

任冠尧,秦婧,钱小将,等.基于多源光谱数据和机器学习算法的土壤有机碳反演研究:以内蒙古自治区东北部为例[J].华中农业大学学报,2026,45(2):137-145.DOI:10.13300/j.cnki.hnlkxb.2026.02.014

基于多源光谱数据和机器学习算法的土壤有机碳反演 研究:以内蒙古自治区东北部为例

任冠尧¹,秦婧¹,钱小将²,张克³,闫伟岳¹,冯子昂⁴,郭龙⁴

1. 内蒙古自治区测绘地理信息中心,呼和浩特010010; 2. 湖北省国土测绘院,武汉430014;
3. 国标(北京)检验认证有限公司,北京100010; 4. 华中农业大学资源与环境学院,武汉430070

摘要 为探究外界环境因素(光照、大气条件)和传感器参数(光谱与空间分辨率)对不同来源光谱数据反演土壤有机碳(SOC)精度的影响机制,以内蒙古自治区东北部为研究区,采集160个表层土壤(0~20 cm)SOC样品,并同步获取近端高光谱(室内人造光源与室外太阳光源)以及星载多光谱(Landsat-8、Sentinel-2)与高光谱(ZY1-02D)数据,使用随机森林(RF)与支持向量机(SVM)算法,分别构建SOC反演模型,通过系统比较不同数据源的模型性能,分析环境因素与传感器参数对SOC反演精度的影响。结果显示:室内人造光源因光谱信号稳定、可控性强,其反演精度略优于室外太阳光源,但二者差异较小,表明自然光照波动对SOC反演的影响有限;近地高光谱数据反演精度显著高于卫星多光谱与高光谱数据,主要因卫星数据受大气散射、水汽吸收及混合像元等问题干扰;在卫星数据中,高光谱卫星ZY1-02D的反演精度高于多光谱卫星,而Sentinel-2较Landsat-8的空间分辨率提升(10 m vs 30 m)对模型性能改善有限,说明光谱分辨率对SOC精度的贡献大于空间分辨率。

关键词 多源光谱数据;土壤有机碳;机器学习算法;反演方法;资源一号02D;光谱分辨率

中图分类号 TP181;S153.6 **文献标识码** A **文章编号** 1000-2421(2026)02-0137-09

土壤有机碳(soil organic carbon, SOC)是全球碳循环中的关键组成部分,具有重要的生态功能和潜在的经济价值^[1]。据估计,全球SOC贮存量约为1 500亿~2 000亿t,超过全球植被和水体碳储存总量的2倍,约占陆地生态系统碳库的2/3。面对全球气候变化这一严峻挑战,国际社会亟需共同采取行动,以减少温室气体排放并提升气候适应与缓解能力^[2-4]。在此背景下,SOC作为土壤和生物圈的主要碳库,其固碳作用对减缓气候变化具有重要影响^[5-6]。同时,SOC在农业生产体系中发挥着基础性作用,它能促进土壤团粒结构形成,增强土壤的通气性和保水能力,从而为作物根系创造优良的生长环境,最终提高农作物的产量与品质^[7-9]。因此,实现SOC含量的快速准确反演,既是优化农业精准管理的关键前提,也是动态监测土壤健康的核心手段,更是深化区域及全球碳循环研究的根本支撑。

可见光-近红外光谱(Vis-NIR,波长范围350~

2 500 nm)能够捕捉土壤中有机物质对电磁波的特征响应,被广泛用于SOC含量的快速准确反演^[10-11]。SOC中的C-H、O-H等官能团在特定波段发生分子振动和电子跃迁,形成特征吸收峰(如2 200 nm附近的C-H伸缩振动),使得土壤光谱在这些波段表现出与SOC含量相关的特征,这构成了基于光谱信息反演SOC的理论基础^[12-13]。近年来,机器学习算法在土壤光谱信息挖掘中的应用日益广泛,以随机森林(random forest, RF)、支持向量机(support vector machine, SVM)和深度神经网络为代表的模型,能够从多维光谱数据中提取与SOC含量变化高度相关的特征,实现对土壤空间异质性的快速反演^[14-15]。与传统地统计学方法相比,机器学习方法在建模精度和推广能力方面具有显著优势,已被广泛应用于多种空间尺度的SOC光谱反演与制图工作中^[16-18]。

尽管Vis-NIR技术结合机器学习算法为SOC快速反演提供了有效手段,但在实际应用中,SOC的反

收稿日期:2025-07-16

基金项目:国家重点研发计划项目(2023VD1500102;2023VFC3709801);国家自然科学基金项目(42101064;42361028);中国博士后科学基金独立项目(2021M693244)

任冠尧,E-mail:guanyaoren@sina.com

通信作者:冯子昂,E-mail:2021309210605@webmail.hzau.edu.cn

演精度会因光谱数据的来源不同、光谱成像环境的差异等产生不确定性^[19-20]。土壤光谱数据主要来源于近端传感和星载遥感。近端传感数据具有波段连续、信号受控的优势,准确性高,但其采样范围极其有限,难以满足区域乃至全球尺度SOC动态监测的需求^[21]。星载遥感数据具有大范围、周期性覆盖的能力,然而,其反演过程极易受到复杂环境因素的干扰:气溶胶散射、水汽吸收等大气因素会显著衰减和扭曲地表光谱信号;光谱分辨率、空间分辨率等传感器特性差异直接影响对土壤关键光学特征的捕捉能力^[22-23]。对于近端传感,光照条件(使用人造光源或太阳光)也可能引入不确定性。在中小尺度的SOC反演和制图研究中,如何选取上述数据以获得最优的模型精度是亟需探讨的问题。

针对上述问题,本研究以内蒙古自治区东北部为例,基于近端传感(使用高性能地物光谱仪PSR+3500分别在室内人造光源和室外太阳光条件下测得)和星载遥感(资源一号02D(ZY1-02D)高光谱相机、Landsat-8陆地成像仪(OLI)和Sentinel-2多光谱成像仪(MSI))土壤光谱数据,使用SVM和RF算法,分别构建SOC反演模型,并制作研究区域内的茂密草原、稀疏草原和农田3个典型代表区的SOC图,旨在系统地探讨基于多源光谱数据构建的SOC反演模型之间的精度差异及其驱动机制,以期为区域土壤有机碳的高精度遥感制图与最优数据源选择提供科学依据。

1 材料与方法

1.1 研究区域

选取内蒙古自治区东北部的呼伦贝尔市、兴安盟、锡林郭勒盟、赤峰市、通辽市为研究区域(41.3°~53.3°N, 115.1°~125.9°E)。该区域属温带大陆性季风气候,受到大陆性气候和复杂地形的显著影响。气候表现出强烈的季节特征:冬季寒冷漫长,夏季短暂炎热。年降水量偏低且时空分布不均,干旱和风沙天气较常见。地貌以高原型为主,地形丰富多样;主要土壤类型包括栗钙土、黑钙土及风沙土,其中,栗钙土有机质含量中等(1.5%~3.5%),黑钙土肥沃深厚(有机质2%~5%),风沙土贫瘠且易受侵蚀(有机质<1%)。内蒙古东部主要以广袤草原为景观特征(与内蒙古西部沙漠景观形成对比),是本研究区的核心地表覆盖类型。耕地是区内重要土地利用类型,主要分布在松嫩平原嫩江右岸及西辽河平原等

相对平坦区域。为了更精细地探究典型土地利用条件下的SOC状况及其空间反演效果,在全区范围内选取了3个具有代表性的重点样区进行详细制图:茂密草原区(49.5°~50.3°N, 119.5°~120.8°E)、稀疏草原区(41.6°~42.3°N, 118.5°~119.4°E)和农田区(44.2°~45.0°N, 115.5°~116.9°E)。为获取具有区域代表性的土壤光谱与属性数据,本研究在全域范围内共采集表层土壤(0~20 cm)样本160个。采样点布局综合考虑了空间分布的均匀性与土地利用类型的典型性,其中草地样本104个、耕地样本33个、林地及其他类型样本23个,样本分布范围与上述研究区地理范围一致。田间采样时,以GPS定位点为中心,采用梅花形五点混合采样法收集土壤,并同步获取样点处的野外光谱数据。所有土样经自然风干、研磨、过2 mm筛处理后,使用元素分析仪测定SOC含量。

1.2 土壤光谱数据

本研究使用的近端传感光谱数据由高性能地物光谱仪PSR+3500采集,其光谱范围覆盖可见光-近红外-短波红外(350~2 500 nm),光谱分辨率分别达3 nm@700 nm、8 nm@1 500 nm和6 nm@2 100 nm。在标准暗室条件下,采用卤素人造光源(光谱稳定性>99%)测量土壤样本;室外测量同步于土壤采样当日正午(12:00)左右进行,以最小化太阳高度角产生的影响。星载遥感数据来自ZY1-02D高光谱相机、Landsat-8 OLI和Sentinel-2 MSI。ZY1-02D高光谱相机含166个波段(400~2 500 nm),空间分辨率为30 m;Landsat-8 OLI和Sentinel-2 MSI均为多光谱传感器,在Vis-NIR范围内分别含有11和13个波段,空间分辨率分别为30和10 m。所有遥感影像获取时间与野外采样同步(±15 d内),经辐射定标、大气校正及正射校正后,提取样点位置裸土像元光谱值。

1.3 土壤有机碳反演模型

本研究使用RF和SVM 2种成熟的机器学习算法构建SOC反演模型^[24-27]。所有模型构建均基于R语言环境实现,其中RF模型通过调用randomForest包完成构建,SVM模型通过调用e1071包完成构建,并结合caret包进行模型训练流程的管理。

1.4 模型验证

本研究使用网格搜索结合十折交叉验证的方法训练机器学习模型,其流程如下:首先基于Kennard-Stone(KS)算法将全体样本划分为10个互斥子集,随后预设超参数搜索空间。针对每个超参数组合执

行10轮迭代训练:每轮选取1个子集作为验证集,剩余9个子集作为训练集,同时在验证集上计算模型性能指标。完成10轮迭代后,计算该超参数组合对应的平均验证集性能指标(10次验证集性能均值)。遍历所有预设超参数组合后,筛选验证集性能指标最高的超参数组合作为该机器学习模型最优解。对于使用星载遥感数据建立的模型,最终基于全体样本重新训练使用最优参数的模型,用于SOC空间制图。

使用 R^2 (R -squared)和RMSE(root mean square error,公式中用 R_{MSE} 表示)2个常用的评估指标衡量模型的拟合程度和预测性能^[17]。 R^2 的取值范围在0~1,越接近1表示模型对观测数据的变异性解释得越好,说明模型能够更好地拟合数据。RMSE是用于度量模型预测值与观测值之间的误差的指标,它计算了模型预测值与实际观测值之间的平方差,并取平均值后再开平方根。RMSE的值越小表示模型的预测能力越好,即模型的预测值与实际观测值之间的差距越小。 R^2 和RMSE的计算公式如下:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

式(1)~(2)中: n 为样本数量; y_i 为第 i 个样本的SOC实测值; \hat{y}_i 为第 i 个样本的SOC预测值; \bar{y} 为所有样本实测值的算术平均值。

2 结果与分析

2.1 基础统计分析

表1和表2分别呈现了研究区域内SOC含量在不同土壤类型及不同土地利用类型下的统计特征。从土壤类型的差异来看(表1),淋溶土的平均有机碳含量最高,达2.79%,这与该类土壤长期稳定的有机质积累环境密切相关;而初育土的平均有机碳含量最低,仅为1.08%,可能与其处于土壤发育初期、有机质积累尚未充分的特性相关。在离散程度与分布特征方面,半淋溶土表现最为突出,其标准差(2.42%)、峰度(2.22)和偏度(4.06)均为表1中的最大值,反映出该类SOC含量波动显著,且数据分布存在明显的极端值倾向,可能与成土条件或人类活动的复杂影响有关。从土地利用类型的视角分析(表2),林地因森林生态系统的固碳优势,平均有机碳含量最高

(3.01%),体现了植被覆盖对土壤有机质积累的促进作用;耕地受耕作活动干扰频繁,平均有机碳含量最低(1.66%),反映了农业管理措施对土壤碳储量的潜在影响。值得注意的是,耕地和草地的偏度值分别达到6.09%和8.27%,显著高于其他土地利用类型,表明这两类用地的有机碳含量分布呈现右偏特征,即中低值区间样本占比较大,而高值样本相对集中,可能与不同利用方式下养分投入或植被凋落物输入的不均衡性相关。此外,草地的样本量(104)为表2中最大,其统计结果对区域土壤碳分布的代表性较强。综合2类统计分析可见,土壤类型通过成土过程和发育阶段的差异作用,以及土地利用方式通过人类干扰和植被管理的综合作用,共同影响着SOC的积累、转化与保持能力,为区域土壤碳循环研究及可持续管理提供了重要的基础数据支撑。

2.2 重要光谱波段分析

为精准筛选反演SOC的敏感光谱波段,本研究基于偏最小二乘回归(PLSR)模型,计算暗室、室外及遥感卫星高光谱数据与SOC的变量投影重要性指数(VIP),相关结果如图1所示。其中,暗室高光谱数据(图1A)显示,VIP指数>1的敏感波段集中于202~756、1 015~1 058、1 516~1 553、1 724~1 802 nm,在454 nm处出现峰值(VIP=1.39),在1 040、1 541、1 765 nm附近出现次级峰值,表明这些波段对SOC预测具有关键作用。室外高光谱数据(图1B)的敏感波段主要分布于200~550、1 500~1 620及2 000~2070 nm,以及在350 nm处(VIP=1.4)、1 550 nm(VIP=2.0)和2 030 nm(VIP=2.2)处的显著峰值,反映了自然光照条件下SOC对光谱的响应特性;ZY1-02D卫星高光谱数据(图1C)显示,敏感波段呈离散分布,集中于第3~12波段(475~610 nm)、第21~22波段(760~775 nm)、第75~82波段(1 600~1 640 nm)、第99~102波段(2 000~2 060 nm)及第121~126波段(2 200~2 300 nm),其中第125波段(2 280 nm, VIP=2.71)为最高值,第6、22、80、125波段处的显著峰值表明卫星遥感波段对SOC反演的重要性受多因素耦合影响。研究采用VIP>1作为筛选阈值^[28],不同数据源的VIP差异揭示了光照条件、环境噪声及传感器特性对光谱响应的综合影响,为多源数据融合估算SOC提供了明确的波段选择依据,确保分析结果的科学性与可靠性。

2.3 土壤有机碳反演

本研究基于近端传感(室内人造光源与室外太

表1 不同土壤类型下土壤有机碳含量统计分析

Table 1 Statistical analysis of soil organic carbon content under different soil types

土纲 Soil order	样本数量 Sample size	平均值/% Mean	标准差/% Standard deviation	峰度 Kurtosis	偏度 Skewness	最大值/% Maximum	最小值/% Minimum
半水成土 Semi-hydromorphic soil	31	1.79	1.42	0.87	-0.38	5.47	0.26
水成土 Hydromorphic soil	12	2.19	1.14	-0.20	-1.17	4.01	0.09
初育土 Primosols	18	1.08	0.66	0.74	-0.53	2.63	0.31
钙层土 Pedocals	75	1.69	1.05	1.47	2.88	5.85	0.05
淋溶土 Leached soil	12	2.79	1.96	1.12	0.24	7.54	0.56
半淋溶土 Semi-leached soil	12	2.13	2.42	2.22	4.06	9.44	0.02

表2 不同土地利用类型下土壤有机碳含量基本统计分析

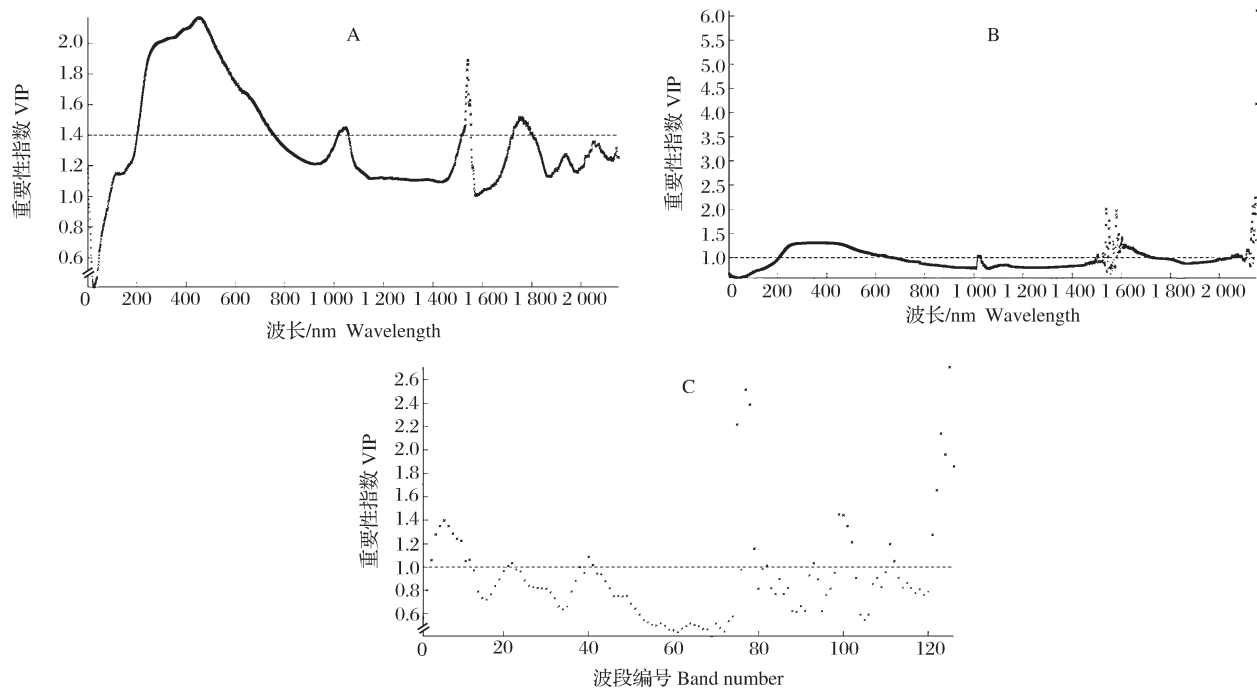
Table 2 Basic statistical analysis of soil organic carbon content under different land use types

土地利用类型 Land use type	样本数 Sample size	平均值/% Mean	标准差/% Standard deviation	峰度 Kurtosis	偏度 Skewness	最大值/% Maximum	最小值/% Minimum
耕地 Cultivated land	33	1.66	1.43	2.13	6.09	7.54	0.02
林地 Forest land	6	3.01	1.00	-0.24	-1.94	4.01	1.60
草地 Grassland	104	1.77	1.39	2.31	8.27	9.44	0.05
其他土地 Others	17	1.75	1.02	1.17	0.53	4.45	0.69

阳光源)和星载遥感(ZY1-02D 高光谱、Landsat-8 OLI多光谱、Sentinel-2 MSI多光谱)共5种光谱数据,结合RF与SVM算法构建SOC反演模型。SOC的反演精度结果如表3所示。从总体上分析,在使用相同机器学习模型时,近端传感的SOC反演精度优于星载遥感。这是因为星载遥感数据需穿透大气层,受气溶胶散射、水汽吸收等影响,导致地表光谱信号严重衰减,部分波段失真。而近端传感在厘米级距离测量,规避了大气路径干扰,光谱保真度更高^[29]。另一个原因是星载传感器空间分辨率有限,存在混合像元问题,即单个像元内常混合土壤、植被等地物,导致光谱无法纯粹表征土壤属性^[30]。此外,使用近端传感光谱数据反演SOC时,相较于易受自然光照波动干扰的室外太阳光光源(RF, $R^2=0.53$; SVM, $R^2=0.50$),光谱稳定性高的室内人造光源通过减少环境噪声干扰显著提升了光谱保真度,从而获得更好的反演结果(RF, $R^2=0.53$; SVM,

$R^2=0.57$)^[31]。

不同来源的卫星光谱数据对SOC的解释能力存在显著差异:高光谱卫星(ZY1-02D)的模型精度(RF, $R^2=0.43$; SVM, $R^2=0.40$)显著优于相同空间分辨率的多光谱卫星(Landsat-8)(RF, $R^2=0.28$; SVM, $R^2=0.37$),表明光谱分辨率提升可有效增强对SOC敏感波段的捕捉能力,进而提高模型精度^[32]。此外, Sentinel-2的SOC反演精度(RF, $R^2=0.34$; SVM, $R^2=0.38$)略优于Landsat-8(RF, $R^2=0.28$; SVM, $R^2=0.37$),表明将空间分辨率从30 m提升至10 m尽管能减轻混合像元的影响,但在实际应用中对SOC反演模型性能的提升有限,若没有更高空间分辨率的遥感数据源,应当优先考虑提升数据的光谱分辨率^[33]。值得注意的是,RF并非在所有条件下模型精度都优于SVM。这表明在较小样本数量的中小尺度SOC反演研究中,应多尝试几种机器学习方法,以构建当前光谱数据集下的最优模型。



A:室内人造光源 Indoor artificial light;B:室外太阳光 Outdoor sunlight;C:ZY1-02D卫星高光谱数据 ZY1-02D satellite hyperspectral data;红色区域表示 VIP>1 的敏感波段,蓝色区域表示 VIP≤1 的波段。ZY1-02D横坐标为波段编号,对应波长范围400~2 500 nm(共166个波段)。Red areas indicate sensitive bands with VIP>1,blue areas represent bands with VIP≤1. The x-axis for ZY1-02D shows band numbers,corresponding to wavelength range 400~2 500 nm(166 bands total).

图1 不同光源与卫星高光谱数据的SOC敏感波段VIP指数

Fig.1 VIP indices of sensitive bands for SOC from different light sources and satellite hyperspectral data

表3 不同光谱数据源下土壤有机碳反演精度比较

Table 3 Comparison of SOC inversion accuracy from different spectral data sources

数据源 Data sources		室内人造光源 Indoor artificial light	室外太阳光 Outdoor sunlight	ZY1-02D 高光谱 ZY1-02D hyperspectral	Landsat-8 多光谱 Landsat-8 multispectral	Sentinel-2 多光谱 Sentinel-2 multispectral
随机森林 Random forest	R^2	0.53	0.53	0.43	0.28	0.34
	RMSE/%	1.08	1.09	1.22	1.56	1.45
支持向量机 Support vector machine	R^2	0.57	0.50	0.40	0.37	0.38
	RMSE/%	1.05	1.10	1.24	1.44	1.42

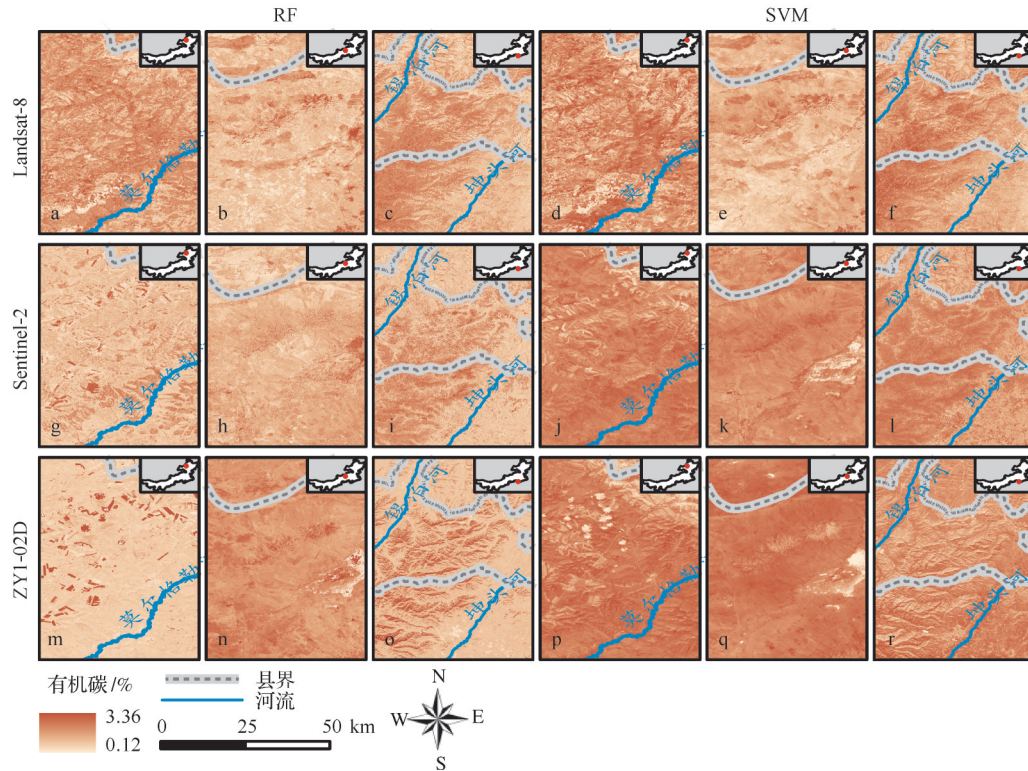
2.4 土壤有机碳制图

选取茂密草原、稀疏草原和农田3个典型区域,基于Landsat-8、Sentinel-2和ZY1-02D卫星光谱数据,使用RF和SVM进行SOC空间反演制图(图2)。SOC含量范围为0.12%~3.36%。结果显示:使用Landsat-8数据时,RF与SVM模型在3类区域的SOC反演结果高度一致,Landsat-8较低的光谱分辨率限制了模型对复杂光谱特征的捕捉能力,导致算法差异被削弱^[34]。对于Sentinel-2和ZY1-02D的SOC制图结果,SVM模型的预测值整体高于RF模型,且SOC空间分布差异显著。这可能是由于使用RBF核函数的SVM模型,相较于使用决策树的集成

学习模型(RF),在处理Sentinel-2和ZY1-02D这类光谱信息更丰富的高维光谱特征时具有更强鲁棒性,不易使模型出现过拟合,模型预测性能更优,能够更好地解释SOC的空间异质性^[35-36]。

3 讨论

本研究聚焦内蒙古自治区东北部区域,基于近端传感和星载遥感土壤光谱数据,使用SVM和RF算法构建SOC光谱反演模型,探讨了大气条件、光照条件、混合像元、空间分辨率和光谱分辨率对模型性能的影响,并制作了茂密草原、稀疏草原和农田3个典型代表区的SOC图。结果表明:受限于大气影响



子图 a, g, m, d, j, p 表示茂密草原; b, h, n, e, k, q 表示稀疏草原; c, i, o, f, l, r 表示农田。Subplots a, g, m, d, j, p represent dense grassland; b, h, n, e, k, q represent sparse grassland; c, i, o, f, l, r represent farmland.

图2 基于不同卫星数据与算法的土壤有机碳空间分布反演结果。

Fig.2 Spatial distribution of SOC inversion results from different satellite data and algorithms

和混合像元问题,星载遥感的SOC反演精度低于近端传感,但其仍是理想的SOC制图数据源,尤其是高光谱遥感数据(ZY1-02D)。室内人造光源因光谱信号稳定,是SOC反演最优数据源;室外太阳光源次之,精度略低但接近室内,主要受自然光照波动影响。光谱分辨率对SOC反演精度的提升具有决定性作用,其重要性高于空间分辨率;在缺乏高空间分辨率数据时,应优先考虑使用高光谱遥感数据。

尽管本研究系统比较了近端传感与卫星遥感在SOC反演中的性能,并可为不同数据源的应用提供实践指导,但仍存在一些值得深入探索的方向。未来的研究可考虑融合多源遥感数据是提升SOC估算精度和稳定性的重要途径。例如,可结合多光谱/高光谱卫星影像、无人机遥感数据等进行协同反演,利用它们各自在光谱分辨率、空间分辨率及覆盖范围上的优势,通过特征级融合技术,构建鲁棒性更强的模型,以弥补单一数据源在特定环境或技术限制下的不足。本研究尚未涉及无人机近地遥感平台的应用探讨,这类平台能有效填补地面测量与卫星遥感之间的尺度间隙,提供厘米级至亚米级的超高空间

分辨率数据,对于刻画农田、草原等生态系统中SOC的细碎空间异质性、验证卫星反演结果以及监测表层土壤碳动态变化具有巨大潜力。目前对亚米级高空间分辨率卫星在SOC制图中的应用研究仍较少,未来可重点研究如何利用该类卫星数据更好地解决地形破碎或地类复杂的区域SOC制图中的混合像元问题。

参考文献 References

- [1] 余玉洋,李晶,曾莉.中国南北过渡带生态系统固碳服务及价值量空间格局[J].陕西师范大学学报(自然科学版),2020,48(1):11-20.YU Y Y, LI J, ZENG L. Spatial pattern of carbon sequestration services and value of ecosystems of the North-South transitional zone in China[J]. Journal of Shaanxi Normal University (natural science edition), 2020, 48(1): 11-20 (in Chinese with English abstract).
- [2] 辛琨,颜葵,李真,等.海南岛红树林湿地土壤有机碳分布规律及影响因素研究[J].土壤学报,2014,51(5):1078-1086. XIN K, YAN K, LI Z, et al. Distribution of soil organic carbon in mangrove wetlands of Hainan Island and its influencing factors [J]. Acta pedologica sinica, 2014, 51 (5) : 1078-1086 (in

- Chinese).
- [3] 杨玉盛,谢锦升,盛浩,等.中亚热带山区土地利用变化对土壤有机碳储量和质量的影响[J].地理学报,2007,62(11):1123-1131. YANG Y S, XIE J S, SHENG H, et al. The impact of land use/cover change on soil organic carbon stocks and quality in mid-subtropical mountainous area of southern China[J]. *Acta geographica sinica*, 2007, 62(11):1123-1131 (in Chinese with English abstract).
- [4] 杨长明,欧阳竹,杨林章,等.农业土地利用方式对华北平原土壤有机碳组分和团聚体稳定性的影响[J].生态学报,2006,26(12):4148-4155. YANG C M, OUYANG Z, YANG L Z, et al. Organic carbon fractions and aggregate stability in an aquatic soil as influenced by agricultural land uses in the Northern China Plain[J]. *Acta ecologica sinica*, 2006, 26(12):4148-4155 (in Chinese with English abstract).
- [5] 曹新光,岳伟鹏,邓洁.北亚热带山地不同海拔土壤有机碳分布特征:以鄂东龟峰山为例[J].广西师范大学学报(自然科学版),2021,39(6):174-182. CAO X G, YUE W P, DENG J. Soil organic carbon distribution characteristics along an altitudinal gradient in the northern subtropical region: a case study in Guifeng Mountains of eastern Hubei, China [J]. *Journal of Guangxi Normal University (natural science edition)*, 2021, 39(6):174-182 (in Chinese with English abstract).
- [6] 刘满强,胡锋,陈小云.土壤有机碳稳定机制研究进展[J].生态学报,2007(6):2642-2650. LIU M Q, HU F, CHEN X Y. A review on mechanisms of soil organic carbon stabilization [J]. *Acta ecologica sinica*, 2007(6):2642-2650 (in Chinese with English abstract).
- [7] MA Y Q, WOOLF D, FAN M S, et al. Global crop production increase by soil organic carbon [J]. *Nature geoscience*, 2023, 16(12):1159-1165.
- [8] 温云杰,张建诚,杨娜,等.长期秸秆还田配施有机肥对土壤有机碳组分和孔隙结构的影响[J].农业工程学报,2024,40(21):74-81. WEN Y J, ZHANG J C, YANG N, et al. Effects of the long-term application of organic fertilization and straw returning on the components of soil organic carbon and pores [J]. *Transactions of the CSAE*, 2024, 40(21):74-81 (in Chinese with English abstract).
- [9] LIU C Z, HAN X Z, LU X C, et al. Response of soil enzymatic activity to pore structure under inversion tillage with organic materials incorporation in a Haplic Chernozem [J/OL]. *Journal of environmental management*, 2024, 370: 122421 [2025-07-16]. <https://doi.org/10.1016/j.jenvman.2024.122421>.
- [10] 包青岭,丁建丽,王敬哲,等.基于随机森林算法的土壤有机质含量高光谱检测[J].干旱区地理,2019,42(6):1404-1414. BAO Q L, DING J L, WANG J Z, et al. Hyperspectral detection of soil organic matter content based on random forest algorithm [J]. *Arid land geography*, 2019, 42(6):1404-1414 (in Chinese with English abstract).
- [11] 周磊,何秀芹,贾德伟,等.基于CWT-CARS的黄土高原东部陕州区农田表层有机碳含量高光谱反演[J].遥感技术与应用,2025,40(1):60-68. ZHOU L, HE X Q, JIA D W, et al. Hyperspectral retrieval of organic carbon content in farmland surface in Shanzhou District of eastern Loess Plateau based on CWT-CARS [J]. *Remote sensing technology and application*, 2025, 40(1):60-68 (in Chinese with English abstract).
- [12] STENBERG B, VISCARRA ROSSEL R A, MOUAZEN A M, et al. Visible and near infrared spectroscopy in soil science [J]. *Advances in agronomy*, 2010, 107:163-215.
- [13] VOHLAND M, BESOLD J, HILL J, et al. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy [J]. *Geoderma*, 2011, 166(1):198-205.
- [14] 甄诚,王海燕,雷相东,等.基于随机森林模型的旺业甸实验林场土壤全氮数字制图[J].华中农业大学学报,2024,43(3):249-257. ZHEN C, WANG H Y, LEI X D, et al. Digital mapping of soil total nitrogen in Wangyedian experimental forest farm based on random forest model [J]. *Journal of Huazhong Agricultural University*, 2024, 43(3):249-257 (in Chinese with English abstract).
- [15] 邹霞,钱海燕,周杨明,等.季节性淹水湿地表层土壤有机碳含量遥感预测及空间分布特征[J].华中农业大学学报,2024,43(3):111-120. ZOU X, QIAN H Y, ZHOU Y M, et al. Remote sensing prediction and spatial distribution characteristics of content of organic carbon in surface soil of seasonal flooded wetlands in Poyang Lake [J]. *Journal of Huazhong Agricultural University*, 2024, 43(3):111-120 (in Chinese with English abstract).
- [16] MINASNY B, MCBRATNEY A B, MALONE B P, et al. Digital mapping of soil carbon [J]. *Advances in agronomy*, 2013, 118:1-47.
- [17] VISCARRA ROSSEL R A, BEHRENS T, BEN-DOR E, et al. A global spectral library to characterize the world's soil [J]. *Earth-science reviews*, 2016, 155:198-230.
- [18] 刘艳芳,宋玉玲,郭龙,等.结合高光谱信息的土壤有机碳密度地统计模型[J].农业工程学报,2017,33(2):183-191. LIU Y F, SONG Y L, GUO L, et al. Geostatistical models of soil organic carbon density prediction based on soil hyperspectral reflectance [J]. *Transactions of the CSAE*, 2017, 33(2):183-191 (in Chinese with English abstract).
- [19] 颜祥照,姚艳敏,张霄羽.土壤有机质遥感制图研究进展与展望[J].中国农业信息,2019,31(3):13-26. YAN X Z, YAO Y M, ZHANG X Y. The progress and prospect of soil organic matter mapping based on remote sensing technology [J]. *China agricultural information*, 2019, 31(3):13-26 (in Chinese).
- [20] 宋玉婷,高小红,宋奇,等.基于Sentinel-2 MSI与Sentinel-1 SAR影像相结合的区域农田土壤有机碳估算:以青海湟水流域黄土丘陵区为例[J].土壤通报,2025,56(3):755-766.

- SONG Y T, GAO X H, SONG Q, et al. The combination of sentinel-2 MSI and sentinel-1SAR to estimate and map soil organic carbon content in farmland: a case study of the typical area of loess hilly area in the Huangshui watershed of Qinghai Province[J]. Chinese journal of soil science, 2025, 56(3): 755-766(in Chinese with English abstract).
- [21] 王迪, 曾妍, 田甜, 等. 基于卫星遥感的耕地质量监测与评价研究进展[J]. 中国农业资源与区划, 2025, 46(12): 203-218. WANG D, ZENG Y, TIAN T, et al. Research progress on cultivated land quality monitoring and assessment by remote sensing[J]. Chinese journal of agricultural resources and regional planning, 2025, 46(12): 203-218(in Chinese with English abstract).
- [22] 李泽鑫, 高爽, 胡登辉, 等. 天基载荷特性光谱分辨率与信噪比对黑土地有机质定量反演影响分析[J]. 遥感技术与应用, 2024, 39(4): 917-926. LI Z X, GAO S, HU D H, et al. Sensitivity of black soil organic matter content prediction to the spectral resolution and signal-to-noise ratio of space-based remote sensing loads[J]. Remote sensing technology and application, 2024, 39(4): 917-926(in Chinese with English abstract).
- [23] 王倩, 毕研盟, 杨忠东. 气溶胶对大气CO₂短波红外遥感探测影响的模拟分析[J]. 物理学报, 2018, 67(3): 301-313. WANG Q, BI Y M, YANG Z D. Simulation analysis of aerosol effect on shortwave infrared remote sensing detection of atmospheric CO₂[J]. Acta physica sinica, 2018, 67(3): 301-313(in Chinese with English abstract).
- [24] CORTES C, VAPNIK V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [25] BREIMAN L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.
- [26] 王雨雪, 杨柯, 高秉博, 等. 基于两点机器学习方法的土壤有机质空间分布预测[J]. 农业工程学报, 2022, 38(12): 65-73. WANG Y X, YANG K, GAO B B, et al. Prediction of the spatial distribution of soil organic matter based on two-point machine learning method[J]. Transactions of the CSAE, 2022, 38(12): 65-73(in Chinese with English abstract).
- [27] 袁玉琦, 陈瀚闾, 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. YUAN Y Q, CHEN H Y, ZHANG L M, et al. 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(in Chinese with English abstract).
- [28] 聂明鹏. 基于排序差异和算法对光谱变量选择与定性分析的研究[D]. 温州: 温州大学, 2019. NIE M P. Study on the variable selection and qualitative analysis of spectrum based on sum of ranking differences algorithm[D]. Wenzhou: Wenzhou University, 2019(in Chinese with English abstract).
- [29] HONG Y S, CHEN Y Y, CHEN S C, et al. Bridging the gap between laboratory VNIR-SWIR spectra and Landsat-8 bare soil composite image for soil organic carbon prediction [J/OL]. Remote sensing of environment, 2025, 328: 114874 [2025-07-16]. <https://doi.org/10.1016/j.rse.2025.114874>.
- [30] 施建成. MODIS亚像元积雪覆盖反演算法研究: 纪念杰出的地理学家、冰川学家施雅风先生逝世一周年[J]. 第四纪研究, 2012, 32(1): 6-15. SHI J C. An automatic algorithm on estimating sub-pixel snow cover from modis[J]. Quaternary sciences, 2012, 32(1): 6-15(in Chinese with English abstract).
- [31] 祝元丽, 王冬艳, 张鹤, 等. 采用无人机载高分辨率光谱仪反演土壤有机碳含量[J]. 农业工程学报, 2021, 37(6): 66-72. ZHU Y L, WANG D Y, ZHANG H, et al. Soil organic carbon content retrieved by UAV-borne high resolution spectrometer [J]. Transactions of the CSAE, 2021, 37(6): 66-72(in Chinese with English abstract).
- [32] 尹成卓, 高小红, 宋奇, 等. 基于Sentinel-2影像的农田土壤有机碳含量估算: 以湟水流域典型区域为例[J]. 激光与光电子学进展, 2025, 62(12): 294-305. YIN C Z, GAO X H, SONG Q, et al. Estimation of soil organic carbon content in farmland based on Sentinel-2 imagery: case study of a typical area in the Huangshui River Basin[J]. Laser & optoelectronics progress, 2025, 62(12): 294-305(in Chinese with English abstract).
- [33] 张益玮. 基于SAR与光学遥感影像的玉米秸秆覆盖度反研究[D]. 鞍山: 辽宁科技大学, 2023. ZHANG Y W. Inverse study on corn stalk coverage based on SAR and optical remote sensing images[D]. Anshan: University of Science and Technology Liaoning, 2023(in Chinese with English abstract).
- [34] POULADI N, GHOLIZADEH A, KHOSRAVI V, et al. Digital mapping of soil organic carbon using remote sensing data: a systematic review [J/OL]. CATENA, 2023, 232: 107409 [2025-07-16]. <https://doi.org/10.1016/j.catena.2023.107409>.
- [35] BROEG T, DON A, WIESMEIER M, et al. Spatiotemporal monitoring of cropland soil organic carbon changes from space [J/OL]. Global change biology, 2024, 30(12): e17608 [2025-07-16]. <https://doi.org/10.1111%2Fgcb.17608>.
- [36] LI Y J, LUO L Y, GUO F, et al. Quality control and improvement of GNSS-IR soil moisture robust inversion model [J]. Advances in space research, 2025, 75(1): 91-107.

Multi-source spectral data and machine learning algorithm-based inversion of organic carbon in soil : taking northeastern Inner Mongolia Autonomous Region , China as an example

REN Guanyao¹, QIN Jing¹, QIAN Xiaojiang², ZHANG Ke³, YAN Weiyue¹, FENG Zi'ang⁴, GUO Long⁴

1.Inner Mongolia Autonomous Region Geomatics Center, Hohhot 010010, China;

2.Hubei Provincial Institute of Land Surveying and Mapping, Wuhan 430014, China;

3.GBI (Beijing) Testing & Certification Co., Ltd., Beijing 100010, China;

4.College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China

Abstract Northeastern Inner Mongolia Autonomous Region, China was used to study the mechanism of external environmental factors including illumination and atmospheric conditions and sensor parameters including spectral and spatial resolutions affecting the accuracy of inverting the organic carbon in soil (SOC) from different sources of spectral data. SOC samples from 160 surface soils (0-20 cm) were collected and proximal hyperspectral data acquired under indoor artificial lighting and outdoor natural sunlight and spaceborne multispectral (Landsat-8 and Sentinel-2) and hyperspectral (ZY1-02D) data were synchronously obtained. Random forest (RF) and support vector machine (SVM) algorithms were used to construct the inversion models of SOC. The effects of environmental factors and sensor parameters on the accuracy of inverting SOC were analyzed by systematically comparing the performance of model across different sources of data. The results showed that proximal data acquired under indoor artificial lighting sources had slightly better accuracy of inversion than that under outdoor sunlight sources due to stable and controllable spectral signals. However, the difference between the two was small, indicating that the effect of natural illumination fluctuations on the inversion of SOC is limited. The inversion accuracy of proximal hyperspectral data was significantly higher than that of spaceborne multispectral and hyperspectral data, mainly due to the interference from atmospheric scattering, water vapor absorption, and mixed pixels in satellite data. Among the spaceborne datasets, the inversion accuracy of hyperspectral satellite ZY1-02D was higher than that of multispectral sensors, while the improvement in spatial resolution of Sentinel-2 had limited effect on the performance of model compared to that of Landsat-8 (10 m vs 30 m), indicating that spectral resolution contributes more to the accuracy of inverting SOC than spatial resolution.

Keywords multi-source spectral data; organic carbon in soil (SOC); machine learning algorithms; methods of inversion; ZY1-02D; spectral resolution

(责任编辑:陆文昌)