

一种基于无人机高光谱影像的土壤墒情检测新方法

葛翔宇^{1, 2, 3}, 丁建丽^{1, 2, 3*}, 王敬哲⁴, 孙慧兰⁵, 朱志强⁶

1. 新疆大学资源与环境科学学院, 新疆 乌鲁木齐 830046
2. 新疆大学绿洲生态教育部重点实验室, 新疆 乌鲁木齐 830046
3. 新疆大学智慧城市与环境建模自治区普通高校重点实验室, 新疆 乌鲁木齐 830046
4. 深圳大学海岸带地理环境监测自然资源部重点实验室, 广东 深圳 518060
5. 新疆师范大学地理科学与旅游学院, 新疆 乌鲁木齐 830054
6. 北京化工大学材料科学与工程学院, 北京 100029

摘要 土壤含水量(SMC)是生物地球化学和大气耦合过程的关键变量,在干旱区农业、生态和环境扮演着重要角色。相较于星载遥感系统,无人机(UAV)具有可控性强、分辨率高等特点从而被广泛应用,为中小尺度地表参量的快速监测提供新的遥感平台。机载高光谱传感器的引入,为UAV遥感系统提供了高维海量、纳米级的数据源。然而基于UAV高光谱数据的研究并未深度挖掘,也尚未形成一个标准的技术方案。该研究立足于新疆维吾尔自治区典型农业区,利用6种预处理方案,包括一阶导数(FDR),二阶导数(SDR),连续体去除(CR)、吸光度(A)、吸光度一阶(FDA)和吸光度二阶(SDA),对所获取的UAV高光谱数据进行处理。在此背景下构建4种类型的适宜光谱指数:差值型指数(DI),比值型指数(RI),归一化型指数(NDI)和垂直型指数(PI),并从光谱机理上讨论指数的合理性。最后利用梯度提升回归树(GBRT)、随机森林(RF)和eXtreme Gradient Boosting(XGBoost)算法,以28个最适光谱指数为独立变量建立SMC估算模型,并通过不同集成学习算法的重要性对变量进行排序,从线性和非线性的角度对所构建光谱指数的效果进行考量评价。结果表明:(1)预处理和最适光谱指数能有效地消除了大气干扰和土壤背景,其中预处理A突出更多的光谱信息,PI相关性显著;(2)通过分析比较相关性系数(r)和集成学习算法的重要性,发现A-PI($|r|=0.773$)是最适光谱指数,在线性和非线性关系中均有较优的表现;(3)在3种基于集成学习的SMC预测模型中,XGBoost估算模型效果拔群($R_{\text{val}}^2=0.926$, RMSEP=1.943和RPD=2.556),其预测值的统计学特征与实测值的最为接近。3种模型效果排序为:XGBoost>RF>GBRT。综上所述,基于UAV高光谱影像,结合不同预处理和光谱指数,为低空遥感监测土壤墒情提出新的方案。该研究的方案具有潜在的高精度,是检测干旱区SMC的有效方法,针对快速易行地监测地表属性提供了崭新视角。相关结果为干旱区精准农业、生态系统给予更好的管理和保护策略。

关键词 UAV; 遥感; 高光谱; 机器学习; 集成学习

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)02-0602-08

引言

土壤含水量(soil moisture content, SMC)影响土壤的理化过程,参与全球生态、环境、水温和气候变化模式^[1]。同时,SMC是约束土壤养分状况的关键因素,是影响精准农业发展的首要因素^[2]。绿洲农业承载着新疆维吾尔自治区的第

一产业,近年来的人类活动加剧干旱区内墒情失衡,使绿洲生态环境暴露在干旱与盐渍化灾害中^[3]。在实施可持续土壤管理实践和精准农业时,了解SMC的空间分布对于确定土壤墒情监测和土壤水盐运移的区域至关重要。因此,快速易行的获取SMC数据对农业监测、产量估算以及合理灌溉具有必要的现实意义。

近十年来遥感技术的迅猛发展,在地表观测的各领域得

收稿日期: 2018-11-22, 修订日期: 2019-04-21

基金项目: 国家自然科学基金项目(41961059, 41771470)资助

作者简介: 葛翔宇, 1994年生, 新疆大学资源与环境科学学院博士研究生 e-mail: xiangyu_gexj@163.com

* 通讯联系人 e-mail: watarid@xju.edu.cn

到了广泛应用^[3-4]。其中无人机(unmanned aerial vehicle, UAV)遥感技术的推广,使得大尺度、高效率地获取 SMC 信息成为可能^[4]。纵观相关研究的探索,光学遥感能有效监测地物参量,其中植被冠层可反映植被的生长状况和健康状况,其光谱特征在不同土壤水分胁迫条件下表现出差异^[5-6]。因此,使用 UAV 衍生的高光谱植被数据可作为准确评估 SMC 的替代方案。然而高光谱会产生信息冗余和背景噪声等问题^[7]。光谱指数可以解决在高光谱定量估算研究中的类似问题,易检测到敏感波长并且进一步增强目标的特定属性和光谱特性之间的相关性^[8]。通过冠层光谱指数模型,讨论受到水胁迫的各类参数与 SMC 具有良好相关性^[9]。然而,未经预处理的数据是几种复合信号与各种重叠数据的组合,很难实现深度数据挖掘。为弥补这一不足,引入预处理来消除部分外部噪声,增强非线性关系并提高地表参数估计模型的准确性。

就光谱建模而言,机器学习的应用提供了较为理想的光谱建模方案,大大提高了估算模型的预测精度和鲁棒性^[7]。集成学习作为机器学习的重要领域,在机器学习和数据挖掘研究中备受瞩目^[10]。以随机森林(random forest, RF)^[11]为代表的估算模型研究屡见不鲜,然而以梯度提升回归树(gra-

dient-boosted regression tree, GBRT)和 XGBoost(eXtreme gradient boosting)等梯度提升算法却鲜有报道。基于此,基于 UAV 高光谱数据,在不同的预处理方案下,探讨 SMC 与不同光谱指数之间的关系。通过预处理和光谱指数构建基于 RF, GBRT 和 XGBoost 的 SMC 定量估算模型,并尝试对 UAV 高光谱图像进行 SMC 数字化绘图,以预测干旱地区农业表层 SMC 的空间分布,为干旱区土壤墒情的低空遥感平台进行有益尝试,进一步为精准农业提供科学方案。

1 实验部分

1.1 数据采集

采用 DJI Matrice600 PRO[®] 六旋翼 UAV, 搭载 Nano-Hyperspec[®] 高光谱传感器。高光谱传感器的波段范围为 400~1 000 nm, 光谱分辨率为 6 nm, 重采样间隔为 2.2 nm。采集时间为 2018 年 4 月 17 日 15:00, 并对传感器进行暗电流校正及光谱定标。后期利用 Hyperspec III 及 Headwall SpectralView 软件处理校正。布设 70 个 50 cm×50 cm 样方, 四点采样封装于铝盒, 并用 GPS 保存各点坐标, 后期在室内利用烘干法(105 °C 的恒温箱, 48 h)测定 SMC。

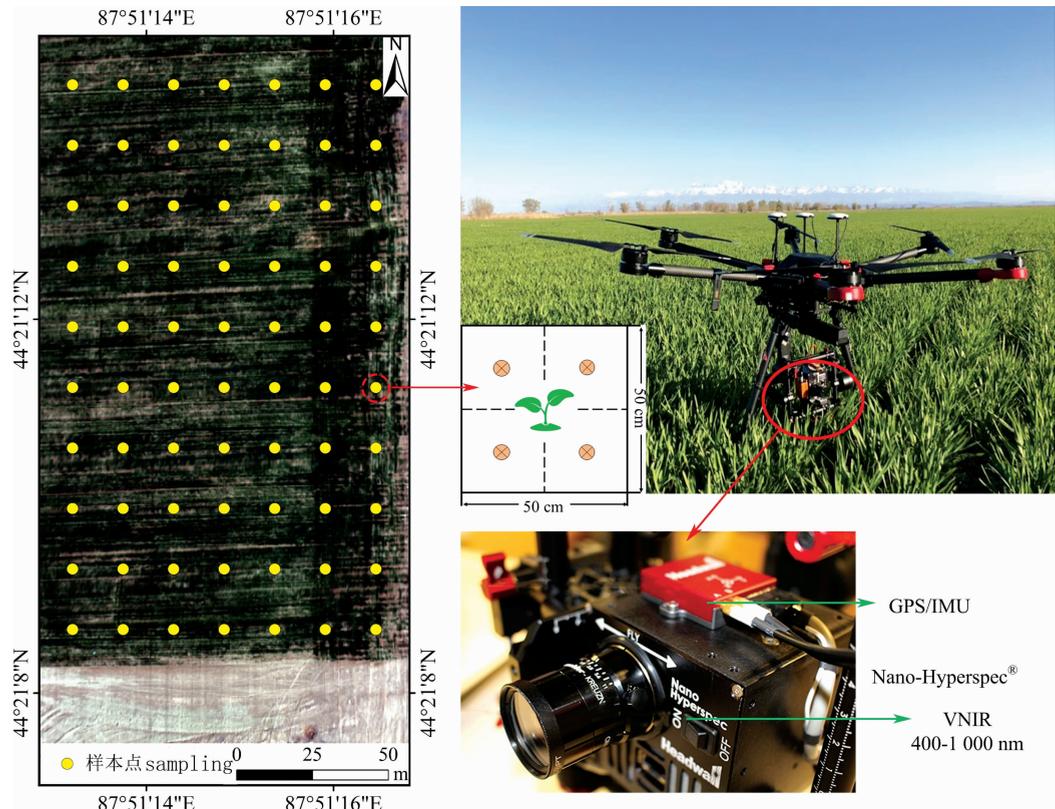


图 1 采样点分布、高光谱传感器说明及 UAV 野外作业

Fig. 1 Sampling site point distribution, hyperspectral sensor description, and drone field work

1.2 数据预处理

在 MATLAB R2016b 环境中基于 Savitzky-Golay(SG)滤波器平滑了高光谱图像。以 SG 平滑的高光谱图像为基础影像(R), 基于 IDL + ENVI 平台对原始数据进行一阶导数

(FDR), 二阶导数(SDR), 连续体去除(CR)、吸光度(A)、吸光度一阶(FDA)和吸光度二阶(SDA)6 种预处理方案。

1.3 适宜光谱指数构建

为了充分利用光谱数据, 根据以往的研究选择差值型指

数(DI), 比值型指数(RI), 归一化型指数(NDI)和垂直型指数(PI)。其中, PI的常数项是根据高光谱影像提取的土壤线的系数 $y=0.4401x+0.3308$, 土壤线方程为。4个光谱指数的数学表达式如式(1)~式(4)

$$DI_{(R_i, R_j)} = R_i - R_j \quad (1)$$

$$RI_{(R_i, R_j)} = R_i / R_j \quad (2)$$

$$NDI_{(R_i, R_j)} = (R_i - R_j) / (R_i + R_j) \quad (3)$$

$$PI_{(R_i, R_j)} = (R_i - 0.4401R_i - 0.3308) / (\sqrt{1 + 0.4401^2}) \quad (4)$$

其中 R_i 和 R_j 是在高光谱传感器的工作范围内(400~1 000 nm)任意获得的 i 和 j 的光谱反射率。使用 MATLAB R2016b 实现指数与 SMC 的相关性及指数的波段组合。

1.4 模型的构建与评价

RF 回归是较常见的机器学习算法之一, 具有理想的估计能力, 特别是对于高维数据集^[9]。它也是一种基于分类回归树(CART)的集成学习算法。该算法关键参数有决策树的数量 ntree(设置 ntree=500)和节点数 mtry(mtry=10, 即总变量数的 1/3)。GBRT 是针对回归和分类问题的机器学习技术^[12]。每次迭代都适合对前一次遗留的残差的决定, 然后通过组合树来完成预测。考虑到同 RF 和 XGBoost 进行比较, 相关参数与 XGBoost 保持一致。

XGBoost 是一个可扩展且灵活梯度增强的优化实现, 有关 XGBoost 的详细说明, 请参阅文献^[13]。为确定最优迭代次数, 防止过拟合, 选用交叉验证的方法, 设定 nfold=3, 测试结果如图 2 所示, 当迭代次数(nround)>300 时验证集的精确度将不再提高, 为了便于同 RF 算法公平对比, 故本文选择 nround=500 作为模型参数。(其他参数设置: 学习速率(eta)=0.01, 每棵树的最大深度(max_depth)=10)。

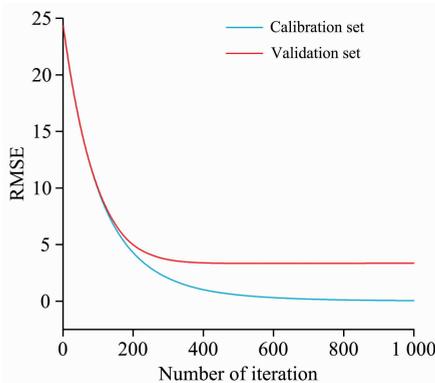


图 2 XGBoost 最优迭代次数选择

Fig. 2 XGBoost optimal iteration number selection

上述 3 个模型基于 R 3.5.0 平台实现。为了量化 3 种集成学习估算模型的性能, 使用以下指标评估模型的效果^[7]: 确定系数(R^2), 均方根误差(RMSE)和相对百分比偏差(RPD)。

2 结果与讨论

2.1 SMC 及光谱特征

样本划分基于联合 $x-y$ 距离 (simple set portioning based on joint $x-y$ distance, SPXY)^[14] 算法进行, 选取 50 个样点作为建模集, 20 个样点作为验证集。全集、建模集和验证集具有以下描述性统计结果(图 3): 全集 SMC 的均值(Mean)为 24.446%, 标准差(SD)为 5.408%。建模集(12.23%~37.63%)的 Mean 和验证集(14.95%~35.62%)的 Mean 分别为 24.499% 和 24.374%。SD 和 Mean 相似表明建模集和验证集与全集 SMC 保持类似的统计分布, 在确保代表性样本的同时尽可能缩减了建模集和验证集中存在偏差的估计。

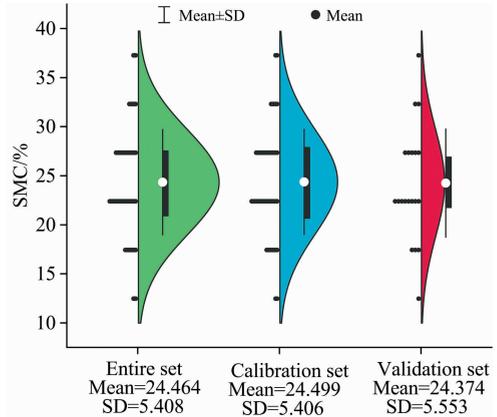


图 3 SMC 样本划分及其描述性统计

Fig. 3 SMC samples and their descriptive statistics

UAV 高光谱影像进行了 R 和 FDR, SDR, CR, A, FDA 和 SDA 共 6 种预处理, 结果如图 4, 红线代表平均光谱, 灰色区域代表平均光谱加减标准差。随着导数的阶数增加, 处理后的光谱的强度也会降低, 而 A 和 CR 增强了某些区域的光谱, 尤其突显了红边信息。

2.2 适宜光谱指数构建

为了可视化表达 SMC 和光谱指数(DI, RI, NDI 和 PI)之间的相关性, 通过二维相关图来表达(图 5)。图 5 色柱表示 SMC 与光谱指数之间的相关系数的平方(r^2), x 轴和 y 轴表示 400~1 000 nm 的波段。暗红色描绘了 SMC 和光谱指数之间的高 r^2 。为了更好地描述, 下面中将 r^2 转换为相关系数的绝对值($|r|$)。

如表 1 所示, 使用 SMC 建立的 28 个光谱指数均通过 0.01 水平的显著性检验(阈值为 ± 0.380), 其中相关性最高的 4 个光谱指数 DI, RI, NDI 和 PI 的 $|r|$ 分布范围为 0.475~0.773。在不同预处理方案下, 光谱指数也表现出明显差异。其中, R_PI, A_PI, CR_NDI 和 CR_RI 都在 0.75 以上, 这四个指数的效果比较显著。基于 FDR, SDR 和 SDA 的 3 种预处理方案构建的光谱指数效果较差。不同的预处理方案可以不同程度地改善光谱指数与 SMC 之间的相关性, 最优光谱指数为 A_PI($|r|=0.773$)。

2.3 构建模型与评价

通过多参量驱动集成学习模型实现较好的 SMC 估算模型, 28 个光谱指数作为独立变量, 70 个 SMC 实测值作为响应变量, 构建基于 GBRT, RF 和 XGBoost 的 SMC 估算模型。建模效果比较(表 2)显示 XGBoost 模型表现出优越的预

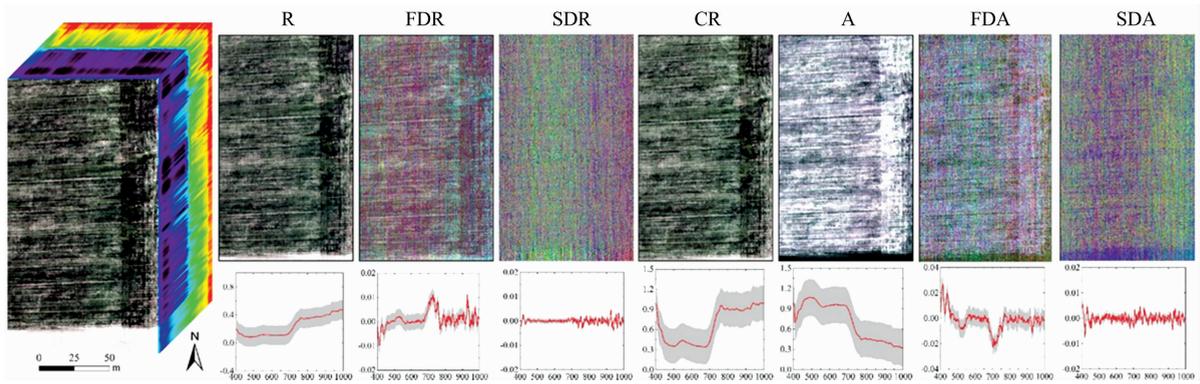


图 4 基于 UAV 高光谱影像的预处理结果

Fig. 4 The hyperspectral images and spectral curves based on different pretreatments

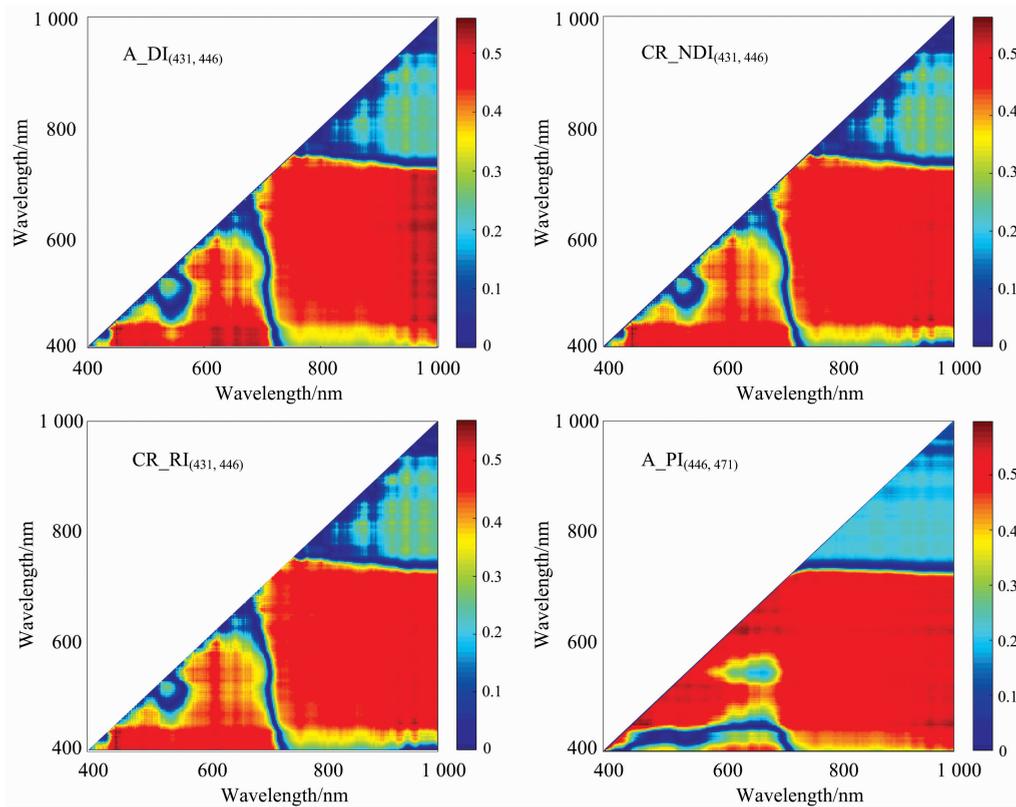


图 5 SMC 与最优光谱指数的 r^2 图

Fig. 5 r^2 maps of spectral indices and SMC

表 1 基于不同预处理的光谱指数与 SMC 相关性 $|r|$

Table 1 $|r|$ between SMC and spectral indices based on different pretreatments

预处理	光谱指数	$ r $									
R	$DI_{(R479, R619)}$	0.724	FDR	$DI_{(R435, R746)}$	0.662	SDR	$DI_{(R710, R753)}$	0.551	CR	$DI_{(R400, R446)}$	0.737
	$NDI_{(R431, R446)}$	0.748		$NDI_{(R702, R726)}$	0.674		$NDI_{(R417, R753)}$	0.487		$NDI_{(R431, R446)}$	0.755
	$RI_{(R431, R446)}$	0.747		$RI_{(R702, R724)}$	0.668		$RI_{(R444, R895)}$	0.475		$RI_{(R431, R446)}$	0.755
	$PI_{(R446, R471)}$	0.772		$PI_{(R435, R744)}$	0.693		$PI_{(R653, R753)}$	0.554		$PI_{(R446, R466)}$	0.746
A	$DI_{(R431, R446)}$	0.748	FDA	$DI_{(R435, R744)}$	0.742	SDA	$DI_{(R579, R753)}$	0.577			
	$NDI_{(R431, R619)}$	0.725		$NDI_{(R513, R726)}$	0.616		$NDI_{(R677, R753)}$	0.561			
	$RI_{(R461, R619)}$	0.720		$RI_{(R420, R726)}$	0.624		$RI_{(R440, R446)}$	0.424			
	$PI_{(R446, R471)}$	0.773		$PI_{(R435, R713)}$	0.738		$PI_{(R753, R946)}$	0.569			

测能力, $R_{\text{val}}^2=0.926$, $\text{RMSEP}=1.943$, $\text{RPD}=2.556$ 。RF 模型的预测能力次之, GBRT 模型预测最不理想。XGBoost 模型建模的拟合效果最好 ($R_{\text{cal}}^2=0.925$, $\text{RMSEC}=2.198$), RF 和 GBRT 两种模型建模拟合效果较为相近。

预测精度越高表明估计值接近实测值, 即实测 SMC 和预测 SMC 之间的散点图接近 1:1 线。如图 6 所示, 其中 XGBoost 模型比其他模型更接近 1:1 线, 而其他两种模型在低值区域出现明显的高估。为了更直观地表示 3 种 SMC 估算模型的效果, 在模型验证散点图的基础上引入泰勒图^[15]。如图 6(d) 所示, 黑线表示 R_{val}^2 , 蓝线表示 SD, 彩色五角星表示 3 种估算模型, 其颜色从蓝绿色到橙红色表示 RMSEP 从小到大。红线表示测量的 SMC。五角星越靠近红线, 模型预测值更接近于测量 SMC, 并具有相似的统计特征。

GBRT 模型到红线距离最远, XGBoost 模型比其他两种模型更接近红线, 同时具有最高的 R_{val}^2 。因此, 综合考量后 3 种 SMC 预测模型的模型效果排序为: XGBoost>RF>GBRT。

表 2 不同 SMC 预测模型建模效果

Table 2 Calibration and validation results for SMC estimation based on different modeling strategies

模型	建模集		验证集		
	R_{cal}^2	RMSEC	R_{val}^2	RMSEP	RPD
GBRT	0.897	2.419	0.887	2.103	2.002
RF	0.892	1.993	0.905	2.058	2.109
XGBoost	0.925	2.198	0.926	1.943	2.556

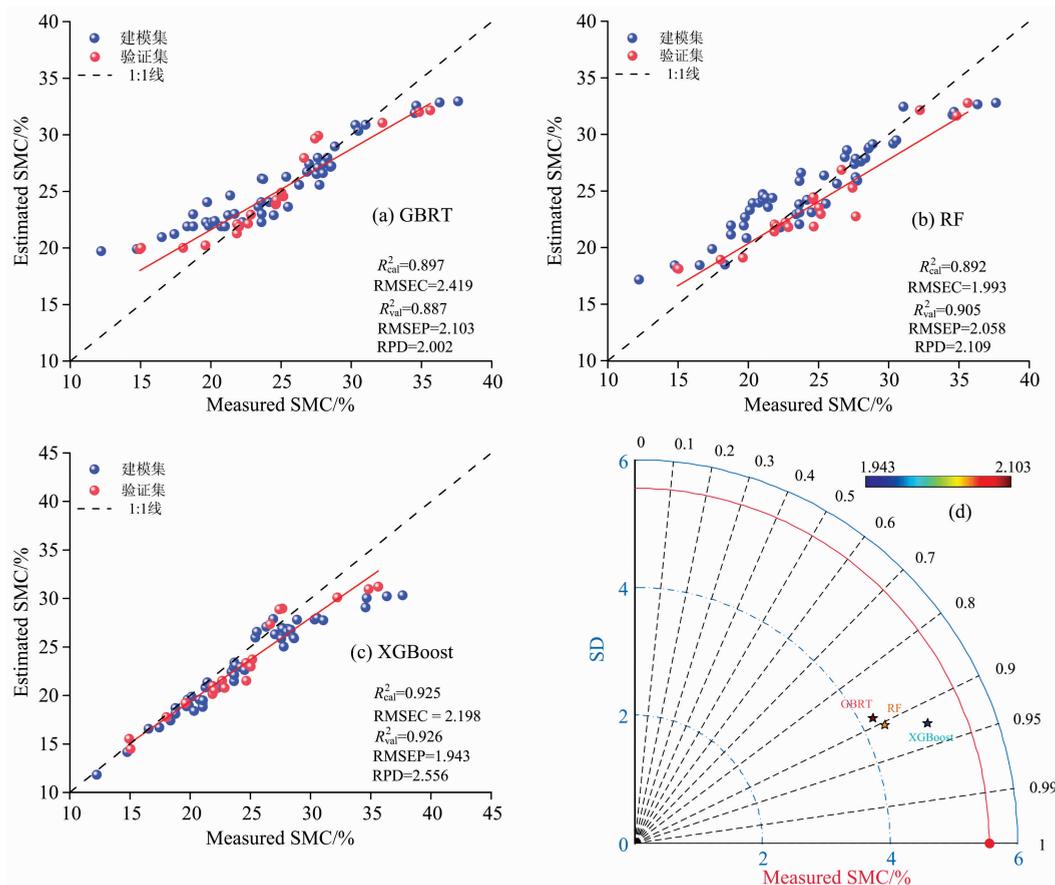


图 6 不同模型的验证结果

Fig. 6 Validation results of measured and predicted SMC based on different modeling methods

此外, 集成学习能通过重要性来演绎各筛选变量对模型结果的影响^[12]。如图 7 所示, 选取了前 12 个变量经行重要性表达, A_PI 指数在 3 种模型中均位列第一, 可见 A_PI 指数的在与 SMC 非线性关系中的重要性。在 4 类适宜光谱指数 (DI, RI, NDI 和 PI) 中, 由于 PI 在构建指数时考虑到土壤背景对植被光谱的干扰, 故而指数的效果在众多指数中名列前茅。DI, NDI 和 RI 指数效果与前人研究具有相似性^[5], 相较而言, PI 对 UAV 高光谱影像消除背景噪声方面略胜一筹。

简单光谱指数仅考虑光谱与物体之间的相互作用, 而不考虑反射光谱之间的相互作用^[9]。由于受到高光谱传感器限制, 主要光谱集中在 400~1 000 nm。农田植被覆盖度高, 难以直接通过土壤光谱信息进行估算。由于干旱区农业植物在不同程度的水分胁迫下, 作物冠层叶绿素随干旱程度波动, 故而土壤水分含量与叶绿素之间存在强烈的正相关关系。基于实际效用考虑, 充分利用植被冠层光谱信息作为媒介, 间接监测土壤含水量。根据适宜光谱指数所得到的波段组合 (表 1) 较集中的处于 420, 440, 460 和 700 nm 附近。420~

460 nm 附近是叶绿素、类胡萝卜素和水分的强吸收带^[5]，由此佐证了指数构建的合理性。本研究中确定的信息光谱带和区域可为传感器的光谱通道设计提供参考，以监测干旱区的农业土壤墒情。

利用 6 种预处理方式 (FDR, SDR, CR, A, FDA 和 SDA) 对 UAV 高光谱影像的处理达到了一定效果。由于一维光谱表达的局限性，引入由 Noda^[16] 开发二维同步相关光谱，

以便检测到更多难以在一维光谱中发现的光谱信息。显然在图 8 中，这些二维同步光谱图的对角线上存在一些自相关峰。他们是证明两者间对外部扰动有协同响应。对比不同预处理下的自相关峰，一阶导数和二阶导数虽能剔除大量无关信息，并同时也损失了更多协同响应的光谱信息。在 R, CR 和 A 中比较，R 对光谱信息表达能力不如 CR；A 方案的光谱信息表达最佳。在 A 中，出

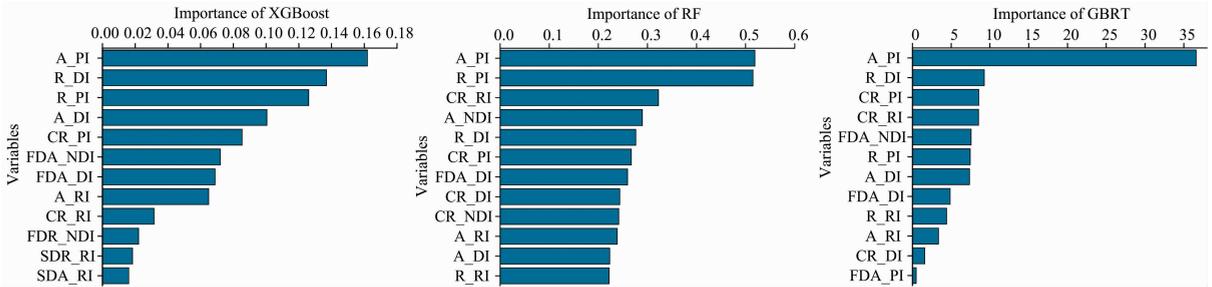


图 7 不同集成学习的变量重要性

Fig. 7 The importance of variables in different model of ensemble learning

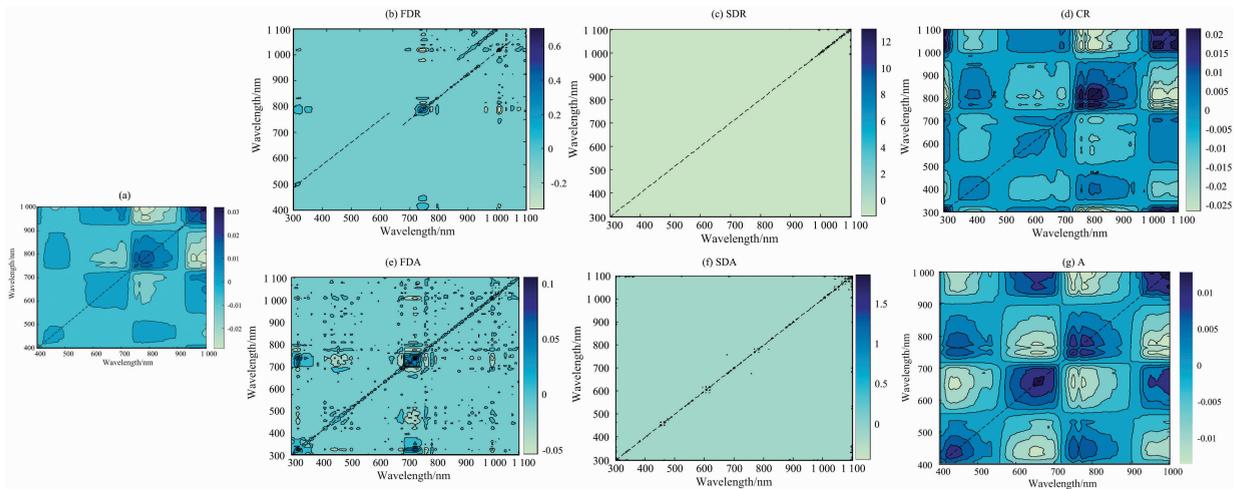


图 8 不同预处理方案的二维同步相关图

Fig. 8 2D synchronization correlation diagrams under different pretreatments

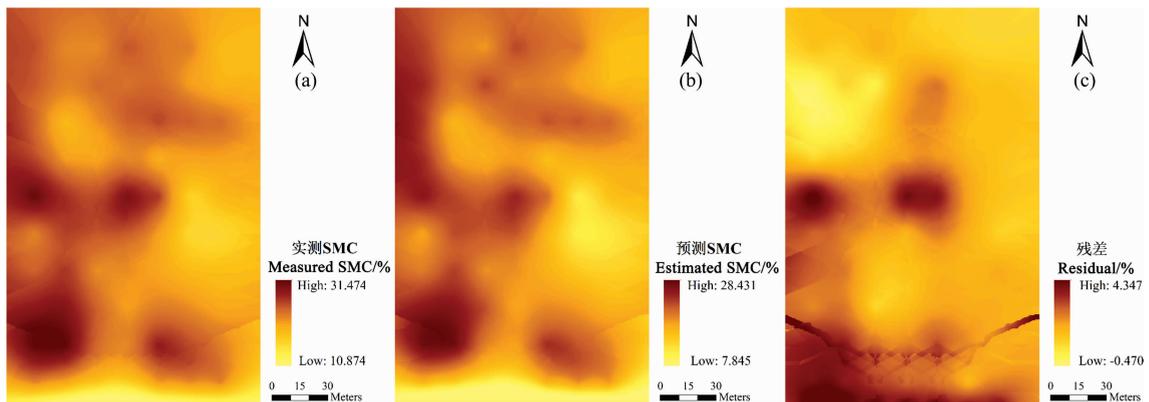


图 9 SMC 空间数字制图

Fig. 9 Spatial distribution maps of (a) the measured SMC, (b) the SMC based on XGBoost prediction, (c) residuals calculated with XGBoost for prediction of the SMC

现四个自相关峰,分别位于 450 nm 附近,670 nm 附近,740 nm 附近以及 980 nm 附近,说明在四个区域植被光谱发生显著变化。这与前文光谱指数合理性的结果相似,在证明预处理效果的同时,佐证了本文光谱指数的响应机制。

机器学习算法被广泛地应用在地表属性的估算中,XGBoost 估算 SMC 模型也在空谱结合的空间数字制图中得到较好的结果(图 9),图 9(a)与(b)表现出相似的空间分布特征,残差最大为 4.347%。这表明该模型可以很好的对 SMC 可视化表达。

利用 UAV 高光谱影像数据,在不同 SMC 估算模型下,收获理想的估算结果,在“星-空-地”遥感系统应用中尝试一种可替代方案,并为低空遥感监测土壤墒情进行了有益探索。因条件所限,未能根据植被物候及土壤养分进行多时相影像数据采集,所建立的 SMC 的机器学习估算模型迁移能力需进一步完善。因此,进一步的科研工作将着手于时间和区域尺度的 SMC 与植被光谱的研究,为应对日益恶化干旱

区生态环境提供精准农业的技术指导,为相关部门提供作物长势监测、估算产量、合理灌溉及早情监测的科学方案。

4 结 论

探讨了在干旱区背景下利用 UAV 高光谱影像估算农业 SMC 的有效方案。利用适宜光谱指数策略,构建集成学习算法估算模型。预处理和光谱指数能有效消除了大气干扰和土壤背景,A_{PI} 具有显著相关性($|r|=0.773$)。总体而言,XGBoost 模型比 RF 和 GBRT 模型有更好的预测结果。XGBoost 模型对于估计 SMC 的具有最佳精度($R_{\text{val}}^2=0.926$, RMSEP=1.943 和 RPD=2.556)。通过 XGBoost 模型预测的 SMC 保持了与实测数据集最接近的统计特征和形态,最优模型将 SMC 从单点尺度扩展到区域尺度。本研究探索出基于 UAV 高光谱影像的快速易行的农业监测方案,为当地环境和农业管理部门提供科学有效手段。

References

- [1] Amani M, Salehi B, Mahdavi S, et al. *Remote Sensing of Environment*, 2017, 197: 1.
- [2] Kumar S V, Dirmeyer P A, Peters-Lidard C D, et al. *Remote Sensing of Environment*, 2018, 204: 392.
- [3] DING Jian-li, WANG Fei(丁建丽, 王 飞). *Acta Geographica Sinica(地理学报)*, 2017, 72(1): 64.
- [4] Jin X, Liu S, Baret F, et al. *Remote Sensing of Environment*, 2017, 198: 105.
- [5] Neto A J S, Lopes D C, Pinto F A C, et al. *Biosystems Engineering*, 2017, 155: 124.
- [6] Mu X, Song W, Gao Z, et al. *Remote Sensing of Environment*, 2018, 216: 44.
- [7] Nawar S, Mouazen M A. *Sensors*, 2017, 17(10): 2428.
- [8] Hong Y, Chen S, Zhang Y, et al. *Science of the Total Environment*, 2018, 644: 1232.
- [9] Jay S, Baret F, Dutartre D, et al. *Remote Sensing of Environment*, 2019, 231: 110898.
- [10] Merentitis A, Debes C. *IEEE Geoscience and Remote Sensing Magazine*, 2015, 3(3): 86.
- [11] Belgiu M, Drăguț L. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24.
- [12] Liu L, Ji M, Buchroithner M. *Remote Sensing*, 2017, 9(12): 1299.
- [13] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proc 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016. 785. 10.1145/2939672.2939785.
- [14] Chen H, Liu X, Jia Z, et al. *Chemometrics and Intelligent Laboratory Systems*, 2018, 182: 101.
- [15] Guevara M, Olmedo G F, Stell E, et al. *SOIL*, 2018, 4(3): 173.
- [16] Noda I. *Journal of Molecular Structure*, 2014, 1069: 3.

A New Method for Predicting Soil Moisture Based on UAV Hyperspectral Image

GE Xiang-yu^{1, 2, 3}, DING Jian-li^{1, 2, 3*}, WANG Jing-zhe⁴, SUN Hui-lan⁵, ZHU Zhi-qiang⁶

1. College of Resources & Environmental Science, Xinjiang University, Urumqi 830046, China

2. Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China

3. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi 830046, China

4. Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of Ministry of Natural Resources, Shenzhen University, Shenzhen 518060, China

5. School of Geographical Science and Tourism, Xinjiang Normal University, Urumqi 830054, China

6. China College of Material Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China

Abstract Soil moisture content (SMC) is a key factor in biogeochemical and atmospheric coupling processes. It plays an important role in areas such as agriculture, ecology and environment in arid region. Compared to the spaceborne remote sensing system, UAV platform with hyperspectral sensors possess higher spatial resolution and maneuverability. With UAV (Unmanned Aerial Vehicle) being increasingly popular, it offers brand new platform of remote sensing. This platform realizes the goal that quickly and quantitatively monitor object in the area. Moreover, hyperspectral sensors contribute to remote sensing when they enrich high dimensional and nanoscale data source. However, there still lacks a standardized research scheme for estimation of UAV by hyperspectral Remote Sensing. In this study, we obtained UAV hyperspectral image from a typically dry-farming region lying in Xinjiang Uygur Autonomous Region. Hyperspectral image was pretreated using six methods of pretreatment, including first-derivative (FDR), second-derivative (SDR), continuum removal (CR), absorbance (A), first-derivative absorbance (FDA) and second-derivative absorbance (SDA). From pretreatment foundation, four types spectral indices were proposed containing the Difference Index (DI), the Ratio Index (RI), the Normalization Index (NDI) and the Perpendicular Index (PI). And the rationality of the spectral index was discussed from the spectral mechanism. Considering the superiority of ensemble learning algorithm rising in recent years, the SMC estimation model was constructed via Gradient Boosted Regression Tree (GBRT), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost). In these models, 28 appropriate spectral indices were used as independent variables and 70 SMC measured values as response variables. Spectral indices were ranked via importance based on ensemble learning model analyzed and compared to make a more comprehensive evaluation. The result indicated that: (1) atmospheric disturbance and soil background were eliminated effectively through various pretreatment schemes and spectral indices. Pretreatment scheme A highlighted more spectral information and PI correlation was significant. (2) Optimum spectral index was A_PI ($|r|=0.773$) that the ranking of importance ranks first, and the correlation coefficient $|r|$ is the highest, and it had excellent performance in both linear and nonlinear relationships. (2) XGBoost prediction model was outstanding in three ensemble learning models, and it yielded the highest R_{val}^2 , the lowest RMSP and the best RPD ($R_{val}^2=0.926$, RMSEP=1.943 and RPD=2.556). The ranking of the predictive performance was XGBoost>RF>GBRT. This proved that this scheme was effective in digital mapping in arid region. In conclusion, there is potential high accuracy for UAV imagery based on hyperspectral imagery. This study afforded an effective method for predicting SMC in arid regions, and it provided a new perspective for quickly and easily monitoring object attributes and it proposed an alternative solution for predicting soil moisture. Ultimately, our program is supporting better management and conservation strategies for precision agriculture and ecosystems in arid regions.

Keywords UAV; Remote Sensing; Hyperspectral image; Machine learning; Ensemble learning

(Received Nov. 22, 2018; accepted Apr. 21, 2019)

* Corresponding author