

融合叶绿素数据的棉花冠层光合速率 高光谱估算建模

侯卜平,王家强,李福庆,石靖,高菊,申栋妍,李克远

(塔里木大学农学院/新疆干旱区特色作物遗传改良与高效生产兵团重点实验室/
新疆绿洲农业资源与环境研究中心,新疆阿拉尔,843300)

摘要:通过设置不同的灌水量梯度,获取棉花5个生育时期(蕾期、初花期、盛花期、花铃期、盛铃期)冠层的光谱反射率、叶绿素密度和叶片净光合速率(P_n)数据,利用支持向量机(SVM)模型和随机森林(RF)模型,建立融合叶绿素密度数据和未融合叶绿素密度数据的冠层光合速率预测模型。结果表明:在水分胁迫下,叶绿素密度与净光合速率呈正相关关系;用CARS+SPA算法重复执行的方式进行特征波段筛选,降维效果显著,剔除冗余波段效率高,盛花期特征波段为332,347,416,466,672,695,711,733,752,848,954 nm和1069 nm。模型监测结果表明,融合叶绿素密度数据的模型拟合度优于未融合叶绿素的模型;比较不同模型的估算能力和模型精度,随机森林(RF)模型均优于支持向量机(SVM)模型;融合叶绿素密度的RF模型5个生育时期的建模集 R^2 分别为0.659、0.676、0.808、0.744和0.633,验证集 R^2 分别为0.635、0.675、0.786、0.725和0.627。与未融合叶绿素密度数据的模型相比,融合叶绿素密度数据模型建模集的 R^2 平均提高5.59%, $RMSE$ 平均降低2.92%, RPD 平均提高7.26%;验证集的 R^2 平均提高4.12%, $RMSE$ 平均降低1.64%, RPD 平均提高5.27%,表明融合叶绿素密度数据的棉花冠层光合速率光谱估测模型具有更高的拟合精度和稳定性。

关键词:棉花;光合速率;叶绿素密度;特征波段选择;光谱估测模型

中图分类号:S562;S127 **文献标志码:**A

Hyperspectral estimation modeling of photosynthetic rate in cotton canopy using chlorophyll data

HOU Buping, WANG Jiaqiang, LI Fuqing, SHI Jing, GAO Ju, SHEN Dongyan, LI Keyuan
(College of Agriculture, Tarim University, Key Laboratory of Genetic Improvement and Efficient Production of
Characteristic Crops in Arid Areas of Southern Xinjiang, Research Center of Oasis Agricultural Resources
and Environment of Southern Xinjiang, Alar, Xinjiang 843300, China)

Abstract: By applying different irrigation gradients, data on spectral reflectance, chlorophyll density, and canopy net photosynthetic rate (P_n) were collected across five growth stages of cotton: budding stage, beginning flower stage, full blossom stage, blossing and boll-forming stage, and full bloom stage. A canopy photosynthetic rate prediction model was developed using support vector machine (SVM) and random forest (RF) algorithms, incorporating both chlorophyll-fused and non-fused data. The results demonstrated a positive correlation between chlorophyll density and the net photosynthetic rate under water stress. The CARS + SPA algorithm was employed to repeatedly perform feature band screening, achieving a remarkable dimension reduction effect and high efficiency in eliminating redundant bands. The feature bands were 332, 347, 416, 466, 672, 695, 711, 733, 752, 848, 954 nm and 1069 nm in full blossom stage. The monitoring results of the model showed that the model fitting degree of

收稿日期:2024-01-13

修回日期:2024-04-09

基金项目:国家自然科学基金(32360547);新疆生产建设兵团科技创新人才计划项目(2022CB001-07);新疆生产建设兵团科技计划项目(2021DB019);阿拉尔市农业绿色高质量发展创新战略联盟(2021BB024)

作者简介:侯卜平(1998-),男,山东莒县人,硕士研究生,研究方向为资源利用。E-mail: 1457591850@qq.com

通信作者:王家强(1977-),男,山东泗水人,教授,主要从事干旱区资源环境遥感研究。E-mail: wjqzky@163.com

the fused chlorophyll data was better than that of the unfused chlorophyll data. Compared with the estimation ability and model accuracy of different models, the random forest (RF) model was superior to the support vector machine (SVM) model. The R^2 of the calibration set of the RF model fusing chlorophyll density in the five growth periods were 0.659, 0.676, 0.808, 0.744 and 0.633, respectively, and the R^2 of the validation set were 0.635, 0.675, 0.786, 0.725 and 0.627, respectively. Compared with the model without chlorophyll density data, the R^2 of the calibration set increased by 5.59% on average, the $RMSE$ decreased by 2.92% on average, and the RPD increased by 7.26% on average. The average R^2 of the validation set was 4.12% higher than that of the unfused chlorophyll data, the $RMSE$ was reduced by 1.64% on average, and the RPD was increased by 5.27% on average. The analysis demonstrated that the spectral estimation model of the cotton canopy photosynthetic rate, integrated with chlorophyll density data, exhibits superior fitting accuracy and stability.

Keywords: cotton; photosynthetic rate; chlorophyll density; feature band selection; spectral estimation model

净光合速率是衡量植被生产力、体现植物整体长势的重要指标,也是作物能量代谢和产物形成的基础,对于植物生长十分重要^[1]。叶绿体是绿色植物进行光合作用的基础细胞器,作为叶绿体的基本组成物质,叶绿素在植物光合作用中至关重要,其密度是植物光合作用能力、营养状况和生长态势的重要指示因子^[2]。叶绿素密度(chlorophyll density, CHD)是指单位面积或单位体积内植物叶片中的叶绿素分布情况,反映了特定区域内的叶绿素浓度,主要用于生态系统或农业中的大面积监测^[3]。在遥感技术中,叶绿素密度常用来评估大范围植被的健康状况或生产力,同时也是衡量植物冠层光合作能力用能力和植株健康的重要指标;而叶绿素含量是指单个叶片或特定植物体内叶绿素的绝对含量,通常以质量表示。叶绿素含量反映单一植物或叶片的光合作用潜力,常用于植物营养状况和光合作用效率的分析。简言之,叶绿素密度主要用于大面积或区域性的植被监测,叶绿素含量则侧重于单一植株或局部的测量。

传统作物光合监测技术均是以气体交换,或是以群体尺度干物质累积量的观测为基础,大都耗时费力,空间代表性差。近年来,遥感技术的发展为作物生理生化参数定量监测带来了更多的可能性,以遥感手段辅助农作物光合参数的定量监测也逐渐被深入研究与应用。在作物遥感监测领域已有众多研究表明,利用高光谱技术监测作物的冠层含水量、生物量、光合色素含量等生理指标,能够大面积、动态且无损地获取植被时空与生长信息^[4-5]。已有研究表明,密度数据与植物光谱存在密切关联。Pinar 等^[6]研究发现,“红边”参数与叶片叶绿素密度具有较好的相关性,说明群体植被光谱“红

边”的位置可以较好地反映植物的叶绿素密度信息;陈燕等^[7]研究表明,反射光谱与微分光谱曲线在红光区、近红外区呈现明显的层次性波形,对应叶绿素密度的变化,表明高光谱数据可以对棉花叶绿素密度进行估算。在作物光合监测方面,吕玮等^[8]基于小麦旗叶高光谱波段反射率与净光合速率(P_n)进行相关性分析确定敏感波段,分别采用 QPSR、PLSR、BPNN 共 3 种方法搭建小麦旗叶的净光合速率反演模型,发现这 3 种方法以及敏感波段对 P_n 的估测是可行的;王月等^[9]基于小波变换的光谱指数分析不同尺度与净光合速率的相关性,以最佳分解层次敏感波段构建光谱指数,建立 BP 神经网络模型;刘聪等^[10]研究发现,使用叶绿素相关植被指数与 PAR_{in} 的乘积反演叶片净光合速率,可以显著提高模型精度。上述研究大都针对单一变量进行估测,鲜少考虑不同生理因子之间的关联性。光合速率与叶绿素密度、温度、 CO_2 浓度等多个因子存在显著相关关系^[11],叶绿素 a (Chl a) 和叶绿素 b (Chl b) 对于初级反应是必需的。Chl a 和 Chl b 吸收不同波长的阳光(Chl a 主要吸收红橙色光,Chl b 主要吸收蓝紫色光),导致叶片叶绿素(Chl a+b)总量和分配比(Chl a/b)变化,直接影响植物的光合能力。因此,对作物的光合速率进行研究不能只考虑外部环境因素,也要考虑植物的内部生理状态。

本研究在构建棉花净光合速率的反演模型时,融合叶绿素密度数据作为主要输入参数,对比分析了融合叶绿素密度数据与未融合叶绿素密度数据的光合速率预测模型,探讨高光谱数据融合叶绿素密度数据反演棉花光合参数的可行性,同时对比不同模型拟合方法的适用性,以期为棉花光合作用的监测及诊断提供技术支撑和模型参考。

1 材料与方法

1.1 研究区概况及试验设置

试验于2022年在中国科学院阿克苏农田生态系统国家野外科学观测研究站(40°37'N, 80°49'E)进行。该研究站位于塔里木盆地,属于典型的暖温带极端大陆性干旱荒漠气候区,夏季炎热,冬季寒冷,日照时间长,降水稀少,蒸发强烈,海拔1 030 m,该地区年平均气温11.2℃,年平均降水量45.2 mm,无霜期211 d,全年日照时数2 940 h。

试验设置5个灌水梯度:900 m³·hm⁻²(T1), 1 800 m³·hm⁻²(T2), 2 700 m³·hm⁻²(T3), 3 600 m³·hm⁻²(T4), 4 500 m³·hm⁻²(T5),冬灌用水量为2 250 m³·hm⁻²,其中T5为常规灌水量,作为对照处理,小区加装水表控制灌水量。试验田0~80 cm土层基本养分状况:碱解氮36.75 mg·kg⁻¹,速效磷10.48 mg·kg⁻¹,速效钾188.06 mg·kg⁻¹,有机质14.15 g·kg⁻¹,总盐5.62 mg·kg⁻¹,pH值8.0。供试棉花品种为‘塔河2号’,采用覆膜点播,于2022年4月上旬播种,一膜六行,宽行60 cm,窄行20 cm,株距10 cm,灌水施肥方式为膜下滴灌水肥一体化,尿素、磷酸氢二铵和硫酸钾的施用量分别为35、10 kg·667m⁻²和25 kg·667m⁻²。在棉花全生育期内,分别于蕾期、初花期、盛花期、花铃期、盛铃期进行叶片光合速率、叶绿素密度以及光谱反射率的测定。

1.2 数据获取

1.2.1 棉花冠层光谱反射率的测定 叶片光谱反射率采用PSR-1100手持式地物光谱仪进行测定,该光谱波段范围介于320~1 100 nm之间,光谱采样间隔1 nm,光谱分辨率3 nm。选择晴朗无风天气的12:00—14:30进行光谱测定。在每个梯度范围内挑选5个采样点,每测3个点进行一次白板校正,每个采样点测3株,每株测3次剔除异常值,取平均值作为该点的光谱反射率数据。

1.2.2 棉花冠层光合速率的测定 采用Li-6400XT型便携式光合仪(LI-COR公司,美国)测定净光合速率(P_n),测定前检查CO₂和H₂O IRGA零点,在晴天12:00—14:30测定。选择连续无损伤且长势均匀的3株棉花功能叶(倒4叶)进行净光合速率测定,取平均值。

1.2.3 棉花冠层叶绿素密度的测定 叶绿素密度(CHD)是指每块土地上的叶绿素总含量,为每平方米叶绿素含量与鲜叶质量之积。将采集光谱后的叶片分装入自封袋中,贴上标签,带回实验室用于

叶绿素含量的测定。测定之前,先用吸水纸将叶片上的水分擦拭干净,对叶片进行分类、编号,去除主脉,剪成1 mm左右的小段。将处理好的叶片和残渣放入10 mL 95%乙醇的密封试管中,静置避光放置24~48 h,直至叶片组织变白。使用95%的乙醇缓冲液提取-可见光光度法测定叶绿素含量,并计算叶绿素密度,具体计算公式如下:

$$A_{\text{chl a}} = 0.01373 A_{663} - 0.000897 A_{537} - 0.003046 A_{647} \quad (1)$$

$$A_{\text{chl b}} = 0.02405 A_{647} - 0.004305 A_{663} \quad (2)$$

$$CHD_a = A_{\text{chl a}} \cdot V \cdot M_a / (1000 \cdot W_f \cdot SLW) \quad (3)$$

$$CHD_b = A_{\text{chl b}} \cdot V \cdot M_b / (1000 \cdot W_f \cdot SLW) \quad (4)$$

$$CHD_{a+b} = CHD_a + CHD_b \quad (5)$$

式中, $A_{\text{chl a}}$ 、 $A_{\text{chl b}}$ 分别为叶绿素a、叶绿素b的吸光度值; A_{663} 、 A_{537} 、 A_{647} 分别为波长663、537、647 nm处的吸光度值; CHD_a 、 CHD_b 分别为叶绿素a、叶绿素b的密度值(mg·m⁻²); CHD_{a+b} 为总密度值(mg·m⁻²); V 为浸提体积(mL); M_a 、 M_b 分别为叶绿素a、叶绿素b的分子量($\mu\text{mol} \cdot \text{mL}^{-1}$); SLW 为比叶重($\mu\text{mol} \cdot \text{mL}^{-1}$); W_f 为叶片鲜质量($\mu\text{mol} \cdot \text{mL}^{-1}$)。

2 数据处理

2.1 光谱预处理方法

高光谱遥感影像数据具有高维和信息冗余等特征,不仅包含样本信息,还包含背景信息和噪声^[12-13]。特征提取通过一定的数学模式将影像数据从高维空间映射到低维空间,降维后的特征可在一定程度上保留原始数据尽可能多的重要信息^[14-15]。在Matlab中使用Savitzky-Golay函数对光谱数据进行平滑降噪处理。本研究最后选取窗口宽度10、多项式阶数2为最佳的平滑处理参数。

2.2 特征波段的选择方法

竞争自适应加权抽样(CARS)算法通过自适应重加权采样技术将回归系数大的波长点保留,将权重小的点去除,模仿了达尔文进化论的“适者生存”原则,通过交叉验证选取均方根误差小的变量组合^[16],其缺点是依然保留了许多变量。连续投影算法(SPA)是通过比较变量投影来选择特征波长的方法,它能够从光谱信息中充分寻找含有最低冗余信息的变量组,能够很好地消除特征波长之间存在的共线性问题,而且避免了重叠信息的重复提取^[17-18]。CARS和SPA的组合可以整合这两种方法的优点,以达到更好的预测效果。因此,本研究将采用了CARS+SPA的方法选择光谱变量,对其进行波数变量优选以减少建模所需的光谱信息,从而降低模型

复杂度,提高检测效率。

2.3 模型构建方法

将棉花冠层原始光谱经 Savitzky-Golay 函数对光谱数据进行平滑降噪处理,采用 CARS、SPA 两种算法筛选光谱特征波段,选择最优量子集提取特征波段。将提取出的特征波段作为输入,利用支持向量机(support vector machine, SVM)模型和随机森林(random forest, RF)模型分别建立融合与未融合叶绿素密度数据的光合速率预测模型。

随机森林(RF)是 Breiman 提出的一种基于分类回归树模型的集成学习算法,具有高效处理大样本数据、不必担心过拟合、可估计某个特征变量以及抗噪音等特点,能够较好地处理高维度样本,并且还可在较大数据集上有效运行,对输入变量进行重要性评估^[19]。支持向量机(SVM)是一种基于统计学习理论的模式识别方法,其核心思想是利用训练样本建立一个样本间间隔最大的分离超平面,并将样本尽可能的逼近超平面,使得偏差至最小^[20]。

2.4 精度验证方法

分别将同步测量 5 个生育时期的 75 个样本数据按 2:1 随机分为建模集和验证集,用验证集样本检验模型精度。选取以下指标对模型精度进行评价:决定系数(determination coefficient, R^2)、均方根误差(root mean square error, $RMSE$)和相对分析误差(relative percent deviation, RPD)。 R^2 越接近 1, $RMSE$ 越小, RPD 越大,说明构建的模型精度越高^[21]。 RPD 介于 1.6~2.0 之间时,模型精度可以接受; $RPD > 2.0$ 时,所建模型有较高可靠性,能够用于棉花净光合速率估测。

$$R^2 = \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (7)$$

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (8)$$

$$RPD = \frac{SD}{RMSE} \quad (9)$$

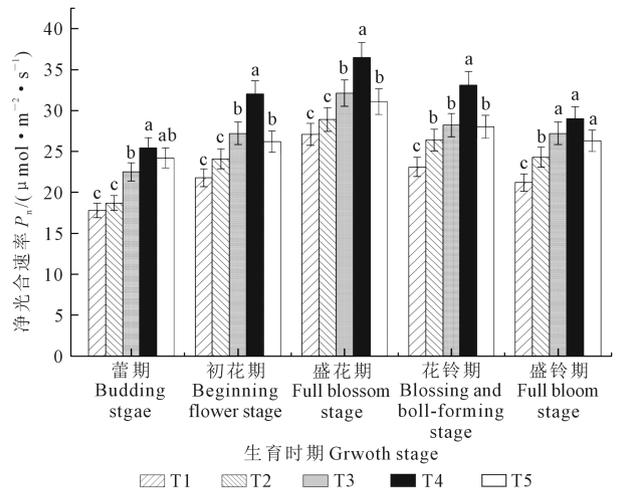
式中, x_i 表示预测值; \bar{x} 表示预测值的平均值; y_i 表示实测值; \bar{y} 表示实测值的平均值; n 表示样本数量; SD 表示标准差。

3 结果与分析

3.1 各生育时期不同灌水量下棉花冠层光合速率及叶绿素密度的变化特征

由图 1 可知,不同水分处理的棉花叶片净光合速率 P_n 随生育时期的推进呈先增后降的变化趋势,在盛花期达到峰值, T1~T5 处理盛花期较蕾期 P_n 的增量分别为 9.3、10.2、9.6、11.4、6.9 $\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$,不同水分处理盛铃期较盛花期的降量分别为 5.9、4.6、4.9、7.5、4.8 $\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ 。可能是盛花期光合作用较强,对水分需求量较大,表现为 P_n 值随着棉花的生长而增大,后期因叶片老化和温度下降, P_n 值也随之下降。

整个生育周期内,随着灌水量的增加,各水分处理组的叶片 P_n 呈现先增加后降低的趋势。T4 处理($3\ 600\ \text{m}^3 \cdot \text{hm}^{-2}$) P_n 最高,在蕾期、初花期、盛花期、花铃期和盛铃期分别比 T3 处理增加 12.98%、17.65%、13.71%、17.38% 和 6.62%;当灌水量增至 $4\ 500\ \text{m}^3 \cdot \text{hm}^{-2}$ (T5) 时, P_n 下降,各生育时期较 T4 处理分别降低 4.96%、22.14%、17.36%、18.21% 和 10.27%。这表明适当增加灌水量可显著提升棉花叶片的净光合速率,而灌水量过高会产生负面影响,影响光合效率。



注:柱上不同小写字母表示同一时期不同处理间差异显著 ($P < 0.05$)。下同。

Note: Different lowercase letters on the column indicated significant differences between treatments in the same growth stage ($P < 0.05$). The same below.

图 1 不同生育时期棉花冠层叶片净光合速率的变化
Fig.1 Changes of net photosynthetic rate of cotton canopy leaves at different growth stages

由图 2 可知,各水分处理的叶绿素密度随生育时期的变化趋势与净光合速率基本一致,在盛花期达到峰值,利于其保持更为持久的光合能力。不同生育时期,叶绿素密度均表现为 T4>T3>T2>T5>T1,即 T4 处理最高,T1 处理最低,且两处理间差异显著。由此分析,在相同水分处理下叶绿素密度的变化与净光合速率有较强的相关性,且存在一定的正相关关系。因而在进行高光谱估算模拟中,可嵌入叶绿素密度生理因子辅助光合监测,采用相同的建模方法,对比分析加入叶绿素密度对模型训练效果和精度的影响。

3.2 特征波段的优化选择

为去除冗余信息,提高建模效率,针对棉花冠层 320~1 100 nm 的光谱,利用 CARS 与 SPA 相结合的方式对 Savitzky-Golay 平滑光谱数据进行特征波段选择。CARS 在特征波长选取中,利用蒙特卡罗交叉验证选取最优潜在变量,设定蒙特卡罗采样次数为 100。图 3a 分别为随采样次数变化所选取的特征波长个数、交叉验证 RMSE 随采样次数的变化以及采样过程中回归系数的变化,对模型贡献率大的变量的回归系数被放大和保留。在迭代到 28 次时, RMSE 达到最小值 2.757,至此共选择 153 个波段(图 3b),占全波段的 19.6%。由于 CARS 提取的特征波长数量仍然较多,使用 SPA 算法对 CARS

提取的 153 个波段进行二次优选,进行第二次数据降维。设定数量范围为 12,当变量数为 12 个时,基于校正模型 RMSE 值从所有待选波长集合中选出符合设定数量范围的特征波长,此时 RMSE 为 2.641(图 3c)。提取到的 12 个特征波长的分布如图 3d 所示,即选择最优变量子集为构建模型的特征波段。根据波段顺序提取到的特征波长依次为 320、332、350、421、550、651、713、750、803、938、1 075、1 100 nm。

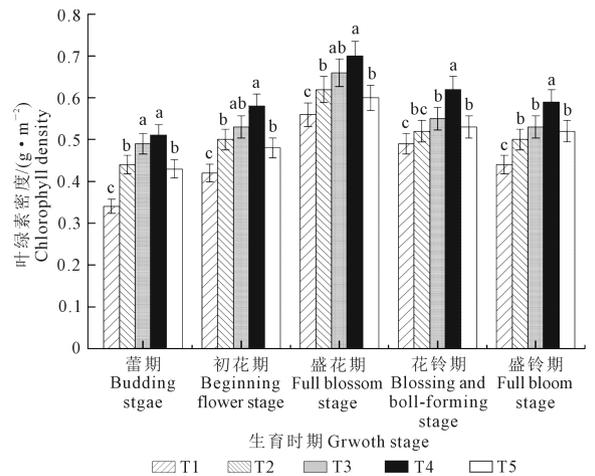


图 2 不同生育时期棉花冠层叶片叶绿素密度的变化
Fig.2 Changes of chlorophyll density in cotton canopy leaves at different growth stages

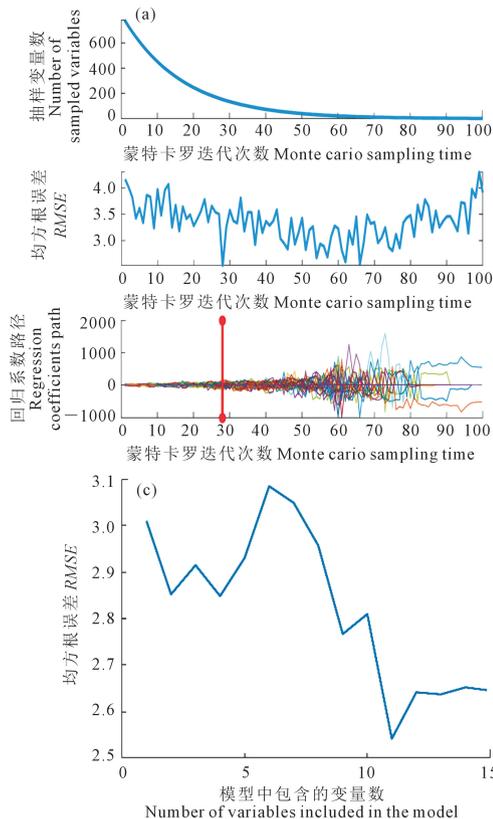
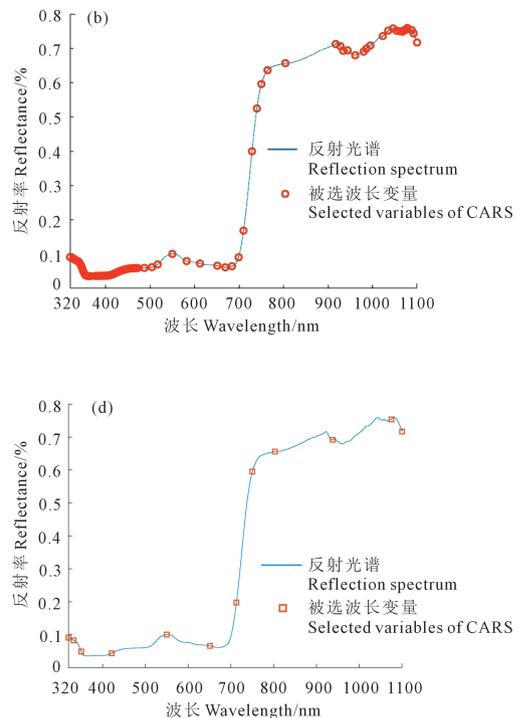


图 3 基于 CARS+SPA 算法的特征波段筛选过程及选择波段

Fig.3 Feature band selection process and band selection based on CARS+SPA algorithm



由表 1 可知,经 SPA 算法进行二次优选,特征波段数量急剧下降,一些相近的波段被剔除。经 CARS 筛选,SPA 二次优选符合设定数量范围的特征波长的变量子集特征波段。在蕾期和花铃期第一次随蒙特卡罗采样次数变化所选取的特征波长个数较多,总体而言,筛选的特征波段分布较为均匀。蕾期和初花期波段筛选过程中,随着蒙特卡罗采样迭代次数分别到 28 次和 40 次, $RMSE$ 达到最小值,分别为 2.757 和 2.081;所选波段个数分别为 153 个和 74 个,占总波段的 19.6%和 9.5%。后经 SPA 再次筛选当符合设定数量范围(12 个)的特征波长时, $RMSE$ 分别为 2.641 和 1.436。盛花期和花铃期随着蒙特卡罗采样迭代次数分别到 38 次和 19 次, $RMSE$ 达到最小值,分别为 1.649 和 1.855;所选波段个数分别为 84 个和 264 个,占总波段的 10.8%和 33.8%;后经 SPA 筛选, $RMSE$ 分别为 1.625 和 1.505。盛铃期蒙特卡罗采样迭代次数为 35 次, $RMSE$ 达到最小值 3.158;所选波段个数为 101 个,占总波段的 12.9%;SPA 筛选 $RMSE$ 为 2.277。综上所述,经过二次优选后,特征波段数量急剧下降,一些相近的波段被剔除,从而大大提高建模效率。

3.3 融合叶绿素密度数据的棉花冠层净光合速率估算模型

将不同生育时期棉花冠层优选 CARS-SPA 提取到的特征波长所对应的光谱数据与净光合速率参量分别构建 RF 模型和 SVM 模型。基于 MATLAB 软件运行随机森林算法(RF)和支持向量机(SVM)回归试验代码。使用 CARS-SPA 筛选的 12 个特征波段作为模型输入,加入生理因子叶绿素密度数据,净光合参量作为输出量,建立融合叶绿素密度数据的棉花冠层光合速率预测模型。为对比分析加入叶绿素密度数据对模型训练效果和精度的影响,采用相同的建模方法,分别对融合叶绿素的试

验样本集 A 和未融合叶绿素的试验样本集 B,建立光合速率估算模型,在 MATLAB 软件中运行随机森林算法和支持向量机回归试验代码。

由表 2 可知,根据 CARS-SPA 优化组合提取到的特征波段,不同生育时期、不同模型的表现能力不同。综合建模表现与验证精度 R^2 、 $RMSE$ 和 RPD 发现,总体上随机森林模型的效果优于支持向量机回归模型。各生育时期基于不同模型的显著性均达到 0.01 极显著水平,说明各模型的拟合性较好。不同生育时期模型估算能力表现为盛花期>花铃期>初花期>蕾期>盛铃期,模型的建模集 R^2 分别为 0.808、0.744、0.676、0.659 和 0.633,验证集 R^2 分别为 0.786、0.725、0.675、0.635 和 0.627。以盛花期的反演效果最佳,反演精度最高,该时期融合叶绿素密度建模集和验证集的 R^2 分别为 0.808 和 0.786, $RMSE$ 分别为 1.639 和 1.433, RPD 分别为 2.257 和 2.165。

由表 3 可知,基于不同模型未融合叶绿素密度数据的建模集和验证集的 RPD 均小于融合叶绿素密度数据模型,在不同模型的表现能力上,随机森林整体优于支持向量机。模型精度以盛花期最高,随机森林模型未融合叶绿素密度数据建模集和验证集的 R^2 分别为 0.721 和 0.720, $RMSE$ 分别为 1.824 和 1.453, RPD 分别为 1.893 和 1.890;建模集和验证集的 RPD 均小于 2,表明该模型可对样本进行粗略估算,但拟合精度差,稳定性不理想。相比之下,融合叶绿素密度数据建模集和验证集的 RPD 均大于 2,其模型精度最高。综上所述,叶绿素密度数据作为输入量融合建模可以提高模型的精度。

图 4 为不同生育时期融合叶绿素密度数据的净光合速率最优监测模型拟合图,综合不同生育时期来看,棉花净光合速率最优模型在盛花期 RPD 精度最高,盛铃期 RPD 精度最低,模型精度整体表现为盛花期>花铃期>初花期>蕾期>盛铃期。从建模集

表 1 不同生育时期 CARS+SPA 算法最佳波段选择
Table 1 Best band selection of CARS+SPA algorithm at different growth stages

生育时期 Growth stage	CARS				SPA		最佳波段选择 Optimal bands selection
	迭代次数 Number of iterations	$RMSE$	波段数量 Number of bands	波段数占比 Proportion of bands/%	波段数量 Number of bands	$RMSE$	
蕾期 Budding stage	28	2.757	153	19.6	12	2.641	320,332,350,421,550,651,713,750,803,938,1075,1100
初花期 Beginning flower stage	40	2.081	74	9.5	12	1.436	337,554,723,764,783,903,935,965,1047,1060,1075,1091
盛花期 Full blossom stage	38	1.649	84	10.8	12	1.625	332,347,416,466,672,695,711,733,752,848,954,1069
花铃期 Blossing and boll-forming stage	19	1.855	264	33.8	12	1.505	362,369,409,576,611,724,994,1035,1048,1078,1087,1092
盛铃期 Full bloom stage	35	3.158	101	12.9	12	2.277	359,554,619,677,707,752,764,811,943,971,1037,1055

表 2 不同生育时期融合叶绿素密度数据的棉花冠层光合速率模型表现

Table 2 Photosynthetic rate model of cotton canopy with chlorophyll density data at different growth stages

生育时期 Growth stage	模型算法 Modeling method	建模集 Calibration set			验证集 Validation set		
		R^2	RMSE	RPD	R^2	RMSE	RPD
蕾期 Budding stage	SVM	0.632	2.884	1.649	0.622	2.384	1.628
	RF	0.659	2.753	1.712	0.635	2.596	1.656
初花期 Beginning flower stage	SVM	0.659	2.578	1.713	0.652	1.947	1.695
	RF	0.676	2.152	1.758	0.675	2.769	1.755
盛花期 Full blossom stage	SVM	0.742	1.785	1.971	0.733	1.529	1.935
	RF	0.808	1.639	2.257	0.786	1.433	2.165
花铃期 Blossing and boll-forming stage	SVM	0.717	1.896	1.881	0.716	1.807	1.877
	RF	0.744	1.877	1.979	0.725	1.697	1.908
盛铃期 Full bloom stage	SVM	0.612	2.782	1.606	0.605	2.375	1.591
	RF	0.633	2.446	1.652	0.627	3.413	1.637

表 3 不同生育时期未融合叶绿素密度数据的棉花冠层光合速率模型表现

Table 3 Photosynthetic rate model of cotton canopy without chlorophyll density data at different growth stages

生育时期 Growth stage	模型算法 Modeling method	建模集 Calibration set			验证集 Validation set		
		R^2	RMSE	RPD	R^2	RMSE	RPD
蕾期 Budding stage	SVM	0.615	3.111	1.613	0.604	2.100	1.590
	RF	0.643	2.769	1.674	0.636	2.544	1.657
初花期 Beginning flower stage	SVM	0.622	2.370	1.626	0.610	2.646	1.602
	RF	0.652	2.189	1.695	0.650	2.818	1.691
盛花期 Full blossom stage	SVM	0.718	1.559	1.884	0.711	2.101	1.861
	RF	0.721	1.824	1.893	0.720	1.453	1.890
花铃期 Blossing and boll-forming stage	SVM	0.645	1.861	1.679	0.635	2.214	1.655
	RF	0.696	1.967	1.813	0.694	1.811	1.808
盛铃期 Full bloom stage	SVM	0.603	2.633	1.558	0.592	2.836	1.565
	RF	0.611	2.445	1.604	0.606	3.480	1.594

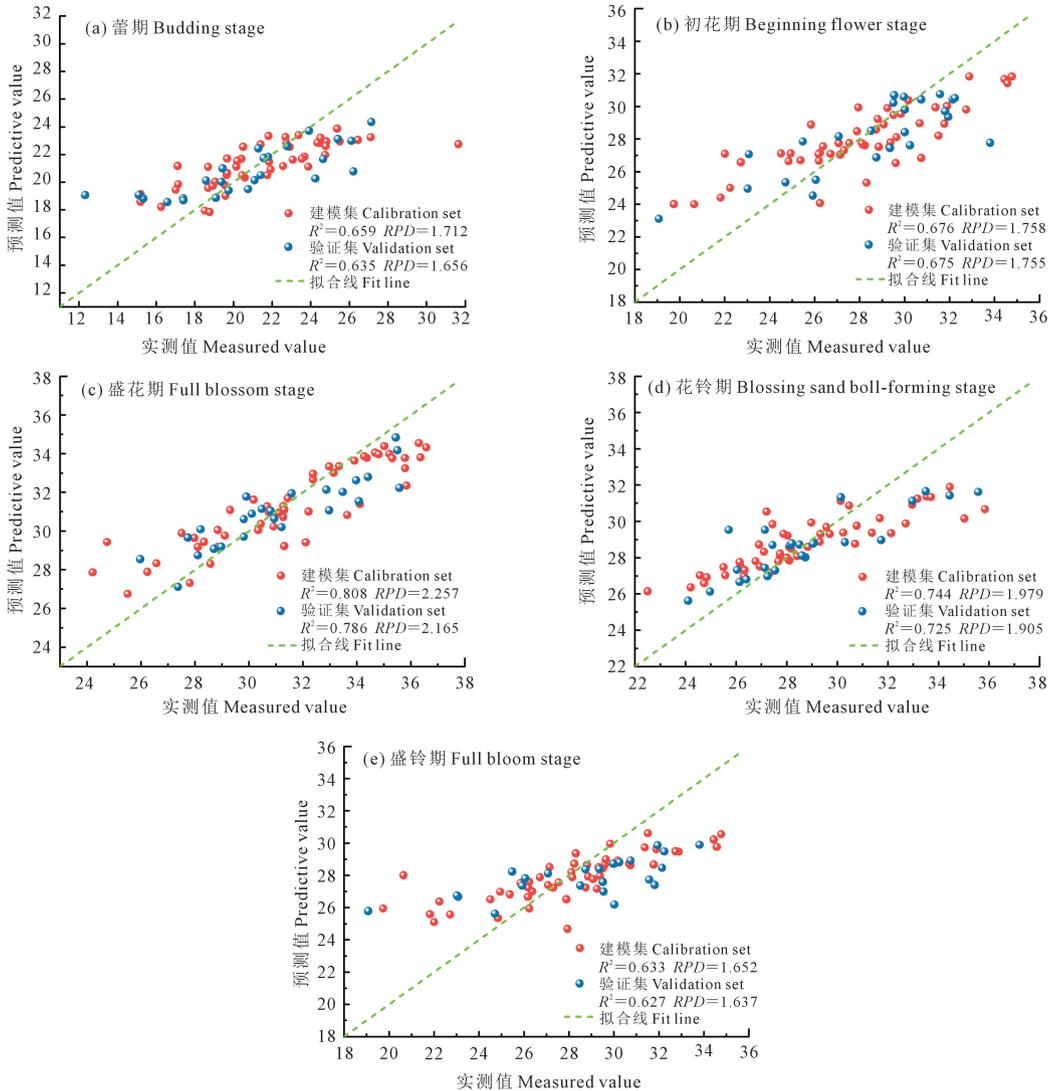


图 4 融合叶绿素密度数据的净光合速率最优监测模型

Fig.4 Optimal monitoring model of net photosynthetic rate fused with chlorophyll density data

与验证集的拟合分析来看,RF 模型的实测值和预测值在 1 : 1 分割线的两侧分布较为均匀, P_n 的估测模型对于真实值的预测范围较为准确,决定系数 R^2 和相对分析误差 RPD 均大于 0.6 和 1.7,很好地解决了单生育时期单个指标不能整体反映棉花光合能力的问题。

4 讨 论

在高光谱遥感监测应用中,光谱特征是作物生理特征引起的变化,而生理特征反映作物的生长情况,因此可以根据光谱响应的差异监测作物的生长状况^[22]。光合作用是植物叶片吸收光能和转换光能的过程,光合有效辐射是反映作物长势和组成预报产量模型的重要特征参数,也是植物进行光合作用形成初级生物量的重要能源,而净光合速率 P_n 是衡量光合作用强弱的主要指标^[23]。光合速率与叶绿素含量、温度、 CO_2 浓度、光照强度、水分等多个因子存在显著相关关系。有研究表明,作物叶绿素含量与净光合速率之间呈正相关关系,即在一定范围内叶绿素含量越高,净光合速率越强^[24]。本研究中随着灌溉梯度增加,棉花叶片净光合速率与叶绿素密度均呈先增后降的变化趋势;其中以 T4 处理最高,灌溉量增至 T5 处理时有所下降。可能是适量灌溉可以保证植物生长,使气孔正常开放, CO_2 能够顺利进入叶片,为光合作用提供原料,同时水分也能保证光合生理过程的正常进行;而灌水量过多会使土壤通气性变差,供氧量减少,植物根系的呼吸作用受抑,其对矿质元素的吸收等生理活动受阻,还可能使叶片生长不良,最终导致光合能力下降^[25]。叶片的叶绿素密度随生育时期推进表现为先增大后减小的变化过程,这与叶片随着伸长,从功能盛期到衰老、叶色由淡转深再转淡的过程相一致^[26]。随着叶绿素含量增加光合强度不断加强,达到高峰后,叶片衰老,叶绿素分解,导致光合强度下降。因此在遥感监测应用中,嵌入叶绿素生理因子,输出因子预测模型提高模型的精度及稳定性是可行的。

前人研究表明,光谱预处理方法可以消除光谱噪声及散射效应,进而增强光谱特征^[27]。光谱数据进行适当的预处理后,特征波段选择至关重要。随着计算数学的快速发展,用于选择特征波段的方法很多,不同研究中使用的特征波段差异很大。Zhao 等^[28]在叶片尺度上进行小麦白粉病监测时,SPA 算法选择的主要波段位于 500~760 nm 之间,分别为 423、528、597、602、645、675、714、737、774 nm 和 1 057 nm。CARS 算法可有效提取全波段的主要光

谱信息,但该算法未充分考虑到多重共线性问题,提取出的相近波段较多。而 SPA 算法通过计算剩余变量与选取变量的投影向量大小来选择特征变量,能够保证选取变量间的线性关系最小,以消除变量间的冗余信息,减弱多重共线性,但其往往运算效率较低^[29]。基于这一现实情况,本研究将 CARS 算法和 SPA 算法联合使用,既可以提取出有效波段,多重共线性较低,又能提升运算效率^[30],从而提高了模型的分类精度和稳健性。本研究筛选的波段主要分布在蓝光、绿光及红光以及近红外区域,确立的敏感波段筛选方法也可为其他相关参数的监测研究提供参考和借鉴。机器学习算法可有效分析和利用信息丰富的数据集以及高维观测数据,已被广泛应用于遥感数据的分析以及建模反演,但针对不同作物种类的机器学习算法存在差异,性能也各不相同^[31]。本研究以特征波段结合光合参数,使用支持向量机和随机森林回归构建棉花冠层光合速率估算模型发现,随机森林模拟结果更加稳定,反演效果最好。此结论与刘娣等^[32]对于多层土壤湿度反演中的研究结果一致。本研究构建的模型中,在输入参量相同时,随机森林回归模型的预测结果均优于支持向量机,且前者的拟合趋势更好,散点分布均匀集中。随机森林模型建模过程简单,容易实现,需要调的参数较少,计算量较小,且其在很多实际数据建模中有很好的性能支持;而向量机模型需要调的参数较多,有时很难找到合适的核函数。同时,随机森林模型处理高维度数据的能力也非常优秀,在训练完成后,能够给出变量的重要性程度^[33]。

本研究对原始光谱数据进行预处理,有效降低了环境对光谱数据的影响,但是野外数据采集依然受到土壤、大气和周边冠层等因素影响,且不同地区的棉花叶片高光谱特征存在差别,导致所选择的特征波段与前人研究中水分的敏感波段有部分差异^[34],比如 CARS 和 SPA 方法所选择的波段位置出现一定的偏移,多集中于蓝光区和近红光区。此外,本研究使用 5 个不同生育时期的数据集对模型进行反复验证,一定程度上克服了以往研究中数据集单一的缺陷。将叶绿素、净光合速率及光谱参数三者紧密联系在一起,搭建光合参数与光谱之间的桥梁,明确不同水分处理下棉花光谱反射率的变化规律,解析棉花生长过程中光合参数与光谱指数之间的量化关系,构建其反演估算模型。利用遥感技术对作物的光合作用进行大面积、动态、非破坏性的遥感监测,对现代农业的生产管理具有指导意义。

5 结 论

本研究以原始光谱数据 SG 平滑后 CARS+SPA 算法筛选特征波段,加入生理因子叶绿素密度数据嵌套试验获得 5 个生育时期数据,并进行归一化处理,输入数据根据有无叶绿素密度分成两类,通过两种机器学习(RF、SVM)进行模型训练,加入叶绿素密度作为输入用于建模和特征提取,以预测棉花冠层净光合速率。结果表明,在考虑融合叶绿素数据的条件下,其训练效果与模型拟合度均优于未融合叶绿素数据的训练模型,且 RF 随机森林方法可以获得更好的模型效果,模拟精度高,其盛花期的建模精度最高。在未融合叶绿素密度数据的模型中,盛花期建模集与验证集的 R^2 分别为 0.721 和 0.720, $RMSE$ 分别为 1.824 和 1.453, RPD 分别为 1.893 和 1.890;在融合叶绿素密度数据模型中,该时期建模集和验证集的 R^2 分别为 0.808 和 0.786, $RMSE$ 分别为 1.639 和 1.433, RPD 分别为 2.257 和 2.165。因此,融合叶绿素密度数据作为输入的训练模型拟合精度高,更具稳定性,可为棉花冠层光合速率反演研究及营养诊断提供数据支撑和模型参考。

参 考 文 献:

[1] 王娣, 佃袁勇, 乐源, 等. 基于高光谱植被指数的叶片净光合速率 P_n 反演[J]. 地理与地理信息科学, 2016, 32(4): 42-48.
WANG D, DIAN Y Y, YUE Y, et al. Net photosynthetic rate inversion based on hyperspectral vegetation indices[J]. Geography and Geo-Information Science, 2016, 32(4): 42-48.

[2] 李保珠, 赵孝亮, 彭雷. 植物叶绿体发育及调控研究进展[J]. 植物学报, 2014, 49(3): 337-345.
LI B Z, ZHAO X L, PENG L. Research advances in the development and regulation of plant chloroplasts[J]. Chinese Bulletin of Botany, 2014, 49(3): 337-345.

[3] LI Y, HEN P, HOU J H, et al. Factors influencing leaf chlorophyll content in natural forests at the biome scale[J]. Frontiers in Ecology and Evolution, 2018, 6: 64.

[4] 落莉莉, 常庆瑞, 武旭梅, 等. 夏玉米叶片光合色素含量高光谱估算[J]. 干旱地区农业研究, 2019, 37(4): 178-183.
LA L L, CHANG Q R, WU X M, et al. Hyperspectral estimation of photosynthetic pigment contents of summer maize leaves[J]. Agricultural Research in the Arid Areas, 2019, 37(4): 178-183.

[5] 齐雁冰, 楚万林, 解飞, 等. 基于高光谱的渭北旱塬区棉花冠层叶面积指数估算[J]. 干旱地区农业研究, 2017, 35(1): 114-121.
QI Y B, CHU W L, XIE F, et al. Estimating leaf area index of cotton canopy by hyperspectral reflectance in Weibei Plateau[J]. Agricultural Research in the Arid Areas, 2017, 35(1): 114-121.

[6] PINAR A, CURRAN P J. Technical note grass chlorophyll and the reflectance red edge[J]. Remote Sensing, 1996, 17(2): 351-357.

[7] 陈燕, 黄春燕, 王登伟, 等. 北疆棉花叶绿素密度的高光谱估算研究[J]. 新疆农业科学, 2006, 43(6): 451-454.
CHEN Y, HUANG C Y, WANG D W, et al. Estimation of cotton chlorophyll density in North Xinjiang by using high spectral data[J]. Xinjiang Agricultural Sciences, 2006, 43(6): 451-454.

[8] 吕玮, 李玉环, 毛伟兵, 等. 基于高光谱的小麦旗叶净光合速率的遥感反演模型的比较研究[J]. 农业资源与环境学报, 2017, 34(6): 582-586.
LV W, LI Y H, MAO W B, et al. Comparison of estimation methods for net photosynthetic rate of wheat's flag leaves based on hyperspectrum[J]. Journal of Agricultural Resources and Environment, 2017, 34(6): 582-586.

[9] 王月, 徐绍棠, 姚海燕, 等. 基于小波分析小麦旗叶净光合速率高光谱遥感反演[J]. 浙江农业科学, 2020, 61(10): 1979-1984.
WANG Y, XU S T, YAO H Y, et al. Hyperspectral remote sensing inversion of net photosynthetic rate of wheat flag leaves based on wavelet analysis[J]. Journal of Zhejiang Agricultural Sciences, 2020, 61(10): 1979-1984.

[10] 刘聪, 彭漪, 方圣辉. 基于高光谱数据的水稻叶片净光合速率反演[J]. 中国农业大学学报, 2020, 25(1): 56-65.
LIU C, PENG Y, FANG S H. Remote estimation of rice leaf net photosynthetic rate based on hyperspectral reflectance[J]. Journal of China Agricultural University, 2020, 25(1): 56-65.

[11] FROMME P, MELKOZERNOV A, JORDAN P, et al. Structure and function of photosystem I: interaction with its soluble electron carriers and external antenna systems[J]. FEBS Letters, 2003, 555(1): 40-44.

[12] HONG G, EL-HAMID H T. Hyperspectral imaging using multivariate analysis for simulation and prediction of agricultural crops in Ningxia, China[J]. Computers and Electronics in Agriculture, 2020, 122: 105355.

[13] 万升, 杨健, 宫辰. 基于图神经网络的高光谱图像分类研究进展[J]. 电子学报, 2023, 51(6): 1687-1709.
WAN S, YANG J, GONG C. Advances of hyperspectral image classification based on graph neural networks[J]. Acta Electronica Sinica, 2023, 51(6): 1687-1709.

[14] 苏红军. 高光谱遥感影像降维: 进展、挑战与展望[J]. 遥感学报, 2022, 26(8): 1504-1529.
SU H J. Dimensionality reduction for hyperspectral remote sensing: advances, challenges, and prospects[J]. Journal of Remote Sensing, 2022, 26(8): 1504-1529.

[15] 杨含成. 高光谱遥感技术在智慧农业中的应用研究[J]. 南方农机, 2022, 53(20): 156-158.
YANG H C. Application of hyperspectral remote sensing technology in intelligent agriculture[J]. South Agricultural Machinery, 2022, 53(20): 156-158.

[16] 唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. 农业工程学报, 2021, 37(2): 105-113.
TANG H T, MENG X T, SU X X, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(2): 105-113.

[17] 付乐天, 李鹏, 高莲. 考虑样本异常值的改进最小二乘支持向量机算法[J]. 仪器仪表学报, 2021, 42(6): 179-190.

- FU L T, LI P, GAO L. Improved LSSVM algorithm considering sample outliers[J]. Chinese Journal of Scientific Instrument, 2021, 42(6): 179-190.
- [18] 高升, 王巧华. 基于可见/近红外透射光谱技术的红提糖度和含水率无损检测[J]. 中国光学, 2021, 14(3): 566-577.
- GAO S, WANG Q H. Non-destructive testing of red globe grape sugar content and moisture content based on visible/near infrared spectroscopy transmission technology[J]. Chinese Journal of Optics, 2021, 14(3): 566-577.
- [19] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [20] WU S, AMARI S I. Conformal transformation of kernel functions; a data-dependent way to improve support vector machine classifiers[J]. Neural Processing Letters, 2002, 15(1): 59-67.
- [21] 韩建文, 冯春晖, 彭杰, 等. 不同分辨率无人机多光谱影像的棉花叶面积指数估测研究[J]. 棉花学报, 2022, 34(4): 338-349.
- HAN J W, FENG C H, PENG J, et al. Estimation of leaf area index of cotton from unmanned aerial vehicle multispectral images with different resolutions[J]. Cotton Science, 2022, 34(4): 338-349.
- [22] 史舟, 梁宗正, 杨媛媛, 等. 农业遥感研究现状与展望[J]. 农业机械学报, 2015, 46(2): 247-260.
- SHI Z, LIANG Z Z, YANG Y Y, et al. Status and prospect of agricultural remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(2): 247-260.
- [23] 田春燕, 黄春燕, 郭晓飞, 等. 基于植被指数的棉花冠层光合有效辐射截获量和叶片净光合速率估算研究[J]. 新疆农业科学, 2017, 54(6): 981-987.
- TIAN C Y, HUANG C Y, GUO X F, et al. Estimation of cotton canopy fractional interception of absorbed photosynthetic active radiation and leaf net photosynthetic rate based on hyperspectral vegetation index[J]. Xinjiang Agricultural Sciences, 2017, 54(6): 981-987.
- [24] 刘贞琦, 刘振业, 马达鹏, 等. 水稻叶绿素含量及其与光合速率关系的研究[J]. 作物学报, 1984, 10(1): 57-62.
- LIU Z Q, LIU Z Y, MA D P, et al. Study on the relationship between chlorophyll content and photosynthetic rate in rice[J]. Acta Agronomica Sinica, 1984, 10(1): 57-62.
- [25] 宁丽云, 张巨松, 王丽, 等. 灌溉定额对不同品种海岛棉光合特性影响及抗旱性评价[J/OL]. 灌溉排水学报, 2024; 1-11(2024-12-03). <https://doi.org/10.13522/j.cnki.ggps.2024232>.
- NING L Y, ZHANG J S, WANG L, et al. Effects of irrigation quota on photosynthetic characteristics and drought resistance evaluation of different varieties of island cotton[J]. Journal of Irrigation and Drainage, 2024; 1-11(2024-12-03). <https://doi.org/10.13522/j.cnki.ggps.2024232>.
- [26] 段鹏飞, 赵地. DPC 与 DTA-6 复配对棉叶衰老的生理特性、产量与品质的影响[J]. 中国农业科技导报, 2018, 20(3): 104-114.
- DUAN P F, ZHAO D. Effects of DPC and DTA-6 on leaf senescence physiological characteristics, yield and quality[J]. Journal of Agricultural Science and Technology, 2018, 20(3): 104-114.
- [27] 汪六三, 黄子良, 王儒敬. 基于近红外光谱和机器学习的大豆种皮裂纹识别研究[J]. 农业机械学报, 2021, 52(6): 361-368.
- WANG L S, HUANG Z L, WANG R J. Identification of soybean seed coat crack based on near infrared spectroscopy and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 361-368.
- [28] ZHAO J L, FANG Y, CHU G M, et al. Identification of leaf-scale wheat powdery mildew (*Blumeria graminis* f. sp. *tritici*) combining hyperspectral imaging and an SVM classifier[J]. Plants (Basel, Switzerland), 2020, 9(8): 936.
- [29] SOARES S F C, GOMES A A, ARAUJO M C U, et al. The successive projections algorithm[J]. TrAC Trends in Analytical Chemistry, 2013, 42: 84-98.
- [30] ZHANG J J, CHENG T, GUO W, et al. Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods[J]. Plant Methods, 2021, 17(1): 49.
- [31] CHAN A, BARNES C, SWINFIELD T, et al. Monitoring ash dieback (*Hymenoscyphus fraxineus*) in British forests using hyper spectral remote sensing[J]. Remote Sensing in Ecology and Conservation, 2020, 7(2): 306-320.
- [32] 刘娣, 孙佳倩, 余钟波. 基于机器学习模型的多层土壤湿度反演[J]. 河海大学学报(自然科学版), 2024, 52(3): 7-14.
- LIU D, SUN J Q, YU Z B. Multi-layer soil moisture inversion based on machine learning models[J]. Journal of Hohai University(Natural Sciences), 2024, 52(3): 7-14.
- [33] 温学兵, 谢维, 姚佳宜. 基于随机森林和支持向量机模型的期刊评价[J]. 沈阳师范大学学报(自然科学版), 2022, 40(2): 174-179.
- WEN X B, XIE W, YAO J Y. Journal evaluation based on random forest and support vector machine model[J]. Journal of Shenyang Normal University (Natural Science Edition), 2022, 40(2): 174-179.
- [34] 李冬旺, 张永江, 刘连涛, 等. 干旱胁迫对棉花冠层光合、光谱和荧光的影响[J]. 棉花学报, 2018, 30(3): 242-251.
- LI D W, ZHANG Y J, LIU L T, et al. Responses of canopy photosynthesis, spectral indices and solar-induced chlorophyll fluorescence in cotton under drought stress[J]. Cotton Science, 2018, 30(3): 242-251.