

基于机器学习和无人机多光谱遥感的苜蓿产量预测

严海军^{1,2}, 卓越¹, 李茂娜³, 王云玲¹, 郭辉¹,
王晶晶¹, 李长硕¹, 丁峰^{1,4}

(1. 中国农业大学水利与土木工程学院, 北京 100083; 2. 农业节水与水资源教育部工程研究中心, 北京 100083;
3. 中国农业大学草业科学与技术学院, 北京 100193; 4. 新疆农业科学院土壤肥料与农业节水研究所, 乌鲁木齐 830091)

摘要: 作物产量快速准确预估对规模化农业生产具有重要意义。该研究在河北省涿州市开展田间试验, 设置 5 个灌水处理及雨养对照区, 使用无人机搭载多光谱相机在苜蓿的分枝期、现蕾期和初花期进行遥感监测, 将光谱参数与苜蓿产量做相关性分析, 分别挑选各生育时期相关性较好的 5 种植被指数, 以去除土壤噪声后的 5 种植被指数和作物表面模型提取的苜蓿株高为输入变量, 通过支持向量回归算法建立各生育时期的苜蓿产量预测模型, 并结合实测产量进行模型评价。结果表明, 苜蓿产量预测模型精度由高到低的生育时期依次为初花期、现蕾期、分枝期, 使用植被指数和株高组合作为输入变量可以提高产量模型的预测精度。在初花期使用植被指数和株高组合输入变量所构建的估产模型最优, 其决定系数、均方根误差和标准均方根误差分别为 0.90、500 kg/hm² 和 14.3%, 可用于苜蓿产量的快速准确评估。研究结果可为苜蓿的规模化生产和精细化管理提供技术支持。

关键词: 无人机; 产量; 植被指数; 苜蓿; 支持向量回归; 生育时期

doi: 10.11975/j.issn.1002-6819.2022.11.007

中图分类号: S127

文献标志码: A

文章编号: 1002-6819(2022)-11-0064-08

严海军, 卓越, 李茂娜, 等. 基于机器学习和无人机多光谱遥感的苜蓿产量预测[J]. 农业工程学报, 2022, 38(11): 64-71.

doi: 10.11975/j.issn.1002-6819.2022.11.007 <http://www.tcsae.org>

Yan Haijun, Zhuo Yue, Li Maona, et al. Alfalfa yield prediction using machine learning and UAV multispectral remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 64-71. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.11.007 <http://www.tcsae.org>

0 引言

苜蓿是国内外广泛种植的重要饲草作物^[1], 具有产草量高、草质优良、营养丰富、易于消化等特点, 被誉为“牧草之王”^[2]。华北地区苜蓿种植面积达 110 万 hm², 占全国苜蓿种植总面积的 22%以上^[3]。在苜蓿规模化种植管理中, 及时准确判断苜蓿长势和预测产量对于苜蓿的种植收获、农业保险、农业用水效率评价等具有重要意义^[4]。传统的苜蓿产量预测通过人工田间采样获得, 不仅费时、费力、具有破坏性, 而且效率低下。Sanderson 等^[5]使用上升板仪, 结合作物高度和植物密度对产量进行预测, 可避免破坏植株, 但仍需田间密集采样。Kayad 等^[6]通过卫星遥感数据估测苜蓿产量, 但存在时空分辨率低、易受云层遮挡等问题, 实际应用受限。

无人机遥感具有容易操作、机动、灵活、时效性强等特点, 近些年被广泛用于作物产量预测^[7]。Zhou 等^[8]利用无人机搭载多光谱相机采集水稻各个生育时期的光谱影像, 构建了多时相产量预测模型, 在多种植被指数

中归一化植被指数与产量的相关性较好, 多时相产量模型的精度优于单生育期模型。朱婉雪等^[9]基于无人机多光谱影像数据构建了冬小麦不同生育时期的产量预测模型, 并指出去除土壤噪声能够提高模型的精度。韩文霆等^[10]基于无人机多光谱遥感数据, 使用最小二乘法构建了玉米单生育期和多生育期的产量预测模型。Maimaitijiang 等^[11]在建模时融合了无人机光谱数据、冠层结构、热成像等多模态信息, 提高了大豆产量预测模型的精度。上述研究利用无人机搭载多种传感器, 使用多元线性回归 (Multiple Linear Regression, MLR)、随机森林 (Random Forest, RF)、人工神经网络 (Artificial Neural Network, ANN) 等方法构建产量预测模型, 然而作物产量受到环境、人为因素等诸多影响, 输入变量与产量之间呈非线性关系, 使用 MLR 构建模型可能会导致模型预测不准, 而 RF、ANN 算法存在小样本数据出现过拟合的问题^[12], 支持向量回归 (Support Vector Regression, SVR) 具有结构简单、泛化性强等特点, 在处理小样本、非线性问题中优势明显, 在很大程度上可以克服模型过拟合的问题^[13], 可作为构建产量预测模型的有效算法。

目前, 无人机遥感估产技术以小麦、玉米、水稻等大田作物的应用为主, 而在苜蓿生产中的应用较少。Feng 等^[4]集成多种机器学习算法, 构建了高光谱植被指数与产量的非线性模型, 但是尚未讨论植被指数与冠层高度等

收稿日期: 2022-02-21 修订日期: 2022-05-13

基金项目: 国家自然科学基金项目 (51939005); 河北省现代农业产业体系草业创新团队专项资金资助项目 (HBCT2018160202); 自治区区域协同创新专项 (科技援疆计划) (2021E02056); 国家牧草产业技术体系专项 (CARS-34)

作者简介: 严海军, 博士, 教授, 博士生导师, 研究方向为节水灌溉技术与装备。Email: yanhj@cau.edu.cn

多模态信息融合的建模效果。Dvorak 等^[14]基于无人机可见光遥感影像建立了苜蓿冠层的三维模型, 成功预测了苜蓿的产量与品质, 但模型的苜蓿成熟度、受病害程度等指标需要人工评价, 因此在实际生产中的应用受到限制, 同时该研究未明确苜蓿估产的最佳生育时期。

综上所述, 本文以河北地区种植的紫花苜蓿为研究对象, 使用无人机多光谱遥感平台对苜蓿的不同生育时期进行遥感监测, 以去除土壤噪声后的植被指数和株高作为输入变量, 利用支持向量回归算法构建苜蓿的产量预测模型, 以期确定苜蓿估产的最优遥感生育时期, 为苜蓿规模化高效生产和智慧管理提供技术支持。

1 材料与方法

1.1 试验地概况

试验地点位于河北省涿州市中国农业大学教学实验场, 地处北纬 39°27', 东经 115°51', 海拔 42 m。该地区属于暖温带半湿润季风区, 大陆性季风气候显著, 季节温差变化大, 光照充足, 夏季高温多雨, 冬季寒冷干燥, 年平均气温为 11.6 °C, 多年平均降雨量在 550~650 mm 之间, 其中 50%~70% 的降雨集中在 7—8 月。根据土壤颗粒分析, 试验地土壤以砂土为主 (国际制), 0~80 cm 土层平均 pH 值为 7.9, 硝态氮含量为 22.15 mg/kg, 有效磷含量为 38.75 mg/kg, 速效钾含量为 47.83 mg/kg, 有机质 8.41 g/kg。试验地种植多年生紫花苜蓿, 品种为 WL363HQ, 于 2019 年秋季播种, 采用条播方式, 行距 30 cm, 播种量为 22.5 kg/hm²。采用圆形喷灌机进行灌溉, 所用喷灌机由三跨加悬臂组成, 整机长度 140 m。

1.2 田间试验设计

试验于 2021 年在建植第 2 年的苜蓿地进行。为了使产量预测模型具有更好的适用性, 试验以 60%田间持水量 (Field Capacity, FC) 作为下限, 90%FC 作为上限进行灌溉, 并将从下限灌至上限的灌水量设为额定灌水量 W , 设置 W1 (55% W)、W2 (70% W)、W3 (85% W)、W4 (100% W)、W5 (115% W) 共 5 个处理, 以雨养处理作为对照。每个处理重复 3 次, 共 18 个 6 m×6 m 的重复小区, 其中共 5 个灌水处理设置在喷灌机喷灌范围内, 雨养处理设置在喷灌范围外。每个小区内均装有土壤水分传感器, 对土壤含水率进行实时监测, 当 W4 处理下 0~40 cm 土层内的土壤含水率平均值低至 60%FC 时, 各处理按照设计灌水量开始灌溉。2021 年 3 月 26 日试验地苜蓿开始返青, 由于土壤墒情较好, 各处理未统一灌返青水。由于第 1 茬内无有效降雨, 5 个处理的总灌水量分别为 103、131、159、187、215 mm。为了保证苜蓿的正常生长, 第 1 茬刈割后在 5 月 31 日统一灌水 54 mm。第 2 茬内有效降雨量为 9 mm, 各处理的总灌水量分别为 81、89、96、104、111 mm。苜蓿在第 3 茬、第 4 茬内有效降雨量分别为 336 和 216 mm, 远超出当地多年平均降雨量, 过多的降雨导致苜蓿出现了严重倒伏和腐烂现象, 影响了不同灌水处理试验的开展, 采集数据不具有代表性, 因此本文选取前 2 茬数据进行建模分析。

1.3 数据采集

1.3.1 无人机数据

试验采用 DJI Phantom 3 Advanced 无人机, 搭载 Micasense RedEdge-M 多光谱相机。该相机可以采集 5 个波段的数据, 分别为蓝光 (475 nm)、绿光 (560 nm)、红光 (668 nm)、红边 (717 nm) 和近红外 (840 nm), 相机焦距 5.5 mm, 视场角 47.2°, 图像分辨率 1 280×960 (像素)。

使用 DJI GS Pro 软件进行航线规划, 航向和旁向重叠度均为 75%, 飞行速度为 5.6 m/s, 飞行高度为 50 m, 地面分辨率为 3.5 cm/像素。数据采集时间为 2021 年 4 月 23 日至 6 月 22 日, 覆盖苜蓿第 1 茬、第 2 茬的分枝期、现蕾期和初花期 3 个关键生育时期, 具体如表 1 所示。选择晴朗无云、无风的天气进行飞行作业, 作业时间 10:00—14:00, 单次作业时长约 13 min, 飞行前拍摄校正白板用于辐射校正。在田间设置 5 个地面控制点, 使用司南导航 M600 mini 高精度 GNSS (Global Navigation Satellite System) 接收机测量地面控制点的精确坐标及高程。

表 1 无人机数据采集日期

茬次 Cutting	分枝期 Branching stage	现蕾期 Budding stage	初花期 Early blooming stage
1	04-23	05-10	05-23
2	06-08	06-17	06-22

1.3.2 田间实测数据

无人机数据采集后使用卷尺测量苜蓿株高, 在 18 个小区内沿对角线取 3 个 1 m×1 m 的样方, 每个样方测量 3 株具有代表性的苜蓿, 取平均值作为该样方的株高, 每个生育期共采集 54 组株高数据。第 1、2 茬分枝期、现蕾期和初花期共测得 324 组株高数据。苜蓿于 2021 年 5 月 24 日、6 月 22 日进行刈割测产, 同样在每 18 个小区内沿对角线取 3 个 1 m×1 m 的样方, 将样方内苜蓿全部刈割并留茬 5 cm, 刈割后立即称取样方鲜草质量。在所取的样方鲜草中随机抽取部分样品后称量, 并放入档案袋中保存, 随后将所取样品放入烘箱, 105 °C 杀青 30 min 后将温度调至 75 °C, 恒温下干燥 48 h, 待样品烘干后称量, 计算出干鲜比, 再根据样方鲜草质量得到干草质量。试验共测得 2 茬、108 个产量数据, 第 1 茬、第 2 茬各处理的产量平均值如表 2 所示, 产量数据整体呈现不同梯度的分布, 具有较好的代表性。

表 2 各处理苜蓿平均产量

茬次 Cutting	雨养 Rainfed	处理 1 Treatment 1 W1 (55% W)	处理 2 Treatment 2 W2 (70% W)	处理 3 Treatment 3 W3 (85% W)	处理 4 Treatment 4 W4 (100% W)	处理 5 Treatment 5 W5 (115% W)
1	2 142	2 744	3 820	5 322	6 233	6 321
2	931	1 857	2 314	3 040	3 909	4 234

注: W 为额定灌水量, mm。

Note: W is rated irrigation amount, mm.

1.4 遥感数据预处理

1.4.1 植被指数的计算与优选

使用 Pix4D Mapper 软件对遥感图像进行处理, 首先

矫正镜头畸变与渐晕效应,随后使用多光谱相机太阳光传感器结合校正白板的标准反射率值进行辐射校正,将多光谱图像的像素从原始值转换为反射率值。辐射校正后进行图像拼接,生成试验区各个波段的正射反射率图,使用各个波段的反射率值计算冠层植被指数。

健康植被的光谱反射率在绿光波段有较低的反射峰,在蓝光波和红光波段存在一个吸收谷,而在近红外波段又达到一个较高的反射峰^[15]。当作物受到水分胁迫或病虫害时,冠层结构、叶片颜色均会发生变化,因此各波段的反射率也会随之变化。为了能够增强遥感影像的解译能力,许多学者根据不同波段的反射率值构建了多种植被指数。本文根据遥感平台搭载相机的波段范围初选 11 种常用于作物产量预测的植被指数,其计算公式如表 3 所示。过多的输入变量参与建模会导致信息冗余,从而降低模型性能,因此需要优选植被指数^[16]。本文预先计算得到苜蓿分枝期、现蕾期和初花期时每种植被指数与实测产量的相关系数,然后选取每个生育时期相关性最高的 5 种植被指数作为输入参数^[17]。

表 3 常用植被指数

Table 3 Commonly applied vegetation indices

植被指数 Vegetation indices	植被指数计算公式 Vegetation indices and formulas	参考文献 References
绿红植被指数 Green-Red Vegetation Index (GRVI)	$GRVI = (G - R) / (G + R)$	[18]
归一化植被指数 Normalized Difference Vegetation Index (NDVI)	$NDVI = (NIR - R) / (NIR + R)$	[19]
绿色归一化植被指数 Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = (NIR - G) / (NIR + G)$	[20]
归一化红边指数 Normalized Difference Red-Edge Index (NDRE)	$NDRE = (NIR - RE) / (NIR + RE)$	[21]
结构不敏感指数 Structure Insensitive Pigment Index (SIPI)	$SIPI = (NIR - B) / (NIR - R)$	[22]
土壤调整植被指数 Soil Adjusted Vegetation Index (SAVI)	$SAVI = 1.5 \times (NIR - R) / (NIR + R + 0.5)$	[23]
优化土壤调整植被指数 Optimized Soil Adjusted Vegetation Index (OSAVI)	$OSAVI = (NIR - R) / (NIR + R + 0.16)$	[24]
改进比值指数 Modified Simple Ratio Index (MSR)	$MSR = (NIR / R - 1) / (NIR / R + 1)^{0.5}$	[25]
改进叶绿素吸收指数 Modified Chlorophyll Absorption in Reflectance Index (MCARI)	$MCARI = [(RE - R) - 0.2 \times (RE - G)] \times (RE / R)$	[26]
转换叶绿素吸收指数 Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	$TCARI = 3 \times [(RE - R) - 0.2 \times (RE - G) \times (RE / R)]$	[27]
增强型植被指数 Enhanced Vegetation Index (EVI2)	$EVI2 = 2.5 \times (NIR - R) / (NIR + 2.4 \times R + 1)$	[28]

注:表中 B、G、R、NIR、RE 分别代表蓝光波段反射率、绿光波段反射率、红光波段反射率、近红外波段反射率、红边波段反射率。

Note: B, G, R, NIR, RE in the table represent the reflectance of blue, green, red, near infrared red and red edge bands, respectively.

1.4.2 剔除土壤背景噪声

苜蓿在分枝期或严重水分胁迫状态下的植被覆盖度较低,图像中土壤像元会占据一定比例。有研究表明,

土壤像元会对估产模型造成干扰,剔除土壤像元可显著提高模型的预测精度^[10]。本文使用阈值分割的大津法^[29],将 NDVI 图像转换为可区分植被和土壤背景的二值图像,使用 QGIS 软件中的波段计算器与区域分析工具计算去除土壤后各样方的植被指数平均值。

1.4.3 作物表面模型与株高计算

将遥感影像导入 Pix4D 软件后,使用运动推断结构 (Structure from Motion, SfM) 方法生成试验区域的稀疏点云,经过点云加密后最终得到数字表面模型 (Digital Surface Model, DSM)^[30]。使用苜蓿刈割后拍摄的遥感影像建立试验区的数字高程模型 (Digital Elevation Model, DEM),将各生育时期的数字表面模型 DSM 与裸地的数字高程模型 DEM 做差得到作物表面模型 (Crop Surface Model, CSM)^[31],再使用 1.4.2 节中生成的二值图剔除土壤像元,由 CSM 提取出苜蓿冠层高度,并与实测株高进行对比。

1.5 苜蓿产量预测模型构建

1.5.1 支持向量回归算法

支持向量机 (Support Vector Machine, SVM) 是一种有监督式的机器学习算法,通过在高维或无限维空间中构造一个或一组超平面来实现分类和回归^[32]。对于分类问题,所构造的最优超平面使两类样本的分类间隔最大化,而对于回归问题,所求的超平面应保证所有样本点到超平面的距离最小。解决非线性问题时通过使用不同类型的核函数,将样本数据映射到高维,常用的有多项式核函数、径向基核函数、双曲正切核函数等,本文采用径向基核函数,因其能够逼近任意非线性函数并且学习收敛速度快^[33]。惩罚系数 C 和核函数系数 γ 对模型精度有较大影响, C 值和 γ 值越大,模型越复杂,越容易出现过拟合现象,而 C 值和 γ 值较小时,模型变得简单,但容易出现欠拟合。为了寻找合适的模型参数,本文通过十折交叉验证和网格参数搜寻方法调整模型参数。

1.5.2 模型构建

已有研究表明,融合光谱信息和冠层结构信息可以提高大麦^[34]、玉米^[35]、大豆^[11]等作物估产模型的精度,因此本文以各生育时期优选的 5 种植被指数和株高组合作为输入变量,将样方产量数据作为输出量建立估产模型,并与单独使用株高、植被指数作为输入变量的模型进行对比,以探究不同输入变量对苜蓿估产模型精度的影响。此外,分别构建分枝期、现蕾期和初花期 3 个具有代表性的估产模型,对比不同生育时期苜蓿估产模型的精度差异。建模时随机选取全部数据的 2/3 作为训练集,其余 1/3 作为测试集。

1.5.3 精度评价

评估模型的精度采用决定系数 (Coefficient of determination, R^2)、均方根误差 (Root Mean Square Error, RMSE) 和标准均方根误差 (Normalized Root Mean Square Error, NRMSE),其中 NRMSE 小于 10% 时表明模型的拟合效果极好,介于 10%~20% 时表明模型的拟合效果较好, >20%~30% 时表明模型的拟合效果一般,大于 30% 时表明模型的拟合效果较差^[13]。

2 结果与分析

2.1 无人机影像植被指数与产量的相关性分析

苜蓿各生育时期植被指数与产量的相关系数如表 4 所示。由表 4 可知，在分枝期时，MSR、MCARI、TCARI 指数与产量的相关性最好，相关系数绝对值均为 0.82，而 NDVI 与产量相关性最差，相关系数为 0.76，相关性绝对值排名前 5 的植被指数分别为 MCARI、MSR、TCARI、GRVI、EVI2。在现蕾期时，MSR 指数与产量的相关性最好，相关系数为 0.93，而 NDRE 指数与产量的相关性最差，相关系数为 0.65，相关性绝对值排名前 5 的植被指数分别为 MSR、MCARI、GNDVI、TCARI、GRVI。在初花期时，EVI2 指数与产量的相关性最好，相关系数为 0.87，而 GRVI、SIPI 指数与产量的相关性最差，相关系数均为 0.68，相关性绝对值排名前 5 的植被指数分别为 EVI2、SAVI、MCARI、OSAVI、TCARI。各生育时期建模时分别选取相关性最高的 5 种植被指数作为输入变量。

表 4 苜蓿各生育时期植被指数与产量的相关系数			
Table 4 Correlation coefficients between alfalfa vegetation indices and yields in each growth stage			
植被指数 Vegetation indices	分枝期 Branching stage	现蕾期 Budding stage	初花期 Early blooming stage
GRVI	0.80**	0.84**	0.68**
NDVI	0.76**	0.81**	0.70**
GNDVI	0.77**	0.91**	0.74**
NDRE	0.76**	0.65**	0.69**
SIPI	-0.71**	-0.74**	-0.67**
SAVI	0.78**	0.77**	0.86**
OSAVI	0.77**	0.79**	0.81**
MSR	0.82**	0.93**	0.74**
MCARI	0.82**	0.92**	0.84**
TCARI	-0.82**	-0.91**	-0.75**
EVI2	0.79**	0.77**	0.87**

注：**表示在 0.01 水平显著。
Note: ** represents significant at 0.01 level.

2.2 株高预测结果与精度评价

将苜蓿两茬分枝期、现蕾期和初花期采集的 324 个株高实测值与预测值进行对比，结果如图 1 所示。可以

看出，模型的误差较小（ $R^2=0.86$ ， $RMSE=6.8\text{ cm}$ ， $NRMSE=14.4\%$ ），表明无人机遥感影像提取苜蓿株高具有较高的精度。

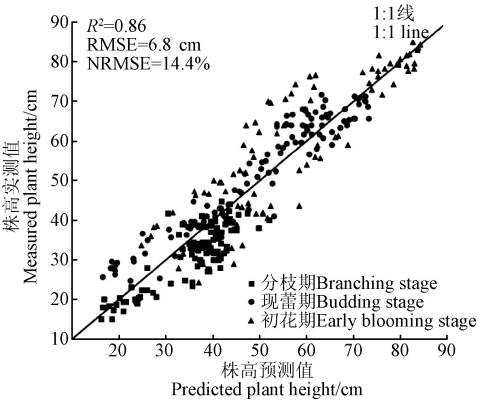


图 1 预测株高与实测株高对比
Fig.1 Comparison of predicted and measured plant heights

2.3 苜蓿产量预测模型精度评价

分别对苜蓿在分枝期、现蕾期和初花期进行产量预测。图 2~图 4 分别为分枝期、现蕾期和初花期各产量预测值与实测值对比。由图 2 可知，分枝期分别以植被指数、株高、以及植被指数和株高组合作为输入变量所构建的 3 种模型的 R^2 在 0.4~0.6 之间，其中以植被指数和株高分别作为输入变量的模型的 $NRMSE$ 均大于 30%，以植被指数和株高组合作为输入变量的模型的 $NRMSE$ 为 29.1%、 $RMSE$ 为 $1\,019\text{ kg/hm}^2$ ，总体上分枝期 3 种模型的预测精度偏低，在实际应用时会出现较大误差。由图 3 可知，现蕾期构建的 3 种模型的 R^2 均超过 0.8， $NRMSE$ 小于 20%，模型精度高于分枝期，其中以植被指数和株高组合作为输入变量的模型精度最高，该模型的 R^2 为 0.87， $RMSE$ 为 564 kg/hm^2 ， $NRMSE$ 为 16.1%。由图 4 可知，初花期以植被指数作为输入变量的产量预测模型 R^2 为 0.72，精度低于现蕾期，但是以株高、以及植被指数和株高组合作为输入变量所构建模型的 R^2 分别为 0.89、0.90，而且 $NRMSE$ 均低于 15%，精度比现蕾期有较大提高。

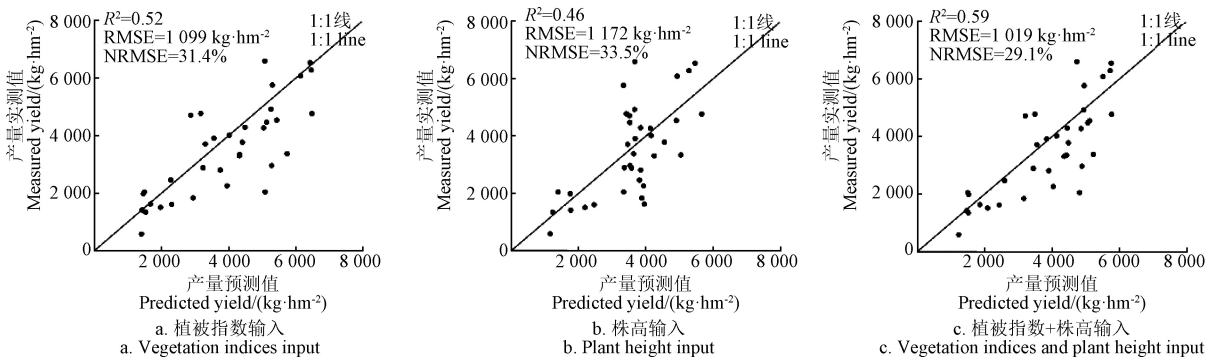


图 2 分枝期预测产量和实测产量比较
Fig.2 Comparison of predicted and measured yields at branching stage

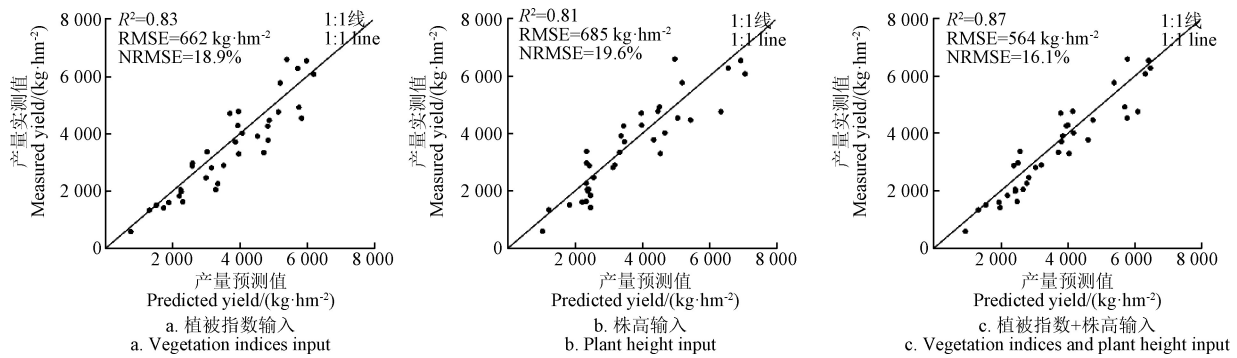


图3 现蕾期预测产量和实测产量比较

Fig.3 Comparison of predicted and measured yields at budding stage

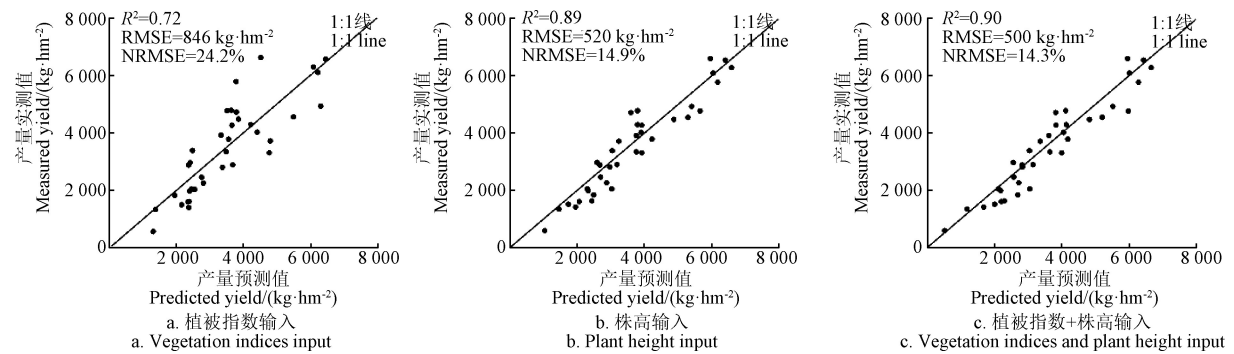


图4 初花期预测产量和实测产量比较

Fig.4 Comparison of predicted and measured yields at early blooming stage

上述研究表明,在同一生育时期,使用植被指数和株高组合作为输入变量的模型预测精度均高于仅使用植被指数或株高所构建的模型。此外,对比苜蓿分枝期、现蕾期和初花期的产量预测结果,发现预测精度随生育时期的发展呈增加趋势,其中分枝期的模型精度最低,初花期的模型精度最高(R^2 为0.90, RMSE为500 kg/hm², NRMSE为14.3%)。

3 讨论

由于苜蓿各生育时期的冠层结构不同,使得无人机获取的光谱信息存在差异,因此构建的各生育时期产量预测模型精度不同。在苜蓿分枝期,分枝数较少、茎秆较细、叶片较小,遥感影像易受杂草、残茬等其他像元的干扰,同时苜蓿叶绿素含量受到水分胁迫后呈现出先上升后下降的趋势^[36],而且植被指数对于干旱响应具有滞后性^[37],因此根据分枝期的遥感影像无法准确预测出干旱造成的减产,从而导致模型精度较低。在苜蓿现蕾期,分枝数增多,冠层覆盖度较高,遥感影像不容易被其他像元干扰,苜蓿受到干旱胁迫后,其株高和植被指数均显著下降,所构建的模型精度相比分枝期有较大提升。从初花期开始,苜蓿进入生长后期的生殖生长阶段,生长速度减缓^[38],所采集的遥感影像能更准确地反映产量,构建模型的精度整体上高于现蕾期。

除了本文使用的支持向量回归(SVR)算法外,人工神经网络(ANN)和随机森林(RF)等算法也常用于产量预测。张智韬等^[13]分别采用ANN和SVR等算法构建不同生育时期植被指数与玉米根系含水率的关系模

型,对比发现SVR算法在各生育时期的建模与预测精度最优。崔日鲜等^[39]使用ANN、SVR和RF算法建立了冬小麦叶片氮累积量的估算模型,指出SVR和ANN方法具有较高的拟合精度。Whitmire等^[40]使用SVR、RF、ANN等算法进行苜蓿估产,认为基于SVR算法构建的产量预测模型精度最高。

苜蓿初花期构建的3种产量预测模型中,分别基于株高、植被指数和株高组合建立的模型精度优于现蕾期,但是基于植被指数的模型精度却低于现蕾期,这是由于苜蓿从现蕾期至初花期叶面积指数逐渐增大、植被指数趋于饱和造成的。Viljanen等^[41]指出基于株高的草地生物量预测模型在生长后期不会出现饱和现象,本研究发现基于株高单变量的估产模型精度在生育期内逐渐递增趋势也证明了这一结论。Wan等^[42]研究认为,使用植被指数、冠层结构信息融合建模能够解决植被指数饱和的问题,并且有效提高模型精度。与上述研究结果相似,本研究使用植被指数和株高组合作为输入变量,分别在分枝期、现蕾期和初花期构建的产量预测模型精度均高于仅使用植被指数或株高的预测精度。

本研究构建的苜蓿产量预测模型使用无人机遥感影像直接获取的植被指数和株高组合作为输入变量,无需人为参与评价,相比于Dvorak等^[14]提出的基于苜蓿株高、成熟度、受病害程度等参数的模型更适用于实际生产,然而本模型在预测倒伏后的苜蓿产量时会出现偏差,此外不同品种苜蓿的茎叶比、分蘖数等指标差异也会限制模型的适用范围。因此今后研究应增加苜蓿品种和区域应用的样本数据量,进一步对模型进行验证和校准,从

而提高模型的预测精度和适用范围。

4 结 论

选取最优机器学习算法、输入变量和最佳生育时期对规模化苜蓿种植的无人机遥感估产具有至关重要的作用。本研究发现, 基于无人机遥感影像提取的苜蓿株高与人工取样实测值的决定系数 R^2 达到 0.86, 可用于有效监测苜蓿株高。以无人机遥感获得的苜蓿初花期植被指数和株高组合作为输入变量, 应用支持向量回归算法构建产量预测模型。就生育时期而言, 苜蓿初花期的预测精度最高, 其决定系数、均方根误差和标准均方根误差分别为 0.90、500 kg/hm² 和 14.3%, 其次是现蕾期、分枝期。本文建立的无人机遥感产量预测模型可以快速精准地评估苜蓿株高和产量, 为苜蓿生产的精细化、智慧化管理提供了快捷高效的技术手段。

[参 考 文 献]

- [1] Feng L, Zhang Z, Ma Y, et al. Multitask learning of alfalfa nutritive value from UAV-based hyperspectral images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.
- [2] 王云玲, 王晓玉, 李茂娜, 等. 喷灌条件下灌水量对建植初期紫花苜蓿产量与品质的影响[J]. 节水灌溉, 2016(8): 104-108.
Wang Yunling, Wang Xiaoyu, Li Maona, et al. Effects of different irrigation amounts on yield and quality of alfalfa planting in the first year under sprinkling irrigation[J]. Water Saving Irrigation, 2016(8): 104-108. (in Chinese with English abstract)
- [3] Li M, Wang Y, Adeli A, et al. Effects of application methods and urea rates on ammonia volatilization, yields and fine root biomass of alfalfa[J]. Field Crops Research, 2018, 218: 115-125.
- [4] Feng L, Zhang Z, Ma Y, et al. Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning[J]. Remote Sensing, 2020, 12(12): 2028.
- [5] Sanderson M A, Rotz C A, Fultz S W, et al. Estimating forage mass with a commercial capacitance meter, rising plate meter, and pasture ruler [J]. Agronomy Journal, 2001, 93(6): 1281-1286.
- [6] Kayad A G, Al-Gaadi K A, Tola E, et al. Assessing the spatial variability of alfalfa yield using satellite imagery and ground-based data[J]. PLoS ONE, 2016, 11(6): e0157166.
- [7] 刘忠, 万炜, 黄晋宇, 等. 基于无人机遥感的农作物长势关键参数反演研究进展[J]. 农业工程学报, 2018, 34(24): 60-71.
Liu Zhong, Wan Wei, Huang Jinyu, et al. Progress on key parameters inversion of crop growth based on unmanned aerial vehicle remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(24): 60-71. (in Chinese with English abstract)
- [8] Zhou X, Zheng H B, Xu X Q, et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 246-255.
- [9] 朱婉雪, 李仕冀, 张旭博, 等. 基于无人机遥感植被指数优选的田块尺度冬小麦估产[J]. 农业工程学报, 2018, 34(11): 78-86.
Zhu Wanxue, Li Shiji, Zhang Xubo, et al. Estimation of winter wheat yield using optimal vegetation indices from unmanned aerial vehicle remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(11): 78-86. (in Chinese with English abstract)
- [10] 韩文霆, 彭星硕, 张立元, 等. 基于多时相无人机遥感植被指数的夏玉米产量估算[J]. 农业机械学报, 2020, 51(1): 148-155.
Han Wenting, Peng Xingshuo, Zhang Liyuan, et al. Summer maize yield estimation based on vegetation index derived from multi-temporal UAV remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(1): 148-155. (in Chinese with English abstract)
- [11] Maimaitijiang M, Sagan V, Sidike P, et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning[J]. Remote Sensing of Environment, 2020, 237: 111599.
- [12] 张宏鸣, 刘雯, 韩文霆, 等. 基于梯度提升树算法的夏玉米叶面积指数反演[J]. 农业机械学报, 2019, 50(5): 251-259.
Zhang Hongming, Liu Wen, Han Wenting, et al. Inversion of summer maize leaf area index based on gradient boosting decision tree algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 251-259. (in Chinese with English abstract)
- [13] 张智韬, 谭丞轩, 许崇豪, 等. 基于无人机多光谱遥感的玉米根域土壤含水率研究[J]. 农业机械学报, 2019, 50(7): 246-257.
Zhang Zhitao, Tan Chengxuan, Xu Chonghao, et al. Retrieving soil moisture content in filed maize root zone based on UAV multispectral remote sensing[J]. Transaction of the Chinese Society for Agricultural Machinery, 2019, 50(7): 246-257. (in Chinese with English abstract)
- [14] Dvorak J S, Pampolini L F, Jackson J J, et al. Predicting quality and yield of growing alfalfa from a UAV[J]. Transactions of the ASABE, 2021, 64(1): 63-72.
- [15] 苏伟, 王伟, 刘哲, 等. 无人机影像反演玉米冠层 LAI 和叶绿素含量的参数确定[J]. 农业工程学报, 2020, 36(19): 58-65.
Su Wei, Wang Wei, Liu Zhe, et al. Determining the retrieving parameters of corn canopy LAI and chlorophyll content computed using UAV image[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(19): 58-65. (in Chinese with English abstract)
- [16] 王李娟, 孔钰如, 杨小冬, 等. 基于特征优选随机森林算法的农耕区土地利用分类[J]. 农业工程学报, 2020, 36(4): 244-250.
Wang Lijuan, Kong Yuru, Yang Xiaodong, et al. Classification of land use in farming areas based on feature optimization random forest algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of

- the CSAE), 2020, 36(4): 244-250. (in Chinese with English abstract)
- [17] 陶惠林, 冯海宽, 杨贵军, 等. 基于无人机数码影像和高光谱数据的冬小麦产量估算对比[J]. 农业工程学报, 2019, 35(23): 111-118.
Tao Huilin, Feng Haikuan, Yang Guijun, et al. Comparison of winter wheat yields estimated with UAV digital image and hyperspectral data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(23): 111-118. (in Chinese with English abstract)
- [18] Tucker C J. Red and photographic infrared linear combinations for monitoring vegetation[J]. Remote Sensing of Environment, 1979, 8(2): 127-150.
- [19] Rouse J W, Haas R H, Schell J A, et al. Monitoring the vernal advancement of retrogradation of natural vegetation[R]. NASA/GSFC Type III Final Report, Greenbelt, MD, 1973.
- [20] Gitelson A A, Kaufman Y J, Merzlyak M N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS[J]. Remote Sensing of Environment, 1996, 58(3): 289-298.
- [21] Gitelson A A, Merzlyak M N. Remote estimation of chlorophyll content in higher plant leaves[J]. International Journal of Remote Sensing, 1997, 18(12): 2691-2697.
- [22] Penuelas J, Baret F, Filella I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance[J]. Photosynthetica, 1995, 31(2): 221-230.
- [23] Huete A R. A soil-adjusted vegetation index (SAVI)[J]. Remote Sensing of Environment, 1988, 25(3): 295-309.
- [24] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996, 55(2): 95-107.
- [25] Chen J M. Evaluation of vegetation indices and a modified simple ratio for boreal applications[J]. Canadian Journal of Remote Sensing, 1996, 22(3): 229-242.
- [26] Daughtry C S T, Walthall C L, Kim M S, et al. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance[J]. Remote Sensing of Environment, 2000, 74(2): 229-239.
- [27] Haboudane D, Miller J R, Tremblay N, et al. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture[J]. Remote Sensing of Environment, 2002, 81(2/3): 416-426.
- [28] Jiang Z, Huete A, Didan K, et al. Development of a two-band enhanced vegetation index without a blue band[J]. Remote Sensing of Environment, 2008, 112(10): 3833-3845.
- [29] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [30] Su B, Xue J, Xie C, et al. Digital surface model applied to unmanned aerial vehicle based photogrammetry to assess potential biotic or abiotic effects on grapevine canopies[J]. International Journal of Agricultural and Biological Engineering, 2016, 9(6): 119-130.
- [31] Bendig J, Bolten A, Bennertz S, et al. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging[J]. Remote Sensing, 2014, 6(11): 10395-10412.
- [32] 黄茜, 杨伟才, 毛晓敏. 东北春玉米单株茎流变化规律及其农田尺度提升方法[J]. 农业工程学报, 2020, 36(16): 94-102.
Huang Xi, Yang Weicai, Mao Xiaomin. Variations of sap flow rate and its farmland scale improving method for spring maize in northeast China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(16): 94-102. (in Chinese with English abstract)
- [33] 王璨, 李志伟. 利用融合高度与单目图像特征的支持向量机模型识别杂草[J]. 农业工程学报, 2016, 32(15): 165-174.
Wang Can, Li Zhiwei. Weed recognition using SVM model with fusion height and monocular image features[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(15): 165-174. (in Chinese with English abstract)
- [34] Bendig J, Yu K, Aasen H, et al. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley[J]. International Journal of Applied Earth Observation and Geoinformation, 2015, 39: 79-87.
- [35] Geipel J, Link J, Claupein W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system[J]. Remote Sensing, 2014, 6(11): 10335-10355.
- [36] 刘军, 齐广平, 康燕霞, 等. 不同灌溉处理下紫花苜蓿光合特性、叶绿素荧光参数及生物量的变化[J]. 草地学报, 2019, 27(6): 1569-1576.
Liu Jun, Qi Guangping, Kang Yanxia, et al. Photosynthetic characteristics, chlorophyll fluorescence parameters and biomass of alfalfa under different irrigation treatments[J]. Acta Agrestia Sinica, 2019, 27(6): 1569-1576. (in Chinese with English abstract)
- [37] 黄健熙, 张洁, 刘峻明, 等. 基于遥感 DSI 指数的干旱与冬小麦产量相关性分析[J]. 农业机械学报, 2015, 46(3): 166-173.
Huang Jianxi, Zhang Jie, Liu Junming, et al. Correlation analysis between drought and winter wheat yields based on remotely sensed drought severity index[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(3): 166-173. (in Chinese with English abstract)
- [38] 周春生, 龚萍, 刘伟, 等. 改性膨润土对沙地土壤改良及紫花苜蓿生物效应的影响[J]. 灌溉排水学报, 2021, 40(7): 16-22.
Zhou chunsheng, Gong ping, Liu wei, et al. Amending sandy soils with modified bentonite to improve its physical properties and crop growth[J]. Journal of Irrigation and Drainage, 2021, 40(7): 16-22. (in Chinese with English abstract)
- [39] 崔日鲜, 刘亚东, 付金东. 基于机器学习和可见光光谱的冬小麦叶片氮积累量估算[J]. 光谱学与光谱分析, 2016, 36(6): 1837-1842.
Cui Rixian, Liu Yadong, Fu Jindong. Estimation of winter wheat leaf nitrogen accumulation using machine learning algorithm and visible spectral[J]. Spectroscopy and Spectral Analysis, 2016, 36(6): 1837-1842. (in Chinese with English abstract)

abstract)

- [40] Whitmire C D, Vance J M, Rasheed H K, et al. Using machine learning and feature selection for alfalfa yield prediction[J]. *AI*, 2021, 2(1): 71-88.
- [41] Viljanen N, Honkavaara E, Näsä R, et al. A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone [J]. *Agriculture*, 2018, 8(5): 70.
- [42] Wan L, Cen H, Zhu J, et al. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer-a case study of small farmlands in the South of China [J]. *Agricultural and Forest Meteorology*, 2020, 291: 108096.

Alfalfa yield prediction using machine learning and UAV multispectral remote sensing

Yan Haijun^{1,2}, Zhuo Yue¹, Li Maona³, Wang Yunling¹, Guo Hui¹, Wang Jingjing¹, Li Changshuo¹, Ding Feng^{1,4}

(1. College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China; 2. Engineering Research Center of Agricultural Water-Saving and Water Resources, Ministry of Education, Beijing 100083, China; 3. College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China; 4. Research Institute of Soil, Fertilizer and Agricultural Water Conservation, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China)

Abstract: Alfalfa has been one of the most widely grown forage crops around the world. The “king of forage”, alfalfa can be known as the high production of grass rich in nutrients. Timely and accurate monitoring of alfalfa growth and yield can be a high demand for large-scale agricultural production in recent years. In this study, a series of yield prediction models were established for the alfalfa using Unmanned Aerial Vehicle (UAV) remote sensing and machine learning. A field experiment was also performed in Zhuozhou City, Hebei Province, China. Five irrigation modes and a rainfed treatment were set for different alfalfa growth statuses and yield prediction. The UAV multispectral platform was used to monitor the alfalfa during the growing stage of branching, budding, and early blooming. A correlation analysis was then made to determine the alfalfa yields and 11 spectral parameters (vegetation indices). The top five correlation indexes were picked out for each growth stage. Then, a Structure from Motion (SfM) imaging technology was used to reconstruct the plant height of alfalfa. The high accuracy of prediction was achieved with the determination coefficient (R^2) of 0.86, the Root Mean Square Error (RMSE) of 6.8 cm, and the Normalized Root Mean Square Error (NRMSE) of 14.4%, compared with the measurement. Eventually, three yield prediction models were established by Support Vector Regression (SVR). Specifically, the measured yield was always used as the output, but the input was used the vegetation indices, the plant height, as well as the combination of vegetation indices and plant height. Soil noise was also removed from the vegetation indices for the best performance. The results showed that there was the most significant correlation of vegetation indices with the measured alfalfa yield during the branching, budding, and early blooming stage. In the branching stage, the R^2 values of the three models were between 0.4 and 0.6. The NRMSEs of the models were greater than 30% with the vegetation indices and plant height as the input variables. In the budding stage, the R^2 values were above 0.8, and the NRMSEs were less than 20% for all three models. The highest accuracy was achieved in the model with the combination of vegetation indices and plant height as the input variables, with the R^2 of 0.87, the RMSE of 564 kg/hm², and the NRMSE of 16.1%. The R^2 value was 0.72 in the yield prediction model with the vegetation indices only as the input variable in the early blooming stage, indicating a lower accuracy than that in the budding stage. However, the R^2 values were 0.89 and 0.90 in the models with the plant height only, and the combination of vegetation indices and plant height as the input variables, respectively, while the NRMSEs were lower than 15%, indicating a higher accuracy than that in the budding stage. A higher accuracy of the yield prediction model was found, as the alfalfa grew. The combination of plant height and vegetation indices as the input variables can also be expected to improve the accuracy of the yield prediction model. The best yield prediction was achieved in the model with the combination of five vegetation indices and plant height as the inputs at an early blooming stage, with the R^2 of 0.90, the RMSE of 500 kg/hm², and the NRMSE of 14.3%. The optimal model can be strongly recommended for the rapid and accurate prediction of alfalfa yield. The finding can provide technical support to the large-scale production and precision management of alfalfa.

Keywords: unmanned aerial vehicle; yield; vegetation index; alfalfa; support vector regression; growth stage