

## 基于无人机高光谱影像的黑土区玉米农田土壤有机质估算

夏晨真<sup>1, 2, 3</sup>, 姜艳艳<sup>4</sup>, 张星宇<sup>1, 2, 3</sup>, 沙野<sup>5</sup>, 崔帅<sup>1, 2, 3</sup>,  
米国华<sup>5</sup>, 高强<sup>1, 2, 3</sup>, 张月<sup>1, 2, 3\*</sup>

1. 吉林农业大学资源与环境学院, 吉林 长春 130118
2. 吉林省商品粮基地土壤资源可持续利用重点实验室, 吉林 长春 130118
3. 秸秆综合利用与黑土地保护教育部重点实验室, 吉林 长春 130118
4. 松辽水利委员会松辽流域水土保持监测中心站, 吉林 长春 130021
5. 中国农业大学资源与环境学院, 北京 100083

**摘要** 土壤有机质(SOM)作为土壤的重要组成部分,其含量高能够反映土壤的肥力和质量状况。相较于传统 SOM 的测定方法,利用无人机高光谱影像可快速、精准获取田块尺度的 SOM 含量。为探究基于高光谱数据建立的线性和非线性模型对作物覆盖下土壤有机质估算精度的差异,以东北黑土区的玉米试验田为研究区,分别采集了拔节期和吐丝期的土壤样本及同时期无人机高光谱影像作为数据源,分析作物覆盖条件下土壤光谱反射率与有机质含量的相关关系,并根据其响应波段构建光谱指数。以施肥量和光谱指数作为自变量,通过特征变量的筛选分别建立多元逐步线性回归模型(SMLR)、支持向量机(SVM)、随机森林(RF)和 eXtreme gradient boosting(XGBoost)模型,并验证比较各模型的精度(选用  $R^2$  和 RMSE 为评价指标)。结果表明,作物覆盖条件下土壤有机质含量的响应波段为 450~640 nm。多年长期施用化肥对 SOM 含量有着显著影响,将其作为协变量引入模型明显提高了对 SOM 的估算精度。4 种模型检验精度的对比结果为: XGBoost>RF>SMLR>SVM,其中以拔节期 XGBoost 的估算结果最好(建模集和验证集的  $R^2$ 、RMSE 分别为 0.516、0.253 和 0.590、0.222)。可以利用无人机高光谱技术快速估算田块尺度玉米农田的土壤有机质含量,且 XGBoost 模型是估算作物覆盖条件下土壤有机质含量的较优选择。

**关键词** 无人机; 高光谱; 土壤有机质; 多元逐步线性回归; 机器学习

中图分类号: TP79 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2023)08-2617-10

### 引言

东北黑土区作为我国重要的粮食生产优势区及商品粮供给地,玉米作为其主要作物之一,每年的产量占全国总产量的 30%以上<sup>[1]</sup>。长期以来对黑土的过度开发利用导致土地退化严重,土壤有机质(soil organic matter, SOM)含量作为土壤的重要组成部分,是反映土壤的肥力和质量的主要标志之一,对作物的生长发育有着显著影响<sup>[2]</sup>。因此,快速、精确地监测黑土区玉米关键生育期的 SOM 含量及其空间差异,在农田精准施肥、玉米长势监测、东北黑土地保护及可持续利用等方面具有重要意义。

东北黑土区长期被作物和积雪所覆盖,为快速、精确地监测黑土区的 SOM 带来了挑战,已有学者针对东北黑土区 SOM 的遥感估算开展了大量研究<sup>[3-5]</sup>。目前 SOM 的估算方法主要有传统地面调查法,基于土壤光谱响应特征的遥感估算算法和数字土壤制图法<sup>[6-9]</sup>。无人机遥感兼具时效快、低成本、高空间分辨率、不受云层影响等优点,在中小尺度的 SOM 估算方面受到了部分研究者的青睐<sup>[3, 10]</sup>。Al-Abbas 等<sup>[11]</sup>发现 SOM 与可见光及近红外波段之间的响应关系密切; Krishnan 等<sup>[12]</sup>发现用可见光区域估算 SOM 含量要优于近红外区域;刘焕军等<sup>[3]</sup>发现黑土 SOM 响应波段则在 415~1380 nm。陶培峰<sup>[13]</sup>、刘恬琳<sup>[14]</sup>和 Sami<sup>[15]</sup>的研究表明高光谱传感器获得的影像相较于多光谱数据含有更丰富的光谱信

收稿日期: 2022-04-06, 修订日期: 2022-07-21

基金项目: 国家重点研发计划项目(2021YFD1500800), 国家自然科学基金项目(U19A2061), 中国科学院战略性先导科技专项课题(XDA28080500)和国家地球系统科学数据中心资助

作者简介: 夏晨真, 1997 年生, 吉林农业大学资源与环境学院硕士研究生 e-mail: Summul\_R@163.com

\* 通讯作者 e-mail: lisa\_ling7892002@163.com

息,在 SOM 估算方面具有切实可行性。然而, SOM 估算的研究多见于裸土条件下,应用无人机高光谱影像、在作物覆盖条件下的 SOM 估算研究还鲜有报道。目前已有学者通过从遥感图像中提取的光谱指数,建立光谱指数与 SOM 的关系从而进行 SOM 估算,但针对植被覆盖下的 SOM 估算仍存在许多困难<sup>[6, 15]</sup>。王曦等<sup>[17]</sup>研究发现基于植被光谱数据进行建模比基于裸土建立的小麦拔节期 SOM 估算模型精度更高。Guo 等<sup>[8]</sup>通过不同时期的遥感影像,以光谱波段和从影像中提取的植被指数作为预测因子反映油菜的生长情况,并克服地表植被的影响,成功在田间尺度预测了土壤有机碳含量。若可实现通过覆盖作物间接估算 SOM 含量,将为 SOM 的估算提供新的方法与途径。数字土壤制图法可通过环境变量来进行土壤属性的估算,使得土壤的空间变化可被更详细的表达<sup>[10]</sup>。现代盛行的数字土壤制图主要建立在土壤学、地理学和数学三种理论基础之上,其中以土壤学为理论依据的应用尤为广泛<sup>[18]</sup>。土壤发生理论中 Jenny<sup>[19]</sup>提出的 clorpt 函数式认为:土壤是气候(c 或 cl, climate)、生物(o, organisms)、地形(r, topography)、母质(p, parent material)和时间(a 或 t, age)综合作用的产物。部分学者也发现土壤的理化性状、地形、气候和人为因素等是影响中国农田有机碳密度空间分异的重要因素,并结合这些因素完成了不同尺度下 SOM 含量的估算<sup>[20-23]</sup>。然而,在一些平原或尺度较小的平缓地区,土壤制图是一个难题,其气候差异不明显,地形、坡度、母质等在大尺度区域的影响因素可能并不适用<sup>[25]</sup>。而在大多数研究中,遥感影像通常是在一个或几个特定时间内采集的一幅或几幅图像,地表植被覆盖使得土壤反射率的获取变得更加困难,干扰辅助变量的准确收集,影响了制图的准确性<sup>[26]</sup>。在这种情况下,获取对其有价值的环境变量尤为重要,可能需要选取一个或多个有价值的变量用于平原或尺度较小的平缓地区的 SOM 制图。人类活动因子对于土壤空间变化的影响越来越显著,逐渐受到研究者的关注,同时,通过遥感数据提取环境变量以辅助估算低起伏地区土壤属性变化的应用也逐渐广泛<sup>[15, 24]</sup>。而农田 SOM 的空间异质性不但受到几种成土因素的影响,还受到耕作、施肥、灌溉等管理方式的影响<sup>[27]</sup>。长期施用化肥对土壤有机质含量有着显著的影响<sup>[28-29]</sup>,利用施肥量作为自变量进行小尺度 SOM 含量估算的研究却鲜有报道。

在对 SOM 进行模型估算时,建模方法不同,其稳定性也不同。考虑到作物状况和 SOM 之间的非线性关系,因此传统回归模型如多元逐步线性回归(stepwise multiple linear regression, SMLR)和偏最小二乘回归(partial least squares regression, PLSR)等可能并不是估算 SOM 的最佳选择<sup>[30]</sup>。随着大数据发掘和人工智能的发展,已经开发了一系列机器学习算法来解决非线性关系和多种类型的数据<sup>[30-31]</sup>。例如,支持向量机(support vector machine, SVM)和人工神经网络的反向传播技术(back propagation neural network, BPNN)已被用作监督学习模型,用于从大量高光谱数据或各种环境变量中估算土壤性质<sup>[32-33]</sup>。Xu<sup>[10]</sup>、刘恬琳<sup>[14]</sup>和 Gu<sup>[34]</sup>等基于高光谱遥感影像的有机质估算模型研究表明,大多数情况下机器学习模型的估算效果优于统计分析模型,尤其以 RF

的估算精度较优。葛翔宇<sup>[35]</sup>研究发现 XGBoost (eXtreme Gradient Boosting)在土壤含水量的估算方面效果超群,要优于 RF 的模型,在 SOM 含量的估算方面却少有研究,其表现也尚未可知。

基于 SOM 的估算研究多见于裸土条件下,将施肥量作为协变量进行低起伏地区 SOM 含量估算的研究鲜有报道。因此,本研究以东北黑土区的试验田块为研究区,在田块尺度上进行土壤样本采集、理化性质分析、光谱测量和处理,以光谱指数和施肥量为协变量,采用 SMLR、SVM、XGBoost 和 RF 建立作物覆盖下 SOM 的高光谱估算模型,并确定最佳估算模型。旨在探究利用高光谱技术建立田块尺度东北黑土区玉米农田 SOM 估算模型的可行性,以期对东北黑土地耕地质量的提升与可持续利用提供理论依据与科学参考。

## 1 实验部分

### 1.1 研究区概况

研究区位于吉林省梨树县境内(图 1)四棵树乡的三棵村,地理位置  $43^{\circ}20'17.4''$  N,  $124^{\circ}0'29.1''$  E,属北温带半湿润大陆性季风气候,土壤类型主要为黑土(黏化湿润均腐土),四季分明。年均气温  $6.5^{\circ}\text{C}$ ,年日照时数  $2541\text{ h}$ ,作物生长期(5月—9月)内为  $1192\text{ h}$ 。年积温  $>0^{\circ}\text{C}$  积温  $3244^{\circ}\text{C}$ ,  $>10^{\circ}\text{C}$  积温  $3030^{\circ}\text{C}$ ;无霜期  $155\text{ d}$ <sup>[36]</sup>。

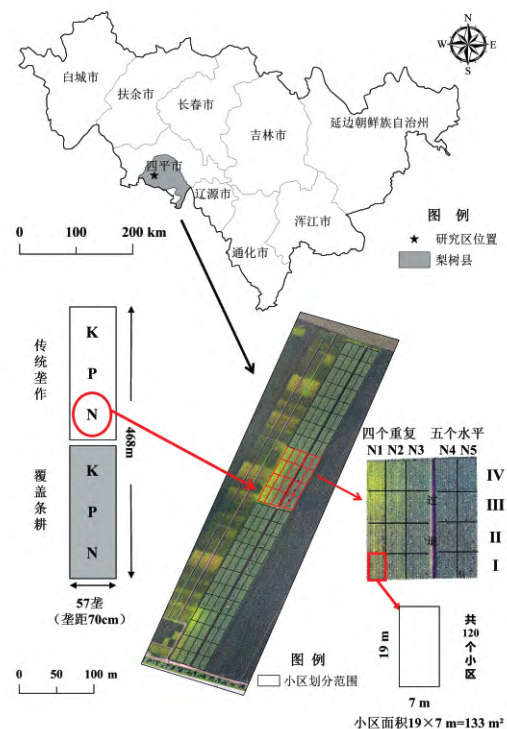


图 1 研究区的地理位置

Fig 1 Location of the study area

如图 1 所示,试验田为已经进行多年的长期定位施肥试验,设置为玉米连作区的单因子裂区试验。裂区分别设置为

传统耕作和条耕秸秆覆盖两种耕作方式, 每种耕作方式下包含3种施肥处理, 分别为氮梯度、磷梯度和钾梯度, 每种处理设置4次重复。氮梯度为尿素(46% N)控制, 其中磷肥和钾肥施用量分别为75和90 kg·hm<sup>-2</sup>, 氮肥施用设置5个水平: N1(0 kg·hm<sup>-2</sup>)、N2(60 kg·hm<sup>-2</sup>)、N3(120 kg·hm<sup>-2</sup>)、N4(180 kg·hm<sup>-2</sup>)、N5(240 kg·hm<sup>-2</sup>); 磷肥为过磷酸钙(18% P<sub>2</sub>O<sub>5</sub>)控制, 其中氮肥和钾肥施用量分别为180和90 kg·hm<sup>-2</sup>, 磷肥施用设置5个水平: P1(0 kg·hm<sup>-2</sup>)、P2(25 kg·hm<sup>-2</sup>)、P3(50 kg·hm<sup>-2</sup>)、P4(75 kg·hm<sup>-2</sup>)、P5(100 kg·hm<sup>-2</sup>); 钾肥为氯化钾(50% K<sub>2</sub>O)控制, 其中氮肥和磷肥施用量分别为180和75 kg·hm<sup>-2</sup>, 钾肥施用设置5个水平: K1(0 kg·hm<sup>-2</sup>)、K2(30 kg·hm<sup>-2</sup>)、K3(60 kg·hm<sup>-2</sup>)、K4(90 kg·hm<sup>-2</sup>)、K5(120 kg·hm<sup>-2</sup>)。共计120个小区, 小区面积为133 m<sup>2</sup>, 行距设置为70 cm, 种植密度62 000株·hm<sup>-2</sup>。

## 1.2 土壤数据的采集和分析

根据玉米的生长状况及天气情况, 采样时间分别为2019年玉米的拔节期(6月23日)和吐丝期(8月1日)。根据五点采样法, 在每个小区进行样品采集, 每个生育期采集120个样本, 两生育期共计采集240个样本, 样本采集的同时利用GPS记录每个样本的坐标, 以便之后在高光谱数据中获取与采样点对应的植被冠层光谱反射率值( $R$ )。然后在室内进行土壤样品的称量后研磨、风干, 过2 mm筛, 用重铬酸钾容量法<sup>[37]</sup>测定SOM含量。

为增加建模结果的代表性, 分别进行了单生育期和两生育期的数据分析及建模。将不同数据集按照7:3的比例进行5次随机划分, 其中70%作为估算模型的建模集(单生育期为84个样本, 两生育期为168个样本), 其余30%则作为验证集(单生育期为36个样本, 两生育期为72个样本)。然后求取5次建模结果的平均值, 使其结果更可靠。通过SPSS Statistics 22软件对各生育期采样点的SOM含量进行统计分析, 结果见表1。

表1 土壤样本有机质含量的统计特征

Table 1 Descriptive statistics of soil organic matter content

生育期	数据集	最大值/%	最小值/%	均值/%	标准差/%	变异系数	峰度	偏度
拔节期	总集	4.086 1	2.336 4	3.119 3	0.347 7	0.121 0	-0.108 0	0.419 0
	建模集	4.013 9	2.336 4	3.114 8	0.347 7	0.121 0	-0.148 0	0.378 0
	验证集	4.086 1	2.503 7	3.129 7	0.352 3	0.124 0	0.139 0	0.535 0
吐丝期	总集	4.680 7	2.316 5	3.246 8	0.447 9	0.201 0	0.912 0	0.552 0
	建模集	4.680 7	2.316 5	3.254 6	0.466 7	0.218 0	1.053 0	0.628 0
	验证集	4.256 1	2.435 1	3.231 5	0.412 2	0.170 0	0.093 0	0.216 0
两生育期	总集	4.680 7	2.316 5	3.183 1	0.405 2	0.164 0	0.937 0	0.614 0
	建模集	4.548 8	2.316 5	3.177 6	0.401 2	0.161 0	0.736 0	0.534 0
	验证集	4.680 7	2.364 9	3.196 4	0.419 5	0.176 0	1.393 0	0.777 0

## 1.3 无人机高光谱影像的采集

选择晴朗无云, 光照充足的无风或微风天气, 于北京时间10:00—15:00进行高光谱影像的采集(太阳高度角大于45°)。利用中国深圳市大疆创新科技有限公司生产的大疆Matrice 600 Pro六旋翼无人机进行影像采集, 无人机飞行载重6 kg, 无风状态最大水平飞行速度65 km·h<sup>-1</sup>, 搭载H185高光谱相机得到研究区影像。该相机光谱范围为450~950 nm, 带宽为4 nm, 通道数126个。起飞前对无人机进行航线设定和白板校正, 设定航高为120 m, 空间分辨率为5 cm, 航向重叠度80%, 旁向重叠度60%, 光谱仪镜头选择聚焦25 mm, 对应的视场角约为13°。影像拼接利用Pix4D Mapper (Pix4D SA, 瑞士)软件完成, 随后进行辐射定标, 辐射定标采用伪标准地物辐射校正法[式(1)], 利用1.5 m×1.5 m的白板作为参考, 将影像中各像元的亮度值(DN)转化为反射率值。通过ENVI 5.3软件(Exelis Visual Information Solutions公司, 美国)的FLAASH模块对高光谱影像进行大气校正。通过预先布设的控制点对影像进行几何校正和地理配准, 几何校正以无人机RGB正射影像为参考, 在影像上均匀选取若干个参考点, 并以此为依据进行几何校正, 确保影像误差不超过0.5个像元, 最终得到影像中各像元的光谱反射率值。

$$R_{\text{目标}} = \frac{DN_{\text{目标}}}{DN_{\text{参考板}}} \times R_{\text{参考板}} \quad (1)$$

式(1)中,  $R_{\text{目标}}$ 为目标地物的反射率,  $DN_{\text{目标}}$ 为目标地物的DN均值,  $DN_{\text{参考板}}$ 为白色参考板的DN均值,  $R_{\text{参考板}}$ 为白色参考板的反射率值。

## 1.4 基于高光谱影像的光谱指数选取

高光谱影像数据的波段数较多, 其变量之间存在较为严重的多重共线性。而由于作物的覆盖使得我们无法直接获取土壤的光谱反射率。但土壤条件的差异会造成植被长势的不同, 最终体现在植被冠层的光谱特征上。因此利用作物冠层的光谱指数以监测作物在不同生育期的生长特征。选取了增强植被指数(enhanced vegetation index, EVI)、可见光耐大气指数(visible atmospherically resistant index, VARI)和近红外/近红外(near infrared/near infrared, NIR/NIR)等20种植被指数, 其公式如表2所示。

## 1.5 SOM估算模型的构建与验证

为了比较不同建模方法在植被覆盖条件下估算田块尺度土壤有机质精度的差异, 根据划分的建模集和验证集, 以SOM为因变量, 基于高光谱数据得到的光谱指数和施肥量作为自变量, 构建了各生育期的SOM预测模型。分别采用SMLR、RF、SVM和XGBoost四种建模方法构建SOM含量

表 2 本研究所用的 20 种光谱指数  
Table 2 The 20 spectral indices used in this study

光谱变量	公式	来源
EVI (enhanced vegetation index)	$2.5((R_{798}-R_{670})/(R_{798}-(6R_{670})-(7.5R_{474})+1))$	[38]
VARI (visible atmospherically resistant index)	$(R_{530}-R_{674})/(R_{530}+R_{674}-R_{450})$	[39]
NIR/NIR (near infrared/near infrared)	$R_{778}/R_{738}$	[40]
ND705 (normalized difference vegetation index 705)	$(R_{750}-R_{705})/(R_{750}+R_{705})$	[41]
mNDblue (modified normalized difference)	$(R_{450}-R_{730})/(R_{450}+R_{850})$	[42]
OSAVI2 (optimized soil-adjusted vegetation index 2)	$(1+0.16)(R_{750}-R_{705})/(R_{750}+R_{705}+0.16)$	[43]
Maccioni	$(R_{780}-R_{710})/(R_{780}-R_{680})$	[44]
MTCI (MERIS terrestrial chlorophyll index)	$(R_{750}-R_{730})/(R_{730}+R_{675})$	[45]
MSR (modified simple ratio)	$(R_{800}-R_{670}-1)/((R_{800}+R_{670})\times0.5+1)$	[46]
DCNI (double-peak canopy nitrogen index)	$(R_{720}-R_{700})/(R_{700}-R_{670})/(R_{720}-R_{670}+0.03)$	[47]
RTVI (red-edge triangular vegetation index)	$(100\times(R_{750}-R_{730})-10\times(R_{750}-R_{550})\times\sqrt{R_{700}/R_{670}})$	[47]
GNDSI (green normalized difference vegetation index)	$(R_{800}-R_{550})/(R_{800}+R_{550})$	[48]
CI2 (red-edge chlorophyll index 2)	$R_{740}/R_{550}-1$	[48]
CIgreen (green chlorophyll index)	$R_{850}/R_{530}-1$	[49]
MCARI (modified chlorophyll absorption reflectance index)	$(R_{700}-R_{600}-0.2\times(R_{700}-R_{550}))\times(R_{700}/R_{670})$	[49]
MCARI/OSAVI RE (red edge-based modified chlorophyll absorption in reflectance index/optimized soil-adjusted vegetation index)	$((R_{750}-R_{705})-0.2\times(R_{750}-R_{705}))(R_{750}/R_{705})/((1+0.16)\times(R_{750}-R_{705})/(R_{750}+R_{705}+0.16))$	[50]
MCARI2/OSAVI2	$((R_{700}-R_{670})-0.2\times(R_{700}-R_{550}))(R_{700}/R_{670})/((1+0.16)\times(R_{750}-R_{705})/(R_{750}+R_{705}+0.16))$	[51]
MCARI/OSAVI	$((R_{700}-R_{600})-0.2\times(R_{700}-R_{550}))(R_{700}/R_{670})/((1+0.16)\times(R_{800}-R_{670})/(R_{800}+R_{670}+0.16))$	[44]
TCARI/OSAVI	$3\times(((R_{700}-R_{670})-0.2\times(R_{700}-R_{550}))(R_{700}/R_{670}))/((1.16\times(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)))$	[52]
MTVI2 (modified triangular vegetation index 2)	$1.5\times(1.2\times(R_{800}-R_{500})-2.5\times(R_{670}-R_{550}))/\sqrt{((2R_{800}+1)^2-(6R_{800}-5)\sqrt{R_{670}+0.5})}$	[53]

的估算模型。建模前，利用皮尔逊(Pearson)相关分析方法和重要性评分筛选特征变量。Pearson 相关分析通过 SPSS Statistics 22 软件实现，重要性评分通过 R 软件实现，其变量的相对重要性以 %IncMSE(increased in mean squared error, %)和残差平方和(increase in node purity, IncNodePurity)来度量[式(2)]。

$$w_i = 0.5A_i / \sum_{i=1}^n A_i + 0.5B_i / \sum_{i=1}^n B_i \quad (2)$$

式(2)中： $w_i$  表示第  $i$  个变量的相对重要性； $A_i$  和  $B_i$  分别表示第  $i$  个变量的 %IncMSE 和 IncNodePurity； $n$  表示变量的总数。%IncMSE 为增长的错误率平方均值，可理解为对目标变量预测准确的贡献度；IncNodePurity 则代表了每个变量对分类树每个节点上观测值异质性的影响，从而比较变量的重要性，该值越大表示该变量的重要性越大。

(1)多元逐步线性回归模型

SMLR 的基本思想是按变量的方差贡献度，从 21 种协变量中逐一挑选重要变量并引入回归模型，每次引入后都会对未选入的所有变量重新进行预测误差判断，使得先前剔除的变量在新变量引入后对模型在具有贡献时又被重新选入。直至筛选出最优参数、建立最优反演模型为止，回归方程中始终只保留重要的变量<sup>[26]</sup>。

(2)随机森林模型

RF 是将 Bagging 与决策树算法进行结合所得到的集成

学习算法<sup>[14-15, 34]</sup>。该算法可以有效避免模型训练时的过拟合现象，同时有效抑制噪声的负面影响，有利于地物高光谱特征的综合利用<sup>[34]</sup>。为防止过度拟合，初始的 ntree (树的数量) 设置为 500，mtry (每个拆分中可供选择的变量数量) 则设置为所有自变量的个数。

(3)支持向量机模型

SVM 是一种二分类模型的机器学习方法，它在高维或无限维空间中构造一个或一组超平面对样本进行分割，可用于分类或回归。通过不同类型的核函数(如线性、径向、Sigmoid 和多项式)实现超平面之间的良好分离<sup>[11, 15]</sup>。本研究选择了线性核函数，基于带宽代价参数和不敏感损失函数对支持向量机模型进行了调整。

(4)XGBoost 模型

XGBoost 是在梯度下降树 (gradient boosting decision tree, GBDT) 的基础上对 boosting 算法进行改进的树结构增强模型<sup>[35]</sup>。其基本思想是：首先构建多个 CART(classification and regression trees)模型对数据集进行预测，随后将这些树集成为一个新的树模型，通过不断地迭代提升，每次迭代生成的新树模型都会拟合前一棵树的残差，直到达到最佳训练效果<sup>[54]</sup>。其常用的关键参数为 booster、eta、max—depth、objective 和 nround，其中 booster 和 objective 均采用默认值。

最后，选用决定系数 (determination coefficients,  $R^2$ )、

均方根误差 (root mean squared error, RMSE) 和归一化均方根误差 (normalized root mean squared error, nRMSE) 3 个参数进行上述 4 种估算模型的精度评价。其中,  $R^2$  用来评价模型的拟合优度, 取值范围在 0~1 之间, 越接近 1, 说明拟合程度越好。RMSE 和 nRMSE 用来衡量 SOM 预测值与观测值之间的偏差, 其值越小, 说明预测效果越好。其公式如式(3)、式(4)和式(5)所示

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}} \quad (4)$$

$$\text{nRMSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2 / n}{\bar{X}}} \times 100\% \quad (5)$$

式中:  $X_i$  和  $Y_i$  分别为 SOM 实测值和 SOM 预测值;  $\bar{X}$  为 SOM 实测值的均值;  $\bar{Y}$  为 SOM 预测值的均值;  $n$  为样本数。

## 2 结果与讨论

### 2.1 SOM 与高光谱数据的相关分析

为探究作物覆盖条件下 SOM 含量与高光谱影像数据的相关性, 分别提取了拔节期、吐丝期各采样点在各个波段的反射率, 并进行了光谱数据与 SOM 含量的 Pearson 相关分析(图 2)。发现在波长 450~640 nm 处, 拔节期和吐丝期的 SOM 含量与高光谱反射率具有显著的相关性, 并随着波长的增加其相关性逐渐减弱。在 450~530 nm 处, 拔节期 SOM 含量与高光谱反射率的相关性较高, 470 nm 处相关系数最高, 为 0.237。而吐丝期的 SOM 含量则在 560~610 nm 处与高光谱反射率的相关性较高, 在 610 nm 处相关系数达到峰值, 为 0.244。

### 2.2 SOM 与协变量的相关分析

为构建 SOM 含量的估算模型, 根据土壤有机质在作物覆盖条件下的敏感波段以及 SOM 含量的光谱特性, 选取了

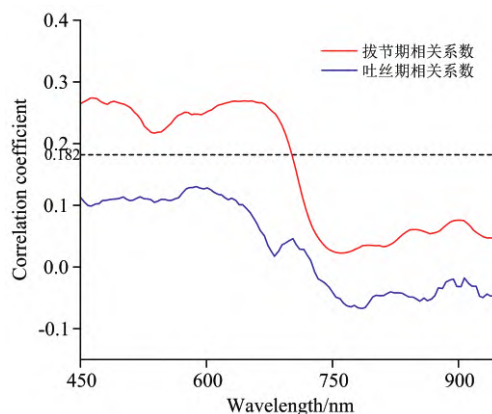


图 2 SOM 与各波段光谱反射率的相关系数

Fig 2 Correlation coefficient between SOM and spectral reflectance

表 2 中的 20 种光谱指数。长期施用化肥对土壤有机质含量有着显著影响<sup>[28-29]</sup>, 所以也将施肥量作为协变量参与估算模型的构建。由于施肥量与光谱指数的度量单位并不相同, 因此分别对各生育期的光谱指数及施肥量进行归一化处理, 并与对应的 SOM 含量进行相关分析, 针对每个生育期的相关分析结果如图 3 所示。

从图 3 可以看出, SOM 含量与施肥量呈现显著的相关关系, 然而光谱指数与 SOM 含量的相关性却并不显著, 这可能是由于 Pearson 系数只能表征一元变量的线性相关程度, 而对多元变量间的线性或非线性相关关系却很难通过 Pearson 系数反映。如图 3(a), 拔节期光谱指数与 SOM 含量多为正相关, 而图 3(b)则显示吐丝期的光谱指数与 SOM 含量则多为负相关。在拔节期, 光谱指数中 MCARI/OSAVI 与 SOM 含量的相关系数最大, MTVI2 与 SOM 含量的相关系数最小; 在吐丝期, 光谱指数中 OSAVI 与 SOM 含量的相关系数最大, VARI 与 SOM 含量的相关系数最小。由图 3(c)中可以看出, 将两个生育期一起进行相关分析, ND705 与 SOM 含量的相关系数最大, 其相关性相较于单生育显著增加, 且随着时间的推移呈增加趋势。各指数之间也具有很强的相关性, 因此后续需要进行特征变量的筛选以防其共线性过大。

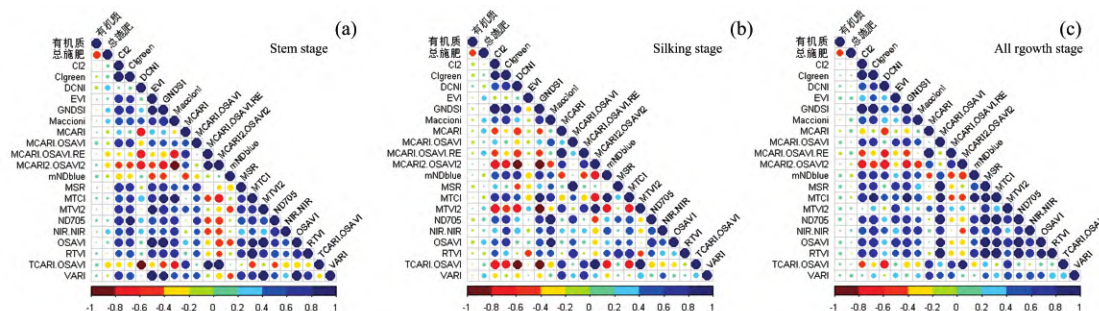


图 3 SOM 与各协变量的相关系数

Fig 3 Correlation matrix heatmaps for SOM with covariates

### 2.3 SOM 估算模型的变量筛选

利用 SMLR、RF、SVM 和 XGBoost 四种建模方法进行

SOM 含量估算模型的构建。在这四种建模方法中, 一类是线性模型如 SMLR, 另一类是以非线性模型为核心的数据挖掘



技术,包括 RF、SVM 和 XGBoost。建模时,将 SOM 作为估算模型的因变量,施肥量和 20 种光谱指数作为模型的自变量。针对拔节期的建模集进行 SMLR 的模型构建时,模型首先选入了施肥量作为最重要的变量,然后将施肥量引入后,再引入 NIR/NIR 作为变量会使模型的精度提高,故最终引入这两个变量参与模型构建。针对吐丝期的建模集进行 SMLR 的模型构建时,并没有变量被引入,因此没有建模结果。而将两生育期数据集组合进行 SMLR 的模型构建时,模型引入了施肥量和 EVI 两个变量。

进行非线性建模时,发现仅用 Pearson 系数无法筛选出用于建模的变量,因此通过 5 次 10 折交叉验证的最高准确性来确定不同生育期模型的最优变量个数,并采用相对重要性评分进行变量的筛选(图 4)。最终在拔节期筛选出了总施肥和 MCARI/OSAVI2 共计两个特征变量,其中总施肥的重要性达到 27.84%,而 MCARI/OSAVI2 的重要性为 7.57%;在吐丝期筛选出了总施肥和 Cigreen 共计两个特征变量,其中总施肥的重要性达到 17.85%,而 Cigreen 的重要性为 7.99%;两生育期组合则筛选出了总施肥、TCARI/OSAVI、

MCARI2/OSAVI2 和 MCARI/OSAVI RE 共计 4 个特征变量,其相对重要性分别为 25.24%、5.87%、5.09%和 4.75%。

2.4 SOM 估算模型的构建和验证

将上述特征变量作为估算模型的自变量,分别建立各生育期的 SMLR、RF、SVM 和 XGBoost 回归模型。其建模集的拟合结果如表 3 所示。结果表明利用线性核函数的 SVM 模型相较于 SMLR 的估算结果较差,且在吐丝期 SVM 模型的  $R^2$  较低, RMSE 则较高,无法对样本进行估算。RF 和 XGBoost 的  $R^2$  和 RMSE 值相较于 SMLR 和 SVM 较好,其对样本估算的结果也更好。而相较于其他 3 种建模方式, XGBoost 不论在单生育期还是两生育期,均表现出更优秀的性能,其估算精度也更稳定。其中 XGBoost 在拔节期的估算结果最优( $R^2=0.516$ ; RMSE=0.253),同时该建模方法针对两生育期 SOM 的估算也具有较优的精度( $R^2=0.439$ ; RMSE=0.308)。而且相较于仅用光谱指数作为自变量的建模结果,发现添加施肥量作为自变量可以显著提高模型的估算精度。

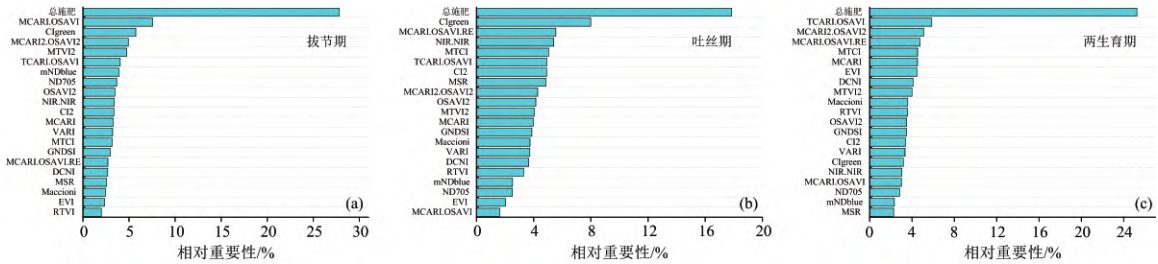


图 4 特征变量在不同生育期的重要性

Fig 4 The importance of indicators at different growth stages

表 3 不同 SOM 估算模型的精度评价

Table 3 Accuracy assessment of different SOM estimation models

模型	自变量	拔节期		吐丝期		两生育期	
		$R^2$	RMSE/%	$R^2$	RMSE/%	$R^2$	RMSE/%
SMLR	光谱指数	0.215	0.443	—	—	0.167	0.483
	光谱指数+施肥量	0.308	0.272	—	—	0.318	0.393
RF	光谱指数	0.109	0.389	0.183	0.633	0.143	0.455
	光谱指数+施肥量	0.423	0.260	0.222	0.390	0.394	0.323
SVM	光谱指数	0.219	0.339	0.185	0.689	0.213	0.523
	光谱指数+施肥量	0.306	0.288	0.214	0.411	0.313	0.332
XGBoost	光谱指数	0.220	0.410	0.202	0.630	0.214	0.420
	光谱指数+施肥量	0.516	0.253	0.362	0.377	0.439	0.308

利用数据验证集对 SMLR、RF、SVM 和 XGBoost 四种模型验证的结果如图 5—图 8 所示。就拔节期而言,这四种模型验证集的  $R^2$  由大到小排序为 XGBoost>RF>SMLR>SVM。结果表明,在拔节期基于 XGBoost 和 RF 建立的 SOM 估算模型的拟合效果要优于基于 SMLR 和 SVM 建立的高光谱反演模型,其中 XGBoost 相较于其他 3 种模型的精度最优,  $R^2$  和 RMSE 分别为 0.590 和 0.222%,能够较好的对 SOM 含量进行估算。而在吐丝期,只有基于 XGBoost 建立

的 SOM 含量高光谱反演模型具备一定估算能力,其  $R^2$  和 RMSE 分别为 0.380 和 0.357%,其他几种模型的  $R^2$  较低,对样本估算的结果较差。两生育期的数据集相较于单生育期来说数据更为复杂,根据其建立的模型更具代表性,但建模也更加困难。而 XGBoost 模型在这种情况下表现出了较优的估算性能,其验证结果的  $R^2$  和 RMSE 分别为 0.453 和 0.277%。

总体来说,这 4 种 SOM 估算模型在拔节期的估算结果

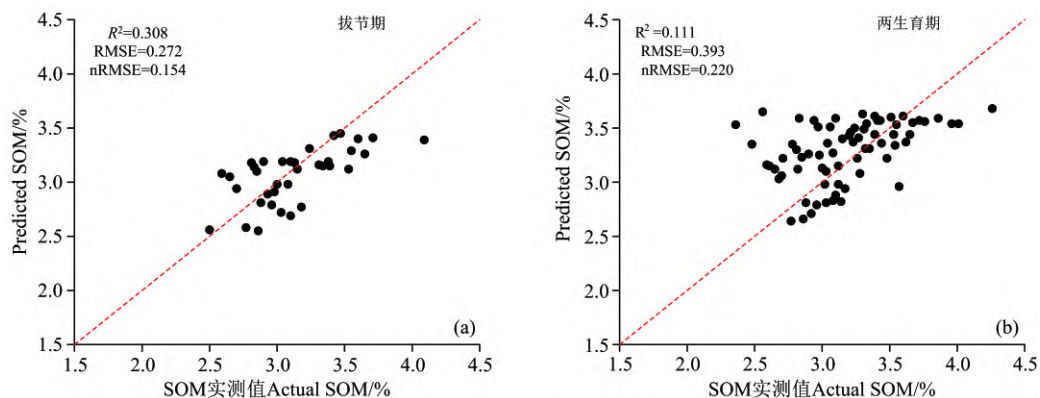


图 5 基于 SMLR 模型的土壤有机质实测值-预测值散点图

Fig 5 Scatter diagrams of measured and predicted SOM values based on SMLR model

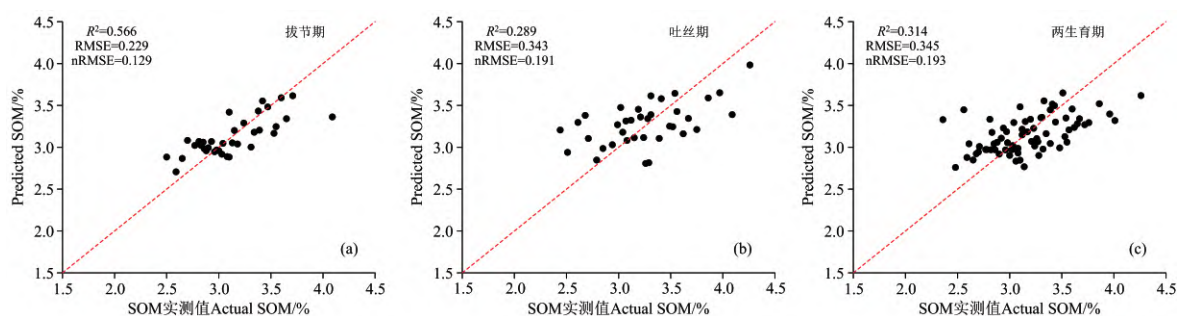


图 6 基于 RF 模型的土壤有机质实测值-预测值散点图

Fig 6 Scatter diagrams of measured and predicted SOM values based on RF model

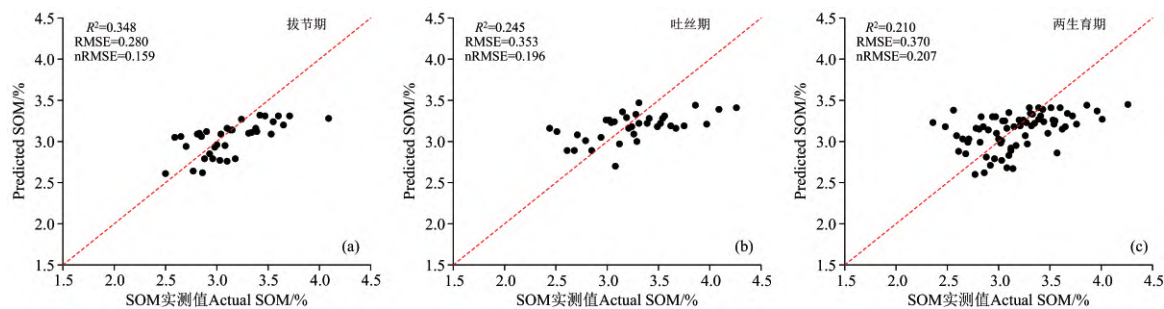


图 7 基于 SVM 模型的土壤有机质实测值-预测值散点图

Fig 7 Scatter diagrams of measured and predicted SOM values based on SVM model

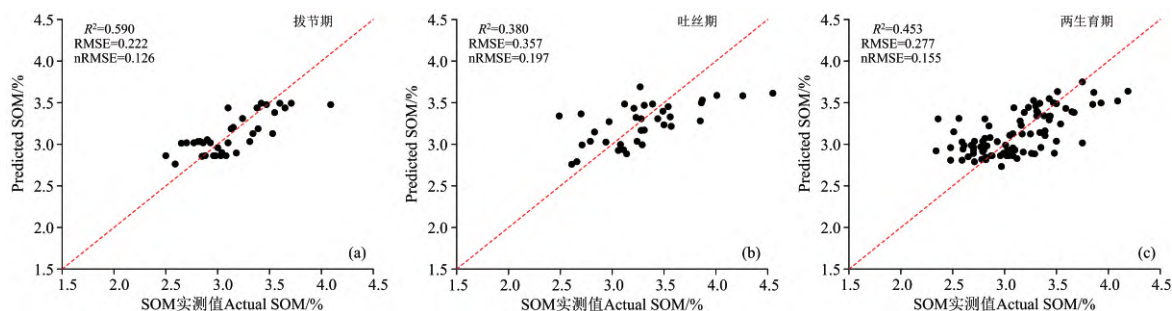


图 8 基于 XGBoost 模型的土壤有机质实测值-预测值散点图

Fig 8 Scatter diagrams of measured and predicted SOM values based on XGBoost model

均优于其他生育期。针对于两生育期组合数据集所建立模型的估算结果, SMLR 和 SVM 的估算精度最低, 模型稳定性相较于 RF 和 XGBoost 较差。XGBoost 相较于其他 3 种模型表现最好, 对于不同生育期 SOM 估算模型的验证结果均有较优的稳定性, 与建模集的结果相一致(表 3)。

### 3 讨 论

无人机技术作为一种新型的遥感手段, 兼具时效快、成本低、空间分辨率高、不受云层影响等优点, 在开展精准农业研究中具有重要意义。无人机高光谱影像含有丰富的光谱信息, 陶培峰<sup>[13]</sup>、刘恬琳<sup>[14]</sup>和 Sami<sup>[15]</sup>的研究表明其在 SOM 估算方面具有切实可行性, 而本研究也发现在田块尺度上进行 SOM 的估算可以充分发挥其作用。然而, SOM 的估算研究大多集中在地形变化明显的大范围区域, 其生物或气候因素差异较大<sup>[20-23]</sup>。本研究区尺度较小, 其气候、地形、坡度和母质等均无明显差异, 增加了 SOM 含量估算的难度。长期施肥对土壤有机质含量有着显著影响, 施用化肥(NPK)会加速 SOM 的分解, 其 SOM 含量从第 3 年开始便下降明显<sup>[27-29]</sup>。而施用钾肥可以提高土壤钾, SOM 对土壤钾具有“稀释效应”, 两者之间存在显著的负相关关系<sup>[55]</sup>。因此, 本研究选用施肥量作为协变量参与建模, 同时通过 SOM 与施肥量的相关分析同样也发现 SOM 与施肥量具有非常显著的负相关关系。

SOM 与可见光及近红外波段之间的响应关系密切, 而黑土有机质响应波段则在 415~1 380 nm<sup>[3]</sup>。本研究通过 SOM 与试验区高光谱数据的相关分析发现, 其响应波段在 450~640 nm 处, 与上述裸土的响应波段存在一定差异。由于地表植被覆盖使得获取地表动态反馈变得更加困难, 土壤有机质与光谱反射率的关系也随之减弱, 而玉米冠层叶绿素含量较高, 使得其与绿波段的相关性更强。吐丝期的 SOM 与光谱反射率的相关性相较于拔节期在 550~640 nm 处的相关关系更弱, 这可能是由于吐丝期相较于拔节期裸露土壤面积减少的原因。土壤有机质与光谱指数的相关性并不太显著, 可能是因为 Pearson 相关系数只能衡量两个分布之间的线性相关关系, 多元变量间线性或非线性的相关关系则不能完全通过 Pearson 相关系数来体现。但通过引入施肥量作为协变量后, 发现施肥量和部分光谱指数与 SOM 的相关性相较于光谱指数与 SOM 的相关性更高。同时, 如表 3 所示, 在田块尺度下将施肥量作为协变量引入可以提高模型的  $R^2$ , 并降低 RSME。

同时, 通过特征重要性评分发现, 施肥量作为协变量对 SOM 估算的贡献率最大, 而各光谱指数的贡献率则相对较小。经过特征变量筛选后所建立的 RF、SVM 和 XGBoost 模型, 根据验证集的验证结果发现 XGBoost 模型在拔节期精度

( $R^2=0.590$ , RMSE=0.222, nRMSE=0.126)、吐丝期精度( $R^2=0.380$ , RMSE=0.357, nRMSE=0.197)和两生育期组合精度( $R^2=0.453$ , RMSE=0.277, nRMSE=0.155), 其估算效果均较好。刘恬琳等<sup>[14]</sup>利用 FieldSpec3 便携式地物光谱仪建立的苹果果园土壤有机质估算模型, 发现 RF 模型反演效果较好( $R^2$  为 0.880, RMSE 为 0.142)。Gu 等<sup>[30]</sup>利用 ASD 高光谱仪进行耕地 SOM 含量的反演研究表明, 使用小波变换和随机森林算法生成的验证模型较好( $R^2$  为 0.748, RMSE 为 0.254)。本研究与其他研究结果相比精度较低, 原因可能是由于本研究选取了不同耕作处理和不同施肥处理两种不同处理条件相结合, 且以整个小区地块为研究基础, 增加了数据的复杂性; 而本研究建立了作物覆盖下土壤有机质的估算模型, 虽然模型的精度有所降低, 但是验证了田块尺度下以施肥量和光谱指数作为协变量所构建模型在植被覆盖下、不同条件田块间的适用性。

XGBoost 是一种兼具线性模型求解器和树的学习算法, 同时支持列抽样, 内置了交叉验证, 相比于 RF 来说更具优势<sup>[31, 50]</sup>。但是在优化模型时, RF 模型只需要调整决策树的数量 ntree 和节点数 mtry 两个关键参数即可, XGBoost 需要调整的参数则更多。为提高模型的估算精度, 今后还需针对模型参数的调整进行尝试和优化, 针对田块尺度的 SOM 估算还需寻找更有价值的辅助变量, 建立更具代表性的 SOM 估算模型。同时, 通过原始波段变换和不同植被指数的选取, 可能会提高有机质含量与光谱特征的相关性, 对模型的精度也会有所影响。此外, 本工作利用高光谱影像数据估算 SOM 含量, 仅利用了同一类型土壤的一年田间采样数据。今后为提高估算模型的精度和适用性, 需要选择不同的黑土类型试验区域并累计多年试验数据进行深入分析。

### 4 结 论

(1)利用无人机高光谱影像在对 SOM 的估算方面具有较大潜力。在植被覆盖条件下, 当波长为 450~640 nm 时, SOM 含量与光谱反射率呈显著相关关系。利用植被冠层的光谱特征间接估算 SOM 含量具有可行性。

(2)在田块尺度下, 研究区域的气候、地形、坡度和母质等因素均无明显差异, 以施肥量和光谱指数作为协变量参与建模可实现 SOM 含量的精准估算。而且, 将施肥量作为协变量可以提高模型的估算精度。

(3)对比 4 种 SOM 估算模型, XGBoost 相较于 SMLR、SVM 和 RF 模型的精度更高, 且针对拔节期 SOM 含量的估算能力最好。本研究可用于耕地质量评价、土壤有机碳库估算, 为农田精准施肥和东北黑土地保护及可持续利用提供理论依据与科学参考。

### References

- [1] WANG Li-gang, YANG Li, HE Mei, et al(王立刚, 杨黎, 贺美, 等). Soil and Fertilizer Sciences in China(中国土壤与肥料), 2016, (6): 1.



- [2] Dotto A C, Dalmolin R S D, Caten A T, et al. *Geoderma*, 2018, 314: 262.
- [3] LIU Huan-jun, PAN Yue, DOU Xin, et al(刘焕军, 潘越, 窦欣, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2018, 34(1): 127.
- [4] WANG Da-ming, QIN Kai, LI Zhi-zhong, et al(汪大明, 秦凯, 李志忠, 等). *Earth Science(地球科学)*, 2018, 43(6): 2184.
- [5] MAO Xin, LIU Lin-jing, LI Chang-an, et al(毛欣, 刘林敬, 李长安, 等). *Earth Science(地球科学)*, 2017, 42(10): 1750.
- [6] ZHOU Wei, XIE Li-juan, YANG Han(周伟, 谢利娟, 杨晗). *Chinese Journal of Soil Science(土壤通报)*, 2021, 52(3): 564.
- [7] ZHAO Ming-song, LIU Bin-yin, LU Hong-liang, et al(赵明松, 刘斌寅, 卢宏亮, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2019, 35(20): 102.
- [8] ZHANG Gan-lin, SHI Zhou, ZHU A-xing, et al(张甘霖, 史舟, 朱阿兴, 等). *Acta Pedologica Sinica(土壤学报)*, 2020, 57(5): 1060.
- [9] Guo L, Fu P, Shi T Z, et al. *Soil and Tillage Research*, 2020, 196: 104477.
- [10] Xu S X, Wang M Y, Shi X Z. *Geoderma*, 2020, 370: 114358.
- [11] Al-Abbas A H, Swain P H, Baumgardner M F. *Soil Science*, 1972, 114(6): 477.
- [12] Krishnan P, Alexander J D, Butler B J, et al. *Soil Society of American Journal*, 1980, 44(6): 1280.
- [13] TAO Pei-feng, WANG Jian-hua, LI Zhi-zhong, et al(陶培峰, 王建华, 李志忠, 等). *Geology and Resources(地质与资源)*, 2020, 29(1): 68.
- [14] LIU Tian-lin, ZHU Xi-cun, BAI Xue-yuan, et al(刘恬琳, 朱西存, 白雪源, 等). *Smart Agriculture(智慧农业(中英文))*, 2020, 2(3): 129.
- [15] Khanal S, Fulton J, Klopfenstein A, et al. *Computers and Electronics in Agriculture*, 2018, 153: 213.
- [16] Nunez-Ramirez F, Santillano Cazares J, Roque Diaz L G, et al. *Terra Latinoamericana*, 2019, 37: 7.
- [17] WANG Xi, LI Yu-huan, WANG Rui-yan, et al(王曦, 李玉环, 王瑞燕, 等). *Chinese Journal of Applied Ecology(应用生态学报)*, 2020, 31(7): 2399.
- [18] SUN Xiao-lin, ZHAO Yu-guo, LIU Feng, et al(孙孝林, 赵玉国, 刘峰, 等). *Chinese Journal of Soil Science(土壤通报)*, 2013, 44(3): 752.
- [19] Jenny H. *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw-Hill, 1941.
- [20] LI Cheng, WANG Rang-hui, LI Zhao-zhe, et al(李成, 王让会, 李兆哲, 等). *Environmental Science(环境科学)*, 2021, 42(5): 2432.
- [21] ZHAO Qing-yue, XU Shi-jie, ZHANG Wu-shuai, et al(赵晴月, 许世杰, 张务帅, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2020, 53(15): 3120.
- [22] Keskin H, Grunwald S, Harris W G. *Geoderma*, 2019, 339: 40.
- [23] WU Cai-wu, ZHANG Yue-cong, XIA Jian-xin(吴才武, 张月丛, 夏建新). *Acta Pedologica Sinica(土壤学报)*, 2016, 53(6): 1568.
- [24] Mosleh Z, Salehi M H, Jafari A, et al. *Environmental Monitoring and Assessment*, 2016, 188(3): 195.
- [25] Zeng C Y, Zhu A X, Liu F, et al. *Ecological Indicators*, 2017, 72: 297.
- [26] Guo S X, Zhu A X, Meng L K, et al. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 49: 126.
- [27] Guo L, Sun X R, Fu P, et al. *Geoderma*, 2021, 398: 115118.
- [28] XU Ming-gang, YU Rong, SUN Xiao-feng, et al(徐明岗, 于荣, 孙小凤, 等). *Journal of Plant Nutrition and Fertilizers(植物营养与肥料学报)*, 2006, (4): 459.
- [29] XU Zhi-qiang, DAI Ji-guang, YU Xiang-hua, et al(徐志强, 代继光, 于向华, 等). *Chinese Journal of Soil Science(土壤通报)*, 2008, (4): 766.
- [30] Yue J B, Feng H K, Yang G J, et al. *Remote Sensing*, 2018, 10(1): 66.
- [31] Saman A, Arie P, Selam A, et al. *Computers and Electronics in Agriculture*, 2018, 148: 250.
- [32] Mojtaba Z, Shamsollah A, Azam J, et al. *Geoderma*, 2018, 338: 445.
- [33] Rodrigues E, Gomes á, Gaspar A R, et al. *Renewable and Sustainable Energy Reviews*, 2018, 94: 959.
- [34] Gu X H, Wang Y C, Sun Q, et al. *Computers and Electronics in Agriculture*, 2019, 167: 105053.
- [35] GE Xiang-yu, DING Jian-li, WANG Jing-zhe, et al(葛翔宇, 丁建丽, 王敬哲, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(2): 602.
- [36] FENG Guo-zhong, WANG Yin, YAN Li, et al(冯国忠, 王寅, 焉莉, 等). *Acta Pedologica Sinica(土壤学报)*, 2017, 54(2): 444.
- [37] BAO Shi-dan(鲍士旦). *Soil Agrochemical Analysis(土壤农化分析)*. Beijing: China Agriculture Press(北京: 中国农业出版社), 2000. 172.
- [38] Huete A, Didan K, Miura T, et al. *Remote Sensing of Environment*, 2002, 83(1-2): 195.
- [39] Gitelson A A, Kaufman Y J, Stark R, et al. *Remote Sensing of Environment*, 2002, 80(1): 76.
- [40] Mistele B, Gutser R, Schmidhalter U, et al. *Validation of Field-Scaled Spectral Measurements of the Nitrogen Status in Winter Wheat 7th International Conference on Precision Agriculture and other Precision Resources Management*. 2004. 1187.
- [41] Gamon J A, Penuelas J, Field C B. *Remote Sensing of Environment*, 1992, 41: 35.
- [42] Sylvain J, Nathalie G, Julien M, et al. *Remote Sensing of Environment*, 2017, 198: 173.

- [43] Wu C Y, Niu Z, Tang Q, et al. *Agricultural and Forest Meteorology*, 2008, 148(8): 1230.
- [44] Maccioni A, Agati G, Mazzinghi P. *Journal of Photochemistry & Photobiology B: Biology*, 2001, 61(1): 52.
- [45] Dash J, Curran P J. *International Journal of Remote Sensing*, 2004, 25(23): 5403.
- [46] Chen J M. *Canadian Journal of Remote Sensing*, 2014, 22(3): 229.
- [47] Chen P F, Haboudane D, Tremblay N, et al. *Remote Sensing of Environment*, 2010, 114(9): 1987.
- [48] Gitelson A A, Merzlyak M N. *Journal of Plant Physiology*, 1996, 148(3): 494.
- [49] Gitelson A A, Vina A, Ciganda V, et al. *Geophysical Research Letters*, 2005, 32(8): L08403, doi: 10.1029/2005GL022688.
- [50] Daughtry C S T, Walthall C L, Kim M S, et al. *Remote Sensing of Environment*, 2000, 74(2): 229.
- [51] Zarco-Tejada P J, Miller J R, Morales A, et al. *Remote Sensing of Environment*, 2004, 90(4): 463.
- [52] Haboudane D, Tremblay N, Miller J R, et al. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46: 423.
- [53] Haboudane D, Miller J R, Pattey E, et al. *Remote Sensing of Environment*, 2004, 90(3): 337.
- [54] Chen T Q, Guestrin C. XGBoost: A Scalable Tree Boosting System Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. arXiv: 1603.02754[CS LG].
- [55] XIE Jia-gui, HOU Yun-peng, YIN Cai-xia, et al(谢佳贵, 侯云鹏, 尹彩侠, 等). *Journal of Plant Nutrition and Fertilizers(植物营养与肥料学报)*, 2014, 20(5): 1110.

## Estimation of Soil Organic Matter in Maize Field of Black Soil Area Based on UAV Hyperspectral Image

XIA Chen-zhen<sup>1, 2, 3</sup>, JIANG Yan-yan<sup>4</sup>, ZHANG Xing-yu<sup>1, 2, 3</sup>, SHA Ye<sup>5</sup>, CUI Shuai<sup>1, 2, 3</sup>, MI Guo-hua<sup>5</sup>, GAO Qiang<sup>1, 2, 3</sup>, ZHANG Yue<sup>1, 2, 3\*</sup>

1. College of Resources and Environment, Jilin Agricultural University, Changchun 130118, China

2. Key Laboratory of Soil Resource Sustainable Utilization for Commodity Grain Bases of Jilin Province, Jilin Agricultural University, Changchun 130118, China

3. Key Laboratory of Straw Comprehensive Utilization and Black Soil Conservation, Ministry of Education, Changchun 130118, China

4. The Monitoring Center of Soil and Water Conservation, Songliao Water Resources Commission, Changchun 130021, China

5. College of Resources and Environment, China Agricultural University, Beijing 100083, China

**Abstract** As an important part of the soil, the soil organic matter (SOM) can reflect soil fertility and quality. Compared with the traditional SOM measurement method, UAV hyperspectral images can quickly and real-time obtain the SOM content at the field-scale, which is of great significance for precision fertilization and sustainable utilization in the black soil region of Northeast China. In order to explore the difference in estimating the accuracy of SOM under crop cover by linear and nonlinear models based on hyperspectral data, the soil samples at the jointing stage and silking stage and UAV hyperspectral images were collected from the experimental corn field in the black soil region of Northeast China as the study area. The correlation between soil spectral reflectance and SOM content under crop cover was analyzed, and the spectral indices were calculated according to their response band. With the fertilizer rates and spectral indices as independent variables, multiple stepwise linear regression models (SMLR), support vector machine (SVM), random forest (RF) and eXtreme gradient boosting (XGBoost) models were established by screening characteristic variables, and the accuracies of the models were verified and compared (select  $R^2$  and RMSE as evaluation indicators). The results showed that the response band of SOM content under crop cover was 450~640 nm. Long-term application of chemical fertilizers had a significant effect on SOM content, and introducing it into the model as a covariate significantly improved the estimation accuracy of SOM. The test accuracies of the four models were: XGBoost>RF>SMLR>SVM, and the estimation result of XGBoost at the jointing stage was the best ( $R^2$  and RMSE of modeling set were 0.516, 0.253, and those of the verification set were 0.590, 0.222, respectively). Therefore, UAV hyperspectral technology can rapidly estimate SOM content in maize fields at field-scale, and the XGBoost model is a preferable option for estimating SOM content under crop cover conditions.

**Keywords** UAV; Hyperspectral; Soil organic matter; Stepwise multiple linear regression; Machine learning

\* Corresponding author

(Received Apr. 6, 2022; accepted Jul. 21, 2022)