RMSE) 指标评估 SOM 反演精度, 计算公式如下:

$$R^2=1-\frac{\sum_{i=1}^n(Q_i-P_i)^2}{\sum_{i=1}^n(Q_i-\bar{Q}_i)^2} \quad (1)$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^n(Q_i-P_i)^2} \quad (2)$$

式中: Q_i 为观测值; P_i 为实测值; n 为样本量。

2 结果与分析

2.1 SOM 含量统计

研究区 SOM 含量范围为 8.60~28.40 g/kg, 中值为 16.50 g/kg, 均值为 16.40 g/kg, 标准差为 3.01 g/kg, 变异系数为 18.36%, 属于中等程度的变异^[18]。SOM 的峰度系数为 -0.65, 偏度系数为 0.41, 其含量分布符合正态分布(图 2), 因此建模时未对测量结果进行转换处理。同时, 为去除异常值对建模的影响, 选择格拉布斯检验(Grubbs Test)方法对 SOM 结果进行异常值检测, 在 $P<0.001$ 的显著性水平下未发现异常值的存在^[22], 因此本研究直接使用去除云影响之后的所有 116 个样本进行建模分析。

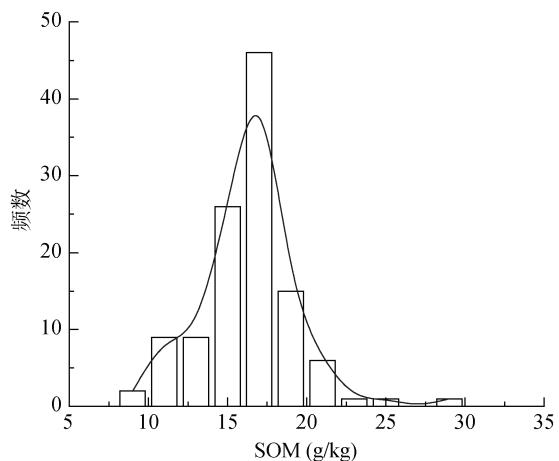


图 2 有机质统计直方图
Fig. 2 Histogram of SOM

2.2 基于影像波段的 SOM 预测

分别以单期影像和两期影像的波段反射率构建模型, SOM 预测的交叉验证结果如表 3 所示。单期影像中, 基于 2014 年影像的 SOM 预测精度较低, R^2 仅为 0.05, RMSE 为 2.88 g/kg; 与之相比, 基于 2017 年影像的 SOM 预测精度相对较高, $R^2=0.30$, RMSE=2.47 g/kg。两期影像同时使用时, 与基于 2017 年单期影像相比, SOM 预测精度并未改善($R^2=0.27$, RMSE=2.52 g/kg)。经筛选变量后, 基于单期影像和两期影像的 SOM 预测精度均有了一定程度的提升,

其中两期影像同时使用时 SOM 预测 R^2 提高到 0.32, RMSE 降低为 2.44 g/kg, 但基于 2014 年影像的 SOM 预测精度仍然较低, $R^2=0.09$, RMSE=2.82 g/kg。

表 3 基于影像波段模型的 SOM 预测交叉验证结果
Table 3 Cross-validation results of SOM predicted by models of image bands

变量	变量筛选方法	变量个数	R^2	RMSE(g/kg)
2014 年影像波段	无	7	0.05	2.88
	LASSO	2	0.09	2.82
2017 年影像波段	无	7	0.30	2.47
	LASSO	4	0.32	2.44
2014 年和 2017 年影像波段	无	14	0.27	2.52
	LASSO	7	0.32	2.44

2.3 基于单时相光谱指数的 SOM 预测

针对单时相光谱指数, 分别利用单期影像和两期影像构建模型, SOM 预测的交叉验证结果如表 4 所示。对于 2014 年影像, 所有光谱指数的 SOM 预测精度均相对较低, $R^2=0.17$, RMSE=3.02 g/kg, 经变量筛选后模型预测 R^2 提升至 0.21, RMSE 降低为 2.63 g/kg。基于 2017 年影像的 SOM 预测精度优于基于 2014 年影像的预测结果, 且变量筛选后模型的预测精度也有所提升, 变量筛选前、后的预测精度分别为: $R^2=0.35$, RMSE=2.38 g/kg 和 $R^2=0.37$, RMSE=2.34 g/kg。当两景影像同时使用时, 特别是经变量筛选后, 模型的 SOM 预测精度显著高于基于单景影像的预测结果, $R^2=0.45$, RMSE=2.20 g/kg。

2.4 基于双时相光谱指数的 SOM 预测

双时相光谱指数经变量筛选前后的 SOM 预测模型评价结果如表 5 所示。与使用波段或者单时相光谱指数的预测结果相比, 有机质的预测精度有了较大程度的提升, 未经变量筛选时, SOM 的预测 R^2 可达

0.45, RMSE 为 2.19 g/kg, 经变量筛选后 SOM 的预测精度更高, R^2 为 0.53, RMSE 降低至 2.01 g/kg。双时相光谱指数模型的有机质预测散点图如图 3 所示, 可以看出, 整体上有机质预测值与实测值较为接近, 特别是经变量筛选后模型的 SOM 预测值更接近 1:1 线, 表现出更优的预测精度。

2.5 基于最优预测模型的 SOM 空间分布预测

高精度的 SOM 空间分布信息对土壤资源高效管理具有重要意义。当前许多研究基于土壤-景观关系, 利用数字土壤制图技术进行土壤空间分布图件的制作, 在相关研究中土壤成土环境因子特别是地形相关变量被广泛采用^[23]。然而, 地形成土环境变量在平缓地区土壤制图中往往面临较大困难。研究区位于黄淮海平原, 其海拔范围在 65 ~ 72.5 m^[24], 变异相对

表 4 基于单时相光谱指数模型的 SOM 预测交叉验证结果

Table 4 Cross-validation results of SOM predicted by models of single-temporal spectral indices

变量	变量筛选方法	变量个数	R^2	RMSE(g/kg)
2014 年影像构建的单时相光谱指数	无	168	0.17	3.02
	LASSO	15	0.21	2.63
2017 年影像构建的单时相光谱指数	无	168	0.35	2.38
	LASSO	16	0.37	2.34
2014 年和 2017 年各自影像构建的单时相光谱指数	无	336	0.35	2.69
	LASSO	30	0.45	2.20

表 5 基于双时相光谱指数模型的 SOM 预测交叉验证结果

Table 5 Cross-validation results of SOM predicted by models of dual-temporal spectral indices

变量	变量筛选方法	变量个数	R^2	RMSE(g/kg)
2014 年和 2017 年两期影像构建的双时相光谱指数	无	196	0.45	2.19
	LASSO	26	0.53	2.01

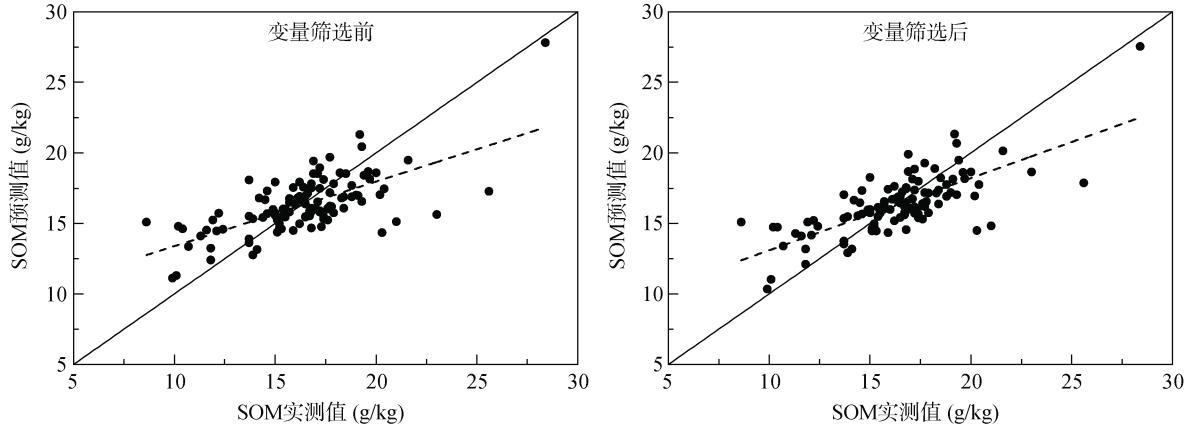


图 3 双时相光谱指数 SOM 预测散点图
Fig. 3 Scatter diagram of SOM predicted by models of dual-temporal spectral indices

较小,因此难以利用地形相关变量进行土壤属性空间分布制图。在与本研究区类似的平缓地区,遥感可以有效获取土壤信息,直接反演土壤属性空间变异。双时相光谱指数的 SOM 预测结果展示出了较高的预测精度(表 5),具有制作 SOM 空间分布图的能力。

利用基于变量筛选后的双时相光谱指数构建 SOM 预测模型对封丘县内耕地土壤进行 SOM 预测,结果如图 4 所示。结果表明,封丘县内 SOM 含量分布差异较大,SOM 含量较高的区域主要分布在封丘县南部与黄河平行一线,以及东北部部分区域,而紧靠黄河沿线的 SOM 含量较低。该预测结果空间上基本与实测值分布一致,统计后发现封丘县内 SOM 预测值的平均结果与 SOM 实测值的平均值相同,均为 16.40 g/kg,表明基于双时相光谱指数的 SOM 空间预测结果较好,相关方法可以用作 SOM 制图。

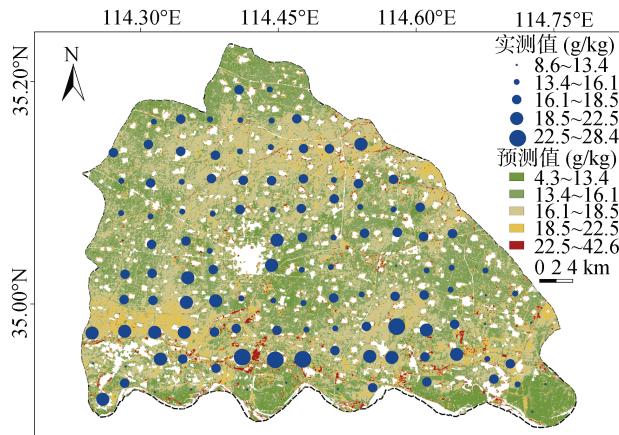


图 4 研究区 SOM 遥感空间预测结果

Fig. 4 Spatial distribution of SOM by remote sensing in study area

3 讨论

利用光学遥感影像进行土壤属性定量遥感反演时容易受地表植被和土壤水分等因素的影响。本研究中选择的遥感影像在潜在裸土期时间窗口内(10月),但直接利用影像波段进行 SOM 预测的精度较低,最高精度模型 $R^2=0.32$, RMSE=2.44 g/kg,这可能是土壤表面覆盖物或者土壤自身状态的影响所致。2014 年 10 月 6 日获取的遥感影像,其 SOM 预测精度较低, R^2 仅为 0.05, RMSE 高达 2.88 g/kg,而 2017 年 10 月 30 日获取的遥感影像预测精度相对较高,两景影像精度差异的原因很可能是由于 10 月 6 日获取的影像受到了玉米秸秆的影响,该影像获取时可能部分区域秸秆尚未收获或者土壤未开始翻耕,而 10 月 30 日时土壤均已翻耕完毕且冬小麦尚未影响到土壤光

谱。两景影像的植被指数差异印证了这一推测,10 月 6 日影像的平均归一化光谱指数为 0.13,而 10 月 30 日的对应值为 0.11。因此,对于黄淮海平原一年两熟区,尽管区域内普遍实行抢收抢种的农田管理模式,但小农户种植管理比重大,因此,某一时刻的土壤表面状况也可能存在较大差异,遥感影像的时相选择显得至关重要。

已有研究发现通过构建光谱指数,能够一定程度上消除外部因素对土壤遥感精度的影响,提高目标土壤属性反演精度^[6,9~10]。本研究中,与直接使用波段反射率构建的模型相比,基于单时相光谱指数的 SOM 预测精度明显更高,当两景影像构建的指数共同使用并经变量筛选后,SOM 的预测精度达到一个相对较高的水平($R^2=0.45$, RMSE=2.20 g/kg)。但与徐夕博^[25]在山东省潍坊市北部平原地区利用 Landsat 8 影像构建单时相光谱指数模型预测 SOM 的结果相比,本研究中的预测精度相对较低,其模型最高精度 $R^2=0.65$, RMSE=2.50 g/kg。邱壑等^[26]在福建省三明市基于 Hyperion 高光谱遥感数据进行 SOM 预测时,同样发现与基于波段反射率建模相比,基于单时相光谱指数建模可以有效提高反演精度,其最高模型精度 $R^2=0.55$, RMSE=8.21 g/kg,而基于影像波段建模最高模型精度 $R^2=0.17$, RMSE=11.14 g/kg。该结果与本研究中的预测精度较为接近,这可能是由于尽管作者使用了高光谱卫星影像,但其研究区与黄淮海平原相比,耕地土壤更易受到作物覆盖和土壤自身状态的影响。本研究发现,与仅利用一景影像的单时相光谱指数相比,两景影像的单时相光谱指数共同使用时 SOM 预测精度得到了提升(表 4),该结果与 Wang 等^[27]的研究类似,其在黑龙江省绥化市基于 Sentinel-2 影像预测 SOM 的研究中发现,同时使用多景影像的单时相光谱指数构建的模型精度($R^2=0.49$, RMSE=3.53 g/kg)优于单景单时相光谱指数的结果($R^2=0.31$, RMSE=3.84 g/kg)。单景影像仅能记录地表的瞬时状态,而多景影像可以反映地表的动态变化过程,因此,更多时相信息的加入可以改善 SOM 预测精度。

对比影像波段模型和单时相光谱指数模型,尽管使用了相同的两景影像,双时相光谱指数 SOM 的反演精度更高,最高模型精度 $R^2=0.53$, RMSE=2.01 g/kg。因此,遥感影像时相信息的利用方式也可能对土壤属性的预测精度产生影响。实际上,双时相光谱指数利用的是目标土壤属性对外部因素(如土壤水分、土壤粗糙度和秸秆覆盖度等)的响应差异来间接反映土壤属性的差异,而单时相光谱指数利用的是土壤属性在

不同波段之间的差异。刘焕军等^[10]在我国东北黑土区,选择了同一年内多景裸土期内的 MODIS 影像构建双时相光谱指数, SOM 预测 R^2 可达 0.72, RMSE 为 4.42 g/kg。在黑土区田块尺度,潘越^[28]利用同一年份的 Landsat 8 和 Sentinel-2 遥感影像,构建双时相光谱指数对 SOM 预测同样获得了较高的精度 ($R^2=0.80$, RMSE=8.98 g/kg)。黑土区为一年一熟区且区域内耕地多由农场统一管理,因此,一定区域内的农田耕作方式比较一致。但与我国东北黑土区不同,本研究区所在的黄淮海平原耕地普遍采用一年两熟的种植模式,两季作物之间的裸土窗口期短,同一年内获取多景裸土期高质量影像难度较大。当选择构建双时相光谱指数的遥感影像时,影像的获取时间一方面受到影像可获取性、云等干扰因素以及裸土期等因素的影响,另一方面还可能会受到土壤属性随时间变化的影响。但对于 SOM,其短期内含量变化较小,影像选择时间可以不局限于采样当年。针对 SOM 预测,本研究中影像选择时间设置为 5 年,但去除受云等干扰因素影响的影像后仅有两景高质量影像,基于这两景非同年获取的影像构建的双时相光谱指数同样得到了较高的 SOM 预测结果。实际上,即便在裸土期较长的东北地区,使用更长时间间隔的影像构建双时相光谱指数有可能得到更好的预测结果。例如,Dou 等^[6]在针对黑土区 SOM 遥感反演预测中,基于 2002 年、2009 年和 2014 年 3 年的 MODIS 影像构建的双时相光谱指数,使用 2002 年影像时 SOM 预测结果最优($R^2=0.76$, RMSE=7.66 g/kg)。因此,针对 SOM 等随时间变化较慢的土壤属性时,可以选择长时序的遥感影像进行反演,但对于随时间变化较快的土壤属性,遥感影像的时相选择是否对预测结果产生影响还需要进一步研究。

利用遥感影像进行土壤属性预测反演制图时,受制于成本限制,通常建模样品数量较少。本研究中,实际利用 116 个土壤样品构建模型,与光谱指数的数量相比,建模样本相对较少,属于小样本高维模型。这类模型构建时需要进行变量筛选,以减少变量共线性的影响,降低模型过拟合概率,提高模型稳定性和精度^[29]。本研究利用 LASSO 算法筛选变量,然后利用 SVM 机器学习方法构建模型,取得了较高的 SOM 反演精度,所构建模型预测的 SOM 空间分布格局与真实分布较为吻合。在今后研究中,可以尝试使用其他筛选变量算法,例如遗传算法和竞争性自适应重加权算法,进一步分析变量筛选对模型精度的影响。此外,本研究针对土壤 SOM 预测反演仅使用了 Landsat

8 影像,在今后研究中可以尝试更高时空分辨率的影像(如 Sentinel-2 影像),进一步研究双时相光谱指数的应用潜力,实现区域 SOM 高精度反演和高空间分辨率预测制图的结合。

4 结论

1)与直接使用影像波段反射率或者单时相光谱指数相比,利用双时相光谱指数可以更好利用影像时相信息的优势,SOM 预测精度更高。针对影像波段和光谱指数,经 LASSO 变量筛选之后,SOM 预测模型精度更优。

2)在与研究区类似的地形平缓区域,利用双时相光谱指数构建的遥感预测模型,可以有效用于区域 SOM 预测与制图。对于 SOM 等随时间变化相对缓慢的土壤属性,遥感影像可以选择不同于采样年份内的(准)裸土期影像。

参考文献:

- [1] Bünemann E K, Bongiorno G, Bai Z G, et al. Soil quality - A critical review[J]. *Soil Biology and Biochemistry*, 2018, 120: 105–125.
- [2] 曹振, 王昌昆, 马海艺, 等. 基于 Landsat-8 数据的土壤颜色预测方法研究[J]. 土壤, 2022, 54(1): 152–160.
- [3] McCarty G W, Reeves J B, Reeves V B, et al. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement[J]. *Soil Science Society of America Journal*, 2002, 66(2): 640.
- [4] Vaudour E, Gomez C, Loiseau T, et al. The impact of acquisition date on the prediction performance of topsoil organic carbon from sentinel-2 for croplands[J]. *Remote Sensing*, 2019, 11(18): 2143.
- [5] 王森, 潘贤章, 解宪丽, 等. 土壤含水量对反射光谱法预测红壤土壤有机质的影响研究[J]. 土壤, 2012, 44(4): 645–651.
- [6] Dou X, Wang X, Liu H J, et al. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China[J]. *Geoderma*, 2019, 356: 113896.
- [7] 杨邵文. 基于多源光谱数据的土壤属性建模优化与实证研究[D]. 淮南: 安徽理工大学, 2021.
- [8] 赵瑞, 崔希民, 刘超. GF-5 高光谱遥感影像的土壤有机质含量反演估算研究[J]. 中国环境科学, 2020, 40(8): 3539–3545.
- [9] 张新乐, 窦欣, 谢雅慧, 等. 引入时相信息的耕地土壤有机质遥感反演模型[J]. 农业工程学报, 2018, 34(4): 143–150, 315.
- [10] 刘焕军, 宁东浩, 康苒, 等. 考虑含水量变化信息的土壤有机质光谱预测模型[J]. 光谱学与光谱分析, 2017, 37(2): 566–570.
- [11] Shi Y Y, Zhao J Y, Song X C, et al. Hyperspectral band selection and modeling of soil organic matter content in a

- forest using the Ranger algorithm[J]. PLoS One, 2021, 16(6): e0253385.
- [12] Yuan J, Wang X, Yan C X, et al. Wavelength selection for estimating soil organic matter contents through the radiative transfer model[J]. IEEE Access, 2020, 8: 176286–176293.
- [13] Wang K C, Liu L, Yuan C J, et al. Software defect prediction model based on LASSO-SVM[J]. Neural Computing and Applications, 2021, 33(14): 8249–8259.
- [14] Wang P, Tan S Q, Zhang G, et al. Remote sensing estimation of forest aboveground biomass based on lasso-SVR[J]. Forests, 2022, 13(10): 1597.
- [15] Ferhatoglu C, Miller B A. Choosing feature selection methods for spatial modeling of soil fertility properties at the field scale[J]. Agronomy, 2022, 12(8): 1786.
- [16] Meng L, Ding W X, Cai Z C. Long-term application of organic manure and nitrogen fertilizer on N₂O emissions, soil quality and crop production in a sandy loam soil[J]. Soil Biology and Biochemistry, 2005, 37(11): 2037–2045.
- [17] 任频频, 黄峰, 李保国. 黄淮海平原旱作农田土壤有机质含量的空间分异特征[J]. 土壤学报, 2022, 59(2): 440–450.
- [18] 尼加提·卡斯木, 茹克亚·萨吾提, 师庆东, 等. 基于优化光谱指数的土壤有机质含量估算[J]. 农业机械学报, 2018, 49(11): 155–163.
- [19] Yan Z B, Yao Y. Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO)[J]. Chemometrics and Intelligent Laboratory Systems, 2015, 146: 136–146.
- [20] Dotto A C, Dalmolin R S D, Grunwald S, et al. Two preprocessing techniques to reduce model covariates in soil property predictions by Vis-NIR spectroscopy[J]. Soil and Tillage Research, 2017, 172: 59–68.
- [21] Smola A J, Schölkopf B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199–222.
- [22] Wilrich P T. Critical values of Mandel's h and k , the Grubbs and the Cochran test statistic[J]. AStA Advances in Statistical Analysis, 2013, 97(1): 1–10.
- [23] 陈荣, 韩浩武, 傅佩红, 等. 基于多时相遥感影像和随机森林算法的土壤制图[J]. 土壤, 2021, 53(5): 1087–1094.
- [24] Wang C K, Li Z T, Pan X Z. Estimating crop residue cover using SPOT 5 data[J]. Journal of Soil and Water Conservation, 2017, 72(4): 343–350.
- [25] 徐夕博. 基于高光谱的潍北平原土壤有机质预测与空间格局研究[D]. 济南: 山东师范大学, 2019.
- [26] 邱壑, 陈瀚阅, 邢世和, 等. 基于 Hyperion 数据的耕地土壤有机质含量遥感反演[J]. 福建农林大学学报(自然科学版), 2017, 46(4): 460–467.
- [27] Wang X, Wang L P, Li S J, et al. Remote estimates of soil organic carbon using multi-temporal synthetic images and the probability hybrid model[J]. Geoderma, 2022, 425: 116066.
- [28] 潘越. 耕地田块尺度土壤有机质遥感反演研究[D]. 哈尔滨: 东北农业大学, 2019.
- [29] 赵明松, 谢毅, 陆龙妹, 等. 基于高光谱特征指数的土壤有机质含量建模[J]. 土壤学报, 2021, 58(1): 42–54.