

文章编号: 1672-3317(2023)01-0024-07

不同施氮处理下无人机光谱感知冬小麦产量

丁凡¹, 陈震², 李长春^{1*}, 程千², 费帅鹏², 李景勃¹, 徐洪刚², 李宗鹏³

(1.河南理工大学, 河南 焦作 454003; 2.中国农业科学院 农田灌溉研究所, 河南 新乡 453002; 3.河南农业大学, 郑州 450002)

摘要:【目的】快速、准确地预测冬小麦产量, 构建最佳产量预测模型, 对精准农业农田管理有重要应用价值。【方法】以抽穗期、开花期和灌浆期3个不同时期的冬小麦为研究对象, 通过无人机搭载的多光谱传感器采集冠层光谱信息并提取植被指数。使用逐步回归与随机森林2种方法, 筛选最优特征并建立产量预测模型。【结果】在抽穗期, 绿(Green, G)、蓝(Blue, B)、修改型土壤调查植被指数(Modified Soil-adjusted Vegetation Index 2, $MSAVI2$)和土壤调节植被指数(Soil-adjusted Vegetation Index, $SAVI$)组合使用随机森林算法对产量的预测效果最好同时 AIC (Akaike information criterion, 赤池信息准则) 较低, R^2 为 0.65。在开花期, 过绿指数(Excess Green, ExG)、近红外光(Near Infrared, NIR)、归一化差异植被指数(Normalized Difference Index, NDI)、蓝(Blue, B)和修改型土壤调查植被指数2(Modified Soil-adjusted Vegetation Index 2, $MSAVI2$)特征变量组合使用随机森林算法对产量的预测效果最好, 同时 AIC 较低, R^2 为 0.71。在灌浆期, 以全部植被特征变量构建的随机森林回归模型对冬小麦产量预测的精度最高, R^2 达到 0.76, 绿(Green, G)、蓝(Blue, B)、过绿指数(Excess Green, ExG)、过红指数(Excess Red, ExR)、归一化植被指数(Normalized Difference Vegetation Index, $NDVI$)、比值植被指数(Ratio Vegetation Index, RVI)和归一化差分植被指数(Normalized Difference Vegetation Index, $TNDVI$)特征变量组合的随机森林回归模型 R^2 较高同时 AIC 较低, R^2 达到 0.73。同时使用冬小麦3个生育期, 以抽穗期蓝(Blue, B)、开花期过绿指数(Excess Green, ExG)、灌浆期近红外光(Near Infrared, NIR)、灌浆期归一化差异植被指数(Normalized Difference Index, NDI)和灌浆期过绿减过红指数(Excess Green-Excess Red, $ExG-ExR$)为特征变量组合构建的随机森林回归模型 R^2 较高, 同时 AIC 较低, R^2 达到 0.76。【结论】通过逐步回归筛选出特征变量间共线性最小的特征变量组合并利用随机森林构建回归模型具有可行性, 能够准确地预测冬小麦产量。

关键词: 多光谱; 植被指数; 逐步回归; 随机森林; 产量预测

中图分类号: S127

文献标志码: A

doi: 10.13522/j.cnki.ggps.2022141

OSID:



丁凡, 陈震, 李长春, 等. 不同施氮处理下无人机光谱感知冬小麦产量[J]. 灌溉排水学报, 2023, 42(1): 24-30.

DING Fan, CHEN Zhen, LI Changchun, et al. Using Unmanned Aerial Vehicle to Evaluate the Effect of Nitrogen Fertilization on Winter Wheat Yield[J]. Journal of Irrigation and Drainage, 2023, 42(1): 24-30.

0 引言

【研究意义】冬小麦是我国重要的粮食作物之一, 其产量与国民经济发展和人民生活质量具有密切的关系^[1-2], 在冬小麦收获前进行产量预测对国家粮食政策制定、田间管理以及宏观调控具有重要意义。

【研究进展】随着无人机技术和轻型多光谱相机技术的发展, 低空无人机多光谱遥感技术的应用愈加广泛, 在作物生长监测中具有低成本高效率的特点^[3]。由于多光谱传感器光谱分辨率的提升, 相邻波段之间的相关性也随之增加, 因而会造成数据间产生冗余信息, 从而增加了对数据处理的复杂性以及不必要的计

算资源消耗。因此, 有必要对多光谱数据进行降维处理^[4], 筛选出对目标性状影响较大的特征。目前, 已有诸多国内外学者通过无人机平台对小麦等农作物进行产量预测。王来刚等^[5]使用随机森林算法, 构建的融合多源时空数据的冬小麦产量预测模型精度较高; Fei等^[6]基于冬小麦无人机多光谱数据, 通过集成学习方法构建的产量预测模型精度较高; 程千等^[7]采集不同水分亏缺条件下冬小麦多光谱图像, 通过偏最小二乘、支持向量机和随机森林的方法, 利用冬小麦多时相植被指数构建产量预测模型, 结果表明多种植被指数随着冬小麦的生长预测精度不断提高, 且随机森林回归模型预测效果最好; 牛庆林等^[8]将可见光与多光谱植被指数相结合, 使用逐步回归和随机森林回归方法对冬小麦叶片 $SPAD$ 值进行预测, 该方法表现出较好的预测结果。逐步回归可从大量已有变量中筛选出对模型影响较大的变量^[9]并剔除冗余变量, 进而构建预测模型。【切入点】随机森林预测模型精度高于线性

收稿日期: 2022-02-09

基金项目: 河南省高校科技创新团队支持计划项目(22IRTSTHN008); 国家自然科学基金项目(41871333)

作者简介: 丁凡(1997-), 男, 硕士研究生, 主要从事农业遥感长势监测与评估研究。E-mail: 212004020067@home.hpu.edu.cn

通信作者: 李长春(1976-), 男, 教授, 主要从事农业遥感长势监测与评估研究。E-mail: lichangchun610@126.com

预测模型^[10], 但随机森林生成的模型不直观, 无法用数学公式表达, 被认为是“黑箱”模型, 但其提供其他方式, 诸如变量的重要性用来协助解释模型, 因此 Prasad 等^[11]称随机森林为“灰箱”模型。

【拟解决的关键问题】本研究以抽穗期、开花期和灌浆期 3 个生育期的冬小麦为研究对象, 通过无人机多光谱影像技术获取 3 个时期冬小麦多光谱数据并构建多种植被指数, 随后利用逐步回归和随机森林方法构建各生育期产量预测模型, 以实现在收获前对冬小麦产量进行精准预测。

1 材料与方 法

1.1 试验区概况与试验设计

试验于 2020—2021 年在中国农业科学院新乡试验基地 (35.2°N, 113.8°E) 进行。该地为平原区域, 地处黄河和海河两大流域, 土壤肥沃且光照充足, 属暖温带大陆性季风气候, 年平均气温为 14 °C, 年平均降水量 573.4 mm, 适宜冬小麦生长。表层土壤质地为轻质壤土, 地下水埋深超过 10 m。采用大型平移式喷灌机变量喷洒水肥, 设置 6 个不同的施氮(N) 肥处理, 氮肥选用尿素, 施肥水平分别为 N0 (不施加氮肥)、N1 (60 kg/hm²)、N2 (120 kg/hm²)、N3 (180 kg/hm²)、N4 (240 kg/hm² 农民习惯用量) 和 N5 (300 kg/hm², 过量施氮), 每 30 个小区 1 个处理, 共 180 个小区, 试验小区布置如图 1 所示。施氮肥在拔节期和抽穗期 2 个时期进行, 各施肥量按 2:1 划分, 施加氮肥的质量浓度为 0.3%。小区规格 3 m×1.4 m, 相邻小区左右间隔 0.2 m, 前后间隔 1 m。于 2020 年 10 月初按试验小区播种, 2021 年 6 月初按试验小区收获, 自然晒干称质量得到各小区实测产量。

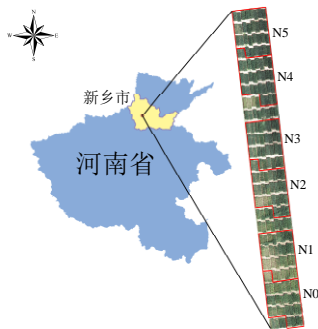


图 1 试验小区布置图

Fig.1 Layout of test area

1.2 无人机影像获取及处理

本试验搭载平台采用大疆经纬 M210 型四旋翼无人机, 无人机质量为 4.69 kg, 最大承载质量 1.45 kg, 续航时间约 20 min。无人机搭载 Red Edge MX 多光谱相机于抽穗期 (2021 年 4 月 17 日)、开花期 (2021 年 4 月 28 日) 和灌浆期 (2021 年 5 月 8 日) 对冬小

麦冠层拍摄多光谱影像, 使用大疆 GSPRO 地面站规划航线, 多光谱相机设置为与地面垂直, 等时间间隔拍照模式。飞行高度为 30 m, 航向重叠率 85%, 旁向重叠率 80%, 影像空间分辨率为 2 mm。光谱影像具有 5 个通道, 分别为红 (Red, *R*)、绿 (Green, *G*)、蓝 (Blue, *B*)、近红外 (Near infrared, *NIR*) 和红边 (Red edge, *RE*), 波段信息如表 1 所示。

表 1 Red Edge MX 型多光谱相机波段信息

Table 1 Red Edge MX multispectral camera band information

波段名称	中心波长/nm	光谱带宽/nm
<i>R</i>	668	10
<i>G</i>	560	20
<i>B</i>	475	20
<i>NIR</i>	840	40
<i>RE</i>	717	10

在无人机飞行作业后利用 Pix4Dmapper 软件对获取的影像进行校正拼接。通过 ArcGIS10.2 计算多光谱植被指数并绘制感兴趣区提取小区植被指数均值。

1.3 植被指数选取

植被指数是由光谱中多个波段组合而成, 在对地面植被状况进行度量方面比单波段具有更好的灵敏性。本试验选取植被提取颜色指数 (Color Index of Vegetation Extraction, *CIVE*)、差异植被指数 (Difference vegetation index, *DVI*)、过绿指数 (Excess Green, *ExG*)、过绿减过红指数 (Excess Green-Excess Red, *ExG-ExR*)、过红指数 (Excess Red, *ExR*)、修改型土壤调查植被指数 (Modified Soil-adjusted Vegetation Index 2, *MSAVI2*)、归一化差异植被指数 (Normalized Difference Index, *NDI*)、归一化植被指数 (Normalized Difference Vegetation Index, *NDVI*)、重归一化植被指数 (Re-normalized Vegetation Index, *RDVI*)、比值植被指数 (Ratio Vegetation Index, *RVI*)、土壤调节植被指数 (Soil-adjusted Vegetation Index, *SAVI*)、归一化差分植被指数 (Normalized Difference Vegetation Index, *TNDVI*) 共 12 个植被指数, 如表 2 所示。

表 2 植被指数及计算式

Table 2 Vegetation index and calculation formula

植被指数	公式	参考文献
<i>CIVE</i>	$0.441R-0.811G+0.385B+18.78745$	[12]
<i>DVI</i>	$NIR-R$	[7]
<i>ExG</i>	$2G-R-B$	[13]
<i>ExR</i>	$1.4R-G$	[14]
<i>ExG-ExR</i>	$ExG-ExR$	[14]
<i>MSAVI2</i>	$\{2NIR+1-[(2NIR+1)^2-8(NIR-R)]^{0.5}\}/2$	[15]
<i>NDI</i>	$(G-R)/(G+R)$	[16]
<i>NDVI</i>	$(NIR-R)/(NIR+R)$	[17]
<i>RDVI</i>	$(NIR-R)/(NIR+R)^{0.5}$	[18]
<i>RVI</i>	R/NIR	[19]
<i>SAVI</i>	$1.5(NIR-R)/(NIR+R+0.5)$	[20]
<i>TNDVI</i>	$[(NIR-R)/(NIR+R)+0.5]^{0.5}$	[21]

1.4 分析方法

本研究通过逐步回归^[22]算法构建的模型筛选特征间影响较小的特征向量组合,再利用随机森林算法^[23]构建预测模型。逐步回归是一种对线性回归模型自变量选择的方法,其基本思想是逐个将特征变量引入回归模型中,在每引入1个特征变量时对模型进行1次检验,若引入的特征变量使模型内其他特征变量的解释显著性降低,则剔除该特征变量,重复此过程,直至不再有对模型贡献大的特征变量引入模型,也没有对模型贡献小的特征变量从模型中被剔除,此时模型为最优模型^[24]。对模型特征选择的标准选用赤池信息准则(Akaike information criterion, *AIC*)^[8], *AIC* 值越小,表示模型内各个特征变量的解释性就越好,每个特征变量之间对彼此重要性的影响程度越小。选用 *AIC* 值最小的模型既为最优模型。随机森林的思想是集成学习,是机器学习的一大分支,由多棵树组成,其基本单元是决策树^[25]。其在训练过程中,从训练集中有放回地随机抽取样本,重复多次,生成多棵决策树,对每棵决策树通过传递随机变量来预测结果,从而得到多种结果,对这些结果计算平均值便得到随机森林回归的最终预测结果。随机森林通过平均决策树的预测结果降低过拟合的风险,并且受噪声影响较小^[26]。

1.5 模型评价指标

随机选取数据的 3/4 作为训练集,剩余 1/4 作为验证集。以验证集决定系数 (R^2)、均方根误差 (*RMSE*)、归一化均方根误差 (*nRMSE*) 和 *AIC* 作为预测模型的精度评价标准。 R^2 越大表明植被指数与产量的拟合性越好;*RMSE* 和 *nRMSE* 越小表明模型预测效果越好;*AIC* 越小表明模型特征最少同时可以最好的解释数据。

2 结果与分析

2.1 植被指数与产量相关性

使用 3 个生育期的 12 个植被指数与产量进行相关性分析。使用 *R*、*G*、*B*、*NIR*、*RE* 和 12 个光谱指数分别建立 3 个生育期产量的随机森林回归模型,详见表 3。

由表 3 可看出,3 个生育期中,灌浆期的大部分植被指数与产量的相关性最高,抽穗期的大部分植被指数与产量的相关性最低。*ExG-ExR*、*ExG*、*NDI* 与产量的相关性较低,除抽穗期的 *ExG-ExR* 与产量无显著相关外,相关系数范围为 0.22~0.45,其余 9 个植被指数在 3 个生育期中与产量的相关性均大于 0.5。在抽穗期和开花期,*ExG* 的相关性均最高,绝对值分别为 0.57 和 0.67;在灌浆期,*MSAVI2* 的相关性最高,

绝对值为 0.59,此外 *DVI*、*RDVI*、*SAVI* 的相关系数均达到 0.58。由此可见,本文选用的 12 个植被指数与产量均具有较强的相关性,可以用来对冬小麦的产量进行估测。

表 3 不同生育期植被指数与产量相关性分析

Table 3 Correlation analysis between vegetation

index and yield in different growth stages

植被指数	抽穗期	开花期	灌浆期
<i>CIVE</i>	0.52	0.65	0.52
<i>DVI</i>	0.54	0.57	0.58
<i>ExG</i>	0.57	0.67	0.54
<i>ExG-ExR</i>	0.05	0.32	0.22
<i>ExR</i>	0.4	0.3	0.32
<i>MSAVI2</i>	0.52	0.57	0.59
<i>NDI</i>	0.45	0.37	0.32
<i>NDVI</i>	0.52	0.53	0.56
<i>RDVI</i>	0.54	0.57	0.58
<i>RVI</i>	0.51	0.52	0.55
<i>SAVI</i>	0.54	0.57	0.58
<i>TNDVI</i>	0.52	0.53	0.55

2.2 基于逐步回归分析方法预测产量

将 *R*、*G*、*B*、*NIR*、*RE* 和各植被指数作为输入变量,以产量为输出变量建立逐步回归模型,结果见表 4。在 3 个生育期, R^2 范围分别在 0.351~0.412、0.436~0.538、0.439~0.486, *RMSE* 值范围分别在 0.569~0.584、0.488~0.588、0.593~0.624 t/hm², *nRMSE* 值范围分别在 16.324%~17.305%、14.396%~17.145%、17.365%~17.993%。综上,开花期的模型精度在包含 5 个输入变量 (*G*、*B*、*ExG*、*RVI*、*ExR*) 时最佳, R^2 为 0.526, *RMSE* 值为 0.488 t/hm², *nRMSE* 值为 14.396%, *AIC* 值为 288.152。

2.3 基于逐步回归筛选变量后的随机森林回归模型精度

通过逐步回归筛选出每个生育期 *AIC* 值最小的 5 个预测模型后,根据每个模型的输入特征变量构建对应的随机森林回归模型,并计算各个预测模型的 R^2 、*RMSE* 和 *nRMSE*,如表 5 所示。

为了筛选最优预测模型,对各个预测模型进行综合评定,以预测模型的 R^2 、*RMSE*、*nRMSE* 和 *AIC* 为评定指标。在 3 个生育期, R^2 范围分别在 0.501~0.669、0.638~0.708、0.629~0.755, *RMSE* 值范围分别在 0.457~0.552、0.453~0.520、0.425~0.522 t/hm², *nRMSE* 值范围分别在 13.053%~15.820%、12.877%~14.584%、12.093%~14.656%。其中,在灌浆期预测模型的综合评定精度最好,精度最好的 2 个预测模型分别包含全部特征变量和 7 个输入特征变量 (*G*、*B*、*ExG*、*ExR*、*NDVI*、*RVI*、*TNDVI*), R^2 分别为 0.755 和 0.734, *RMSE*、*nRMSE* 分别为 0.425 t/hm²、12.093% 和 0.439 t/hm²、12.441%, *AIC* 值分别为 306.180 和 295.005。在开花

期,精度最好的2个预测模型分别包含5个输入特征变量 (*ExG*、*NIR*、*NDI*、*BLUE*、*MSAVI2*)和2个输入特征变量 (*ExG*、*NIR*), R^2 分别为0.708和0.685, $RMSE$ 和 $nRMSE$ 分别为0.471 t/hm²、13.288%和0.453 t/hm²、12.877%, AIC 值分别为293.319和299.338。在抽穗期,精度最好的2个预测模型分别包含4个输入特征变量 (*G*、*B*、*MSAVI2*、*SAVI*)和全部特征变量, R^2 分别为0.645和0.669, $RMSE$ 、 $nRMSE$ 分别为0.496 t/hm²、14.263%和0.481 t/hm²、13.743%, AIC 值分别为315.303和323.129。

表4 冬小麦不同生育期逐步回归线性预测模型

Table 4 Stepwise regression linear prediction model for different growth stages of winter wheat

生育期	植被指数	AIC	训练集			测试集		
			R^2	$RMSE/(t \cdot hm^{-2})$	$nRMSE/\%$	R^2	$RMSE/(t \cdot hm^{-2})$	$nRMSE/\%$
抽穗期	全部特征变量	323.129	0.475	0.558	15.490	0.351	0.570	16.324
	<i>RE</i> 、 <i>NIR</i> 、 <i>NDI</i>	312.776	0.457	0.533	15.598	0.407	0.580	17.205
	<i>ExGExR</i> 、 <i>ExR</i> 、 <i>NDI</i> 、 <i>NDVI</i> 、 <i>RVI</i> 、 <i>TNDVI</i>	312.039	0.478	0.531	15.413	0.412	0.569	16.858
	<i>G</i> 、 <i>B</i> 、 <i>MSAVI2</i> 、 <i>SAVI</i>	315.303	0.464	0.543	15.838	0.385	0.576	17.082
	<i>ExG</i> 、 <i>DVI</i>	315.641	0.450	0.548	16.024	0.375	0.584	17.305
	<i>RE</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>ExG</i>	314.189	0.466	0.533	15.484	0.408	0.575	17.060
开花期	全部特征变量	298.359	0.549	0.488	14.288	0.538	0.573	16.459
	<i>ExG</i> 、 <i>NIR</i>	299.338	0.487	0.548	15.934	0.465	0.553	16.002
	<i>R</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>ExG-ExR</i> 、 <i>NDI</i> 、 <i>RDVI</i>	287.548	0.539	0.500	14.538	0.533	0.562	16.348
	<i>G</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>RVI</i> 、 <i>ExR</i>	288.152	0.535	0.536	15.669	0.526	0.488	14.396
	<i>R</i> 、 <i>G</i> 、 <i>NIR</i>	296.430	0.474	0.562	16.523	0.436	0.552	16.242
	<i>ExG</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>MSAVI2</i>	293.319	0.528	0.506	14.704	0.488	0.588	17.145
灌浆期	全部特征变量	306.180	0.526	0.500	14.650	0.486	0.604	17.365
	<i>G</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i>	301.397	0.491	0.525	15.268	0.479	0.593	17.513
	<i>G</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>ExR</i> 、 <i>NDVI</i> 、 <i>RVI</i> 、 <i>TNDVI</i>	295.005	0.534	0.503	14.656	0.477	0.595	17.544
	<i>B</i> 、 <i>NIR</i> 、 <i>NDI</i>	299.478	0.491	0.525	15.268	0.479	0.593	17.513
	<i>RDVI</i> 、 <i>NDI</i> 、 <i>B</i>	299.627	0.501	0.520	15.121	0.454	0.607	17.993
	<i>G</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>SAVI</i>	301.994	0.495	0.519	15.053	0.439	0.624	17.780

表5 冬小麦不同生育期筛选变量后的随机森林回归模型

Table 5 Prediction model combining stepwise regression and random forest in different growth stages of winter wheat

生育期	植被指数	AIC	训练集			测试集		
			R^2	$RMSE/(t \cdot hm^{-2})$	$nRMSE/\%$	R^2	$RMSE/(t \cdot hm^{-2})$	$nRMSE/\%$
抽穗期	全部特征变量	323.129	0.886	0.253	7.377	0.669	0.481	13.743
	<i>RE</i> 、 <i>NIR</i> 、 <i>NDI</i>	312.776	0.892	0.255	7.586	0.587	0.462	13.239
	<i>ExGExR</i> 、 <i>ExR</i> 、 <i>NDI</i> 、 <i>NDVI</i> 、 <i>RVI</i> 、 <i>TNDVI</i>	312.039	0.888	0.250	7.280	0.592	0.552	15.820
	<i>G</i> 、 <i>B</i> 、 <i>MSAVI2</i> 、 <i>SAVI</i>	315.303	0.891	0.247	7.186	0.645	0.496	14.263
	<i>ExG</i> 、 <i>DVI</i>	315.641	0.863	0.288	8.536	0.501	0.508	14.625
	<i>RE</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>ExG</i>	314.189	0.892	0.255	7.604	0.612	0.457	13.053
开花期	全部特征变量	298.359	0.899	0.238	6.916	0.638	0.520	14.584
	<i>ExG</i> 、 <i>NIR</i>	299.338	0.901	0.235	6.834	0.685	0.453	12.877
	<i>R</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>ExG-ExR</i> 、 <i>NDI</i> 、 <i>RDVI</i>	287.548	0.905	0.230	6.711	0.663	0.490	13.865
	<i>G</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>RVI</i> 、 <i>ExR</i>	288.152	0.883	0.256	7.429	0.668	0.491	13.838
	<i>R</i> 、 <i>G</i> 、 <i>NIR</i>	296.430	0.893	0.244	7.087	0.659	0.491	13.794
	<i>ExG</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>MSAVI2</i>	293.319	0.903	0.232	6.770	0.708	0.471	13.288
灌浆期	全部特征变量	306.180	0.886	0.252	7.315	0.755	0.425	12.093
	<i>G</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i>	301.397	0.874	0.265	7.755	0.690	0.461	13.038
	<i>G</i> 、 <i>B</i> 、 <i>ExG</i> 、 <i>ExR</i> 、 <i>NDVI</i> 、 <i>RVI</i> 、 <i>TNDVI</i>	295.005	0.872	0.267	7.762	0.734	0.439	12.441
	<i>B</i> 、 <i>NIR</i> 、 <i>NDI</i>	299.478	0.882	0.257	7.490	0.642	0.507	14.450
	<i>RDVI</i> 、 <i>NDI</i> 、 <i>B</i>	299.627	0.885	0.254	7.381	0.629	0.522	14.656
	<i>G</i> 、 <i>NIR</i> 、 <i>NDI</i> 、 <i>B</i> 、 <i>SAVI</i>	301.994	0.886	0.252	7.342	0.665	0.484	13.552

同时使用3个生育期的共51个输入特征变量筛选出 AIC 值较小的模型,并构建随机森林回归模型,如表6所示。使用_1、_2和_3后缀以区分不同时期的植被指数,分别对应抽穗期、开花期和灌浆期。

由表6可看出,精度最好的预测模型包含5个输

入特征 (B_1 、 EXG_1 、 NIR_3 、 NDI_3 、 $EXGEXR_3$), R^2 达到0.76, $RMSE$ 为0.402 t/hm², $nRMSE$ 为11.488%,与只考虑单生育期所构建的预测模型3个评价指标相比,均表现出较好的精度。

表 6 冬小麦 3 个生育期筛选变量后的随机森林回归模型

Table 6 Random forest regression model after selecting variables at three growth stages of winter wheat

植被指数	AIC	R ²	RMSE/(t·hm ⁻²)	nRMSE/%
全部特征变量	296.671	0.713	0.445	12.537
<i>B</i> ₁ 、 <i>CIVE</i> ₂ 、 <i>NIR</i> ₃ 、 <i>NDI</i> ₃	285.437	0.756	0.425	12.139
<i>B</i> ₁ 、 <i>EXG</i> ₁ 、 <i>NIR</i> ₃ 、 <i>NDI</i> ₃ 、 <i>EXGEXR</i> ₃	285.081	0.76	0.402	11.488
<i>R</i> ₁ 、 <i>EXGEXR</i> ₂ 、 <i>NIR</i> ₃ 、 <i>NDI</i> ₃	283.749	0.683	0.468	13.267
<i>R</i> ₁ 、 <i>G</i> ₁ 、 <i>EXG</i> ₁ 等 21 个植被指数	280.388	0.74	0.479	13.408

图 2 为不同氮肥处理下冬小麦产量预测值与实测值对比,图 2 中估测值分别选取每个时期估测效果最优的训练模型。抽穗期选用 *G*、*B*、*MSAVI2*、*SAVI* 多特征变量组合模型,开花期选用 *ExG*、*NIR*、*NDI*、*B*、*MSAVI2* 多特征变量组合模型,灌浆期选用 *G*、*B*、*ExG*、*ExR*、*NDVI*、*RVI*、*TNDVI* 多特征变量组合模型。从图 2 可以看出, N4 处理冬小麦产量最高, 平均达到 9.374 t/hm², 随着施氮量的减少, 产量呈下降

趋势,但在 N1 处理时略有提高。N0 处理冬小麦平均产量为 5.964 t/hm², 比 N4 处理降低 36.377%。冬小麦产量预测值与实测值相对比, N1 处理和 N4 处理预测值偏小, 其余施氮处理时, 预测值偏大。N5 处理时预测值的分布范围与实测值的接近, 其余处理下预测值的分布范围均比实测值的小。由于试验所用冬小麦品种较多, 不同品种产量不同, 因而产量数据分布分散, 尤其氮营养亏缺严重时更为明显。

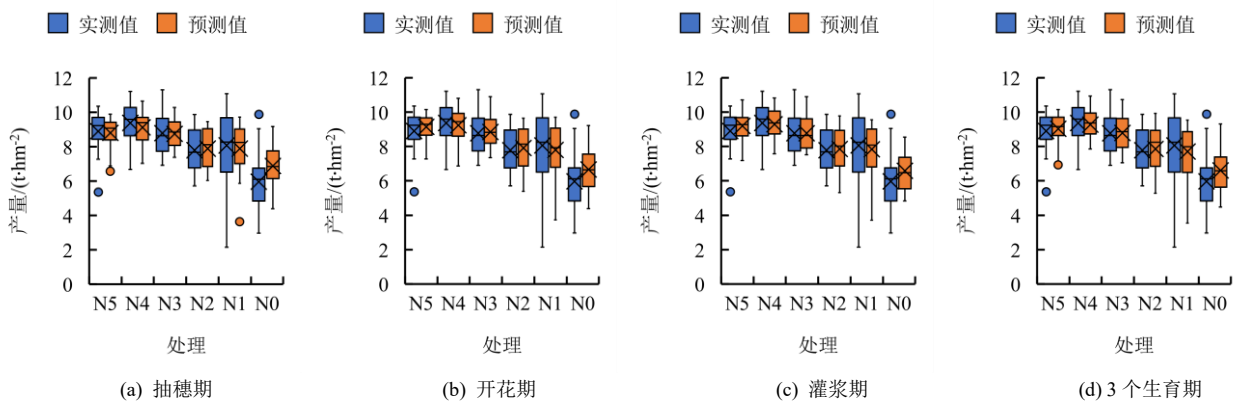


图 2 不同氮肥处理下产量实测值与预测值对比

Fig.2 The measured and estimated yield values were compared under different nitrogen treatments

3 讨论

使用逐步回归筛选出的特征变量作为输入的线性产量回归模型精度较差,但可以对植被特征进行降维,筛选出对模型影响较大的特征组合^[4]。随机森林较传统的回归算法精度和稳定性更高。本研究表明,随机森林应用在作物表型评估中能够提升模型性能,王丽爱等^[27]以 8 个植被指数为输入特征变量分析多个生育期小麦叶片 *SPAD* 值与植被指数间的相关性,构建随机森林回归模型,并以支持向量回归模型和反向传播神经网络回归模型为对照模型,研究表明随机森林回归模型的预测精度最高。王庆等^[28]通过甜菜叶丛冠层的结构特征和光谱特征构建地上部和块根鲜质量和块根含糖率随机森林和偏最小二乘预测模型,研究显示随机森林回归模型在精度上高于偏最小二乘回归模型。基于此,本研究先通过逐步回归筛选特征变量,再使用随机森林算法,在特征变量间相互影响较小的前提下,得到精度较高的随机森林回归模型,并在此前提下,通过对预测模型的分析,能够大致得出哪些特征变量对预测模型更重要且对精度影响更

高,进而从这些特征变量中寻找出与产量之间的联系,探究该特征变量对产量预测精度解释性较强的原因。

在 *R*、*G*、*B*、*NIR* 和 *RE* 这 5 个波段中 *G* 和 *RE* 的建模精度较高, *NIR* 的建模精度较低。原因是绿波段的反射率对叶片叶绿素量敏感,反映了叶片叶绿素量水平,叶绿素量植株光合作用的重要影响因素,在很大程度上决定了作物产量^[29]。*RE* 在作物快成熟时既在灌浆期左右会出现红边位移^[30]现象,作物生长状态好,红边位置会出现红移,反之会出现蓝移,并且红边反射率曲线斜率大,导致红边与产量具有较大的相关性。同时这一结果印证了一些研究者提出红边与叶片叶绿素即绿波段有关系的观点^[30-31]。*RE* 在灌浆期的预测模型精度较高的原因是红边位移现象主要发生在该时期。

开花期和灌浆期冬小麦的营养生长已基本完成,产量也基本定型,叶片对土地覆盖程度高,裸土较少,此时冬小麦的性状与产量相关性较高。开花期之前的生育期受外界因素影响较大,不是对产量进行预测的最佳时期。成熟期冬小麦开始变黄,茎秆叶基本变干,叶片枯萎叶面积减少,裸土增多,因此不建议作为观

测时期。

在逐步回归以及随机森林回归所构建的预测模型中, 灌浆期表现出较高的预测精度。灌浆期时, 冬小麦籽粒长度先达到最大, 随后宽度和厚度明显增加, 是决定粒质量和产量的关键时期^[32], 这一生理特征决定了该时期对产量预测的精准性。

本研究仅使用了多光谱 5 个波段, 没有对多光谱波段以外对产量敏感的指数进行研究, 下一步可以考虑引入高光谱指数对产量的预测模型进行优化, 以期达到更高的预测精度。

4 结论

1) 开花期和灌浆期的植被指数与产量均具有较强的相关性, 所构建的随机森林预测模型最大 R^2 均大于 0.7。

2) 灌浆期的植被指数对产量的预测效果最好, 构建的最优预测模型的 R^2 达到 0.73, 输入特征变量为 G 、 B 、 ExG 、 ExR 、 $NDVI$ 、 RVI 和 $TNDVI$ 组合, 同时 AIC 对比同时期其他特征变量组合较低。

3) 同时考虑 3 个生育期的指标相比单个生育期所构建的随机森林预测模型精度有所提升, 保证 AIC 较低的同时 R^2 可达到 0.76。

4) 本研究中, 过量施氮以及氮肥亏缺均导致冬小麦产量降低。与 $N4$ 处理相比, $N0$ 处理将造成冬小麦产量下降约 1/3。

参考文献:

- FENG P, WANG B, LIU D L, et al. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique[J]. *Agricultural and Forest Meteorology*, 2020, 285: 107-122.
- FENG L W, WANG Y M, ZHANG Z, et al. Geographically and temporally weighted neural network for winter wheat yield prediction[J]. *Remote Sensing of Environment*, 2021, 262: 112-124.
- 孙刚, 黄文江, 陈鹏飞, 等. 轻小型无人机多光谱遥感技术应用进展[J]. *农业机械学报*, 2018, 49(3): 1-17.
SUN Gang, HUANG Wenjiang, CHEN Pengfei, et al. Advances in UAV-based multispectral remote sensing applications[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(3): 1-17.
- KUMAR S, ATTRI S D, SINGH K K. Comparison of Lasso and stepwise regression technique for wheat yield prediction[J]. *Journal of Agrometeorology*, 2019, 21(2): 188-192.
- 王来刚, 郑国清, 郭燕, 等. 融合多源时空数据的冬小麦产量预测模型研究[J]. *农业机械学报*, 2022, 53(1): 198-204, 458.
WANG Laigang, ZHENG Guoqing, GUO Yan, et al. Prediction of winter wheat yield based on fusing multi-source spatio-temporal data[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(1): 198-204, 458.
- FEI S P, ADEEL H M, HE Z H, et al. Assessment of ensemble learning to predict wheat grain yield based on UAV-multispectral Reflectance[J]. *Remote Sensing*, 2021, 13(12): 2 338.
- 程千, 徐洪刚, 曹引波, 等. 基于无人机多时相植被指数的冬小麦产量估测[J]. *农业机械学报*, 2021, 52(3): 160-167.
CHENG Qian, XU Honggang, CAO Yinbo, et al. Grain yield prediction of winter wheat using multi-temporal UAV based on multispectral vegetation index[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(3): 160-167.
- 牛庆林, 冯海宽, 周新国, 等. 冬小麦 SPAD 值无人机可见光和多光谱植被指数结合估算[J]. *农业机械学报*, 2021, 52(8): 183-194.
NIU Qinglin, FENG Haikuan, ZHOU Xinguo, et al. Combining UAV visible light and multispectral vegetation indices for estimating SPAD value of winter wheat[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(8): 183-194.
- GUIMARAES B V C, DONATO S L R, ASPIAZU I, et al. Regression models for productivity prediction in cactus pear cv. Gigante[J]. *Revista Brasileira De Engenharia Agricola E Ambiental*, 2020, 24(11): 721-727.
- LI X, LV X, HE Y F, et al. Application of random forest in identifying winter wheat using Landsat8 imagery[J]. *Engenharia Agricola*, 2021, 41(6): 619-633.
- PRASAD A M, IVERSON L R, LIAW A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction[J]. *Ecosystems*, 2006, 9(2): 181-199.
- 李长春, 牛庆林, 杨贵军, 等. 基于无人机数码影像的大豆育种材料叶面积指数估测[J]. *农业机械学报*, 2017, 48(8): 147-158.
LI Changchun, NIU Qinglin, YANG Guijun, et al. Estimation of leaf area index of soybean breeding materials based on UAV Digital Images[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(8): 147-158.
- CHOLULA U, SILVA J A D, MARCONI T, et al. Forecasting yield and lignocellulosic composition of energy cane using unmanned aerial systems[J]. *Agronomy*, 2020, 10(5): 718.
- ZHAO Y. The Segmentation of plants on RGB images with index based color analysis[C]. *Wuhan: 2021 5th International Conference on Robotics and Automation Sciences (ICRAS)*, 2021: 221-225.
- PISMAN T I, ERUNOVA M G, BOTVICH I Y, et al. Informative value of spectral vegetation indices for assessment of weediness of agricultural crops according to ground-based and satellite data[J]. *Issledovanie Zemli iz Kosmosa*, 2021(3): 55-66.
- 蔡文婷, 赵书河, 王亚梅, 等. 结合 Sentinel-2 光谱与纹理信息的冬小麦作物覆盖度估算[J]. *遥感学报*, 2020, 24(9): 1 108-1 119.
CAI Wenting, ZHAO Shuhe, WANG Yamei, et al. Estimation of winter wheat residue cover using spectral and textural information from Sentinel-2 data[J]. *Journal of Remote Sensing*, 2020, 24(9): 1 108-1 119.
- DARIUS J, DAINIUS S, AURELIJA K, et al. Comparative analysis of the environmental impact of conventional and precision spring wheat fertilization under various meteorological conditions[J]. *Journal of Environmental Management*, 2021, 296: 113-150.
- IHUOMA S O, MADRAMOOTOO C A. Narrow-band reflectance indices for mapping the combined effects of water and nitrogen stress in field grown tomato crops[J]. *Biosystems Engineering*, 2020, 192: 133-143.
- YANG T C, LU J S, LIAO F, et al. Retrieving potassium levels in wheat blades using normalised spectra[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 102: 102-112.
- ATTILA N, ANDREA S, DAVID A O, et al. Wheat yield forecasting for the tisa river catchment using Landsat 8 NDVI and SAVI time series and reported crop statistics[J]. *Agronomy*, 2021, 11(4): 652.
- LAOSUWAN T, UTTARUK Y. Carbon sequestration assessment of the orchards using satellite data[J]. *Journal of Ecological Engineering*, 2017, 18(1): 11-17.
- WANG G Q, ZHANG X J, YINGLAN A, et al. A spatio-temporal cross comparison framework for the accuracies of remotely sensed soil moisture products in a climate-sensitive grassland region[J]. *Journal of Hydrology*, 2021, 597: 126-138.

- [23] RIBEIRO M N, CARVALHO I A, FONSECA G A, et al. Quality control of fresh strawberries by a random forest model. [J]. Journal of the Science of Food and Agriculture, 2021, 101(11): 4 514-4 522.
- [24] 李长春, 施锦锦, 马春艳, 等. 基于小波变换和分数阶微分的冬小麦叶绿素含量估算[J]. 农业机械学报, 2021, 52(8): 172-182.
LI Changchun, SHI Jinjin, MA Chunyan, et al. Estimation of chlorophyll content in winter wheat based on wavelet transform and fractional differential[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8): 172-182.
- [25] NUR H L, TOMOYA W, TSUTOMU M, et al. Machine learning techniques to predict soybean plant density using UAV and satellite-based remote sensing[J]. Remote Sensing, 2021, 13(13): 2 548.
- [26] 李长春, 陈伟男, 王宇, 等. 基于多源 Sentinel 数据的县域冬小麦种植面积提取[J]. 农业机械学报, 2021, 52(12): 207-215.
LI Changchun, CHEN Weinan, WANG Yu, et al. Extraction of winter wheat planting area in county based on multi-sensor sentinel data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(12): 207-215.
- [27] 王丽爱, 马昌, 周旭东, 等. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算[J]. 农业机械学报, 2015, 46(1): 259-265.
WANG Liai, MA Chang, ZHOU Xudong, et al. Estimation of wheat leaf SPAD value using RF algorithmic model and remote sensing data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 259-265.
- [28] 王庆, 车荧璞, 柴宏红, 等. 基于无人机影像的冠层光谱和结构特征监测甜菜长势[J]. 农业工程学报, 2021, 37(20): 90-98.
WANG Qing, CHE Yingpu, CHAI Honghong, et al. Monitoring of sugar beet growth using canopy spectrum and structural characteristics with UAV images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(20): 90-98.
- [29] ZHAO R M, AN L L, SONG D, et al. Detection of chlorophyll fluorescence parameters of potato leaves based on continuous wavelet transform and spectral analysis[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 259: 119 768.
- [30] QIAN B X, HUANG W J, YE H C, et al. Inversion of winter wheat chlorophyll contents based on improved algorithms for red edge position[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(23): 162-170.
- [31] HORLER D N H, DOCKRAY M, BARBER J. The red edge of plant leaf reflectance[J]. International Journal of Remote Sensing, 1983, 4(2): 273-288.
- [32] 吴启侠, 谭京红, 朱建强, 等. 花后渍水对不同耐渍型冬小麦籽粒灌浆特性的影响[J]. 农业工程学报, 2021, 37(18): 74-81.
WU Qixia, TAN Jinghong, ZHU Jianqiang, et al. Effects of waterlogging after anthesis on the grain filling characteristics of winter wheat with different waterlogging tolerances[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(18): 74-81.

Using Unmanned Aerial Vehicle to Evaluate the Effect of Nitrogen Fertilization on Winter Wheat Yield

DING Fan¹, CHEN Zhen², LI Changchun^{1*}, CHENG Qian²,
FEI Shuaipeng², LI Jingbo¹, XU Honggang², LI Zongpeng³

(1. Henan Polytechnic University, Jiaozuo 454003, China; 2. Farmland Irrigation Research Institute, Chinese Academy of Agricultural Sciences, Xinxiang 453002, China; 3. Henan Agricultural University, Zhengzhou 450002, China)

Abstract: **【Objective】** Precision agriculture requires a quick and accurate evaluation of the response of yield to managements at different growing stages, which is challenging at the large scales because of the heterogeneity of plants and soils. Remote sensing technologies can plug this gap, and the purpose of this paper is to investigate the feasibility of using unmanned aerial vehicle (UAV) to evaluate the variation in winter wheat yield in response to change in nitrogen fertilization. **【Method】** Canopy spectral information of winter wheat under different nitrogen fertilizations was measured using a multispectral sensor mounted on a UAV at heading, flowering and filling stages, from which we extracted the vegetation indexes. Stepwise regression and random forest models were used to screen the optimal indexes for estimating the wheat yield. **【Result】** Combination of green (*G*), blue (*B*), modified soil-adjusted vegetation index 2 (*MSAVI2*) and soil-adjusted vegetation index (*SAVI*) at heading stage works best to predict the yield with low *AIC* (Akaike Information Criterion) and $R^2=0.65$. Combination of excess green (*ExG*), near infrared (*NIR*), normalized difference index (*NDI*), *B* and modified soil-adjusted vegetation index 2 (*MSAVI2*) at flowering stage is the best when using the random forest algorithm to predict the yield with low *AIC* and $R^2=0.71$. The random forest regression model using all vegetation features at filling stages gave the most accurate prediction of the yield with $R^2=0.76$, compared with $R^2=0.76$ when using *G*, *B*, *ExG*, excess red (*ExR*), normalized difference vegetation index (*NDVI*), ratio vegetation index (*RVI*) and normalized difference vegetation index (*TNDVI*) only. Prediction using *B* at heading stage, *ExG* at flowering stage, *NIR* at filling stage, *NDI* at filling stage, and excess green-excess red (*EXG-EXR*) at filling stage predicted a yield with $R^2=0.76$. **【Conclusion】** Combination of the vegetation indexes measured using multispectral sensors at different growing stages can predict the ultimate winter wheat yield, but the accuracy varies. Best results are achievable using a combination of different indexes measured at different growing stages.

Key words: multispectral; vegetation index; stepwise regression; random forest; yield prediction

责任编辑: 白芳芳