

水稻产量遥感监测机器学习算法对比

竞霞¹, 张杰^{1,2}, 王娇娇², 明世康², 傅友强³, 冯海宽², 宋晓宇^{2*}

1. 西安科技大学测绘科学与技术学院, 陕西 西安 710054
2. 北京市农林科学院信息技术研究中心, 北京 100094
3. 广东省农业科学院水稻研究所, 广东 广州 510640

摘要 水稻是我国的主要粮食作物, 利用高光谱遥感技术在水稻未成熟之前对水稻产量进行监测, 一方面可以及时调整栽培管理方式, 指导合理追肥, 另一方面, 可以准确掌握水稻的产量信息, 帮助政府提前做出决策。以2019年—2020年广州市白云区钟落潭试验基地氮肥梯度实验为基础, 分别获取水稻分化期和抽穗期冠层高光谱数据、作物群体长势参数(生物量、叶面积指数)及作物养分吸收量, 利用贝叶斯岭回归(BRR)、支持向量回归(SVR)、偏最小二乘回归(PLSR)三种方法建立各生育期的产量监测模型并进行精度对比, 确定水稻产量的最优估算时期和最佳估测模型。结果表明, 三种方法中, BRR和SVR方法更适合产量监测, 在不同时期及不同的参数组合下均有较好的表现($R^2 > 0.82$, $NRMSE < 8.22\%$); 基于2019年与2020年数据, 采用全波段光谱信息进行产量监测时, 分化期最佳监测模型为BRR模型, R^2 为0.90, 抽穗期最优监测模型为SVR模型, R^2 为0.87; 采用全波段光谱协同作物群体长势参数进行产量监测时, 两时期最佳监测模型均为BRR模型, R^2 分别达到0.90和0.92; 相较于BRR模型和SVR模型, PLSR模型在不同时期和不同参数组合下, 最高 R^2 仅为0.75; 基于2020年数据, 以三种不同的参数组合作为输入时, 两时期估算结果均为BRR模型最优, 且分化期建模精度高于抽穗期(R^2 至少增加0.02, $NRMSE$ 至少降低0.61%); 当输入参数组合为全波段光谱协同作物群体长势参数、作物养分吸收量时, BRR模型对产量的估算精度达到最高, R^2 为0.94。分析认为产量的最优监测时期是分化期, 最优监测模型为BRR模型。研究结果可为水稻产量的早期遥感监测提供参考。

关键词 高光谱遥感; 水稻估产; 贝叶斯岭回归; 支持向量回归

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)05-1620-08

引言

水稻是我国种植面积最大、覆盖范围最广的粮食作物。获取高产、高品质的水稻是我国实施精准农业的新要求^[1]。基于高光谱信息的作物生长参数监测, 是农业遥感的研究热点。建立水稻收获前的产量监测模型, 是精准制定田间管理措施、快速评估水稻产量的可靠依据^[2]。

大量学者已利用高光谱遥感对水稻等作物产量预测进行了深入研究^[3]。薛利红等在实测水稻高光谱反射率的基础上, 构建了基于光谱植被指数—叶面积氮指数的复合水稻产量估测模型, 取得了较好的估测精度^[4]。宋红燕等发现水稻植株氮含量的敏感波段是552和890 nm, 并基于氮素敏感波

段构建了比值植被指数(ratio vegetation index, RVI)和绿色归一化植被指数(green normalized vegetation index, GNDVI), 依此构建的产量监测模型决定系数(R^2)达到0.724^[5]。Kawamura等使用迭代逐步消除偏最小二乘法(iterative stepwise elimination partial least squares, ISE-PLS)对400~930 nm之间的波段进行逐步消除选择产量预测敏感波段, 建立水稻收获前产量估算模型, 模型最大决定系数为0.873, 认为水稻孕穗期为最佳估算时期^[6]。冯伟等对水稻叶片氮素参数、冠层光谱等参数进行多时期相关性分析, 认为灌浆前期叶片氮素积累量和叶面积氮指数能够反映水稻籽粒产量状况, 实验结果表明利用水稻灌浆前期特征光谱参数和拔节—成熟期特征光谱指数的累积值能够稳定预报水稻成熟期籽粒产量^[7]。目前关于高光谱数据的研究大多通过相关性分析等

收稿日期: 2021-09-30, 修订日期: 2022-02-25

基金项目: 广东省重点领域研发计划项目(2019B020214002), 国家自然科学基金项目(42171394)资助

作者简介: 竞霞, 1978年生, 西安科技大学测绘科学与技术学院副教授 e-mail: jingxiaxust@163.com

* 通讯作者 e-mail: songxy@nercita.org.cn

方法实现对数据的特征降维,去除高光谱数据中的冗余信息,以较少的特征参数参与模型的构建,在降低模型构建难度的同时,可能会丢失部分有效信息。而机器学习模型具备强大的处理高维数据与冗余数据的能力,能够有效抵抗高维数据中存在的噪声,基于更高效益的数学方法与数据处理方式实现对数据中有效信息的提取^[8-9]。Bao等在小麦品种的快速分类模型的构建中,使用连续投影算法(successive projections algorithm, SPA)、主成分分析算法(principal component analysis, PCA)和随机蛙跳(random frog, RF)三种特征提取方法,从数百个光谱波段中筛选可用于建立分类模型的光谱变量,使用线性判别分析(linear discriminant analysis, LDA)支持向量机(support vector machine, SVM)、极限学习机(extreme learning machine, ELM)三种机器学习算法分别以全波段和经过特征筛选的波段作为输入变量进行小麦品种分类模型构建,以全波段作为输入变量的ELM算法分类精度最优^[10]。

水稻生长过程中,植株本身在生长过程中水分供应、光热吸收、土壤养分固定等都会影响水稻的产量^[11]。本文根据前人研究,选择水稻分化期与抽穗期冠层全波段光谱数据,筛选与产量相关的植株地上生物量、叶面积指数,土壤养分参数等影响因子,利用贝叶斯岭回归(Bayesian ridge regression, BRR)、支持向量回归(support vector regression, SVR)、偏最小二乘回归(partial least square regression, PLSR),分别构建:(1)基于全波段光谱信息的产量监测模型;(2)基于全波段光谱信息协同长势参数的产量监测模型;(3)基于全波段光谱信息协同长势参数、作物养分吸收量的产量监测模型。探索不同算法在不同的输入参数下的适应性,筛选最优变量因子,为水稻产量遥感监测提供依据。

1 实验部分

1.1 试验设计

2019年—2020年在广东省广州市白云区钟落潭试验基地(23°23'24"N—23°23'59"N, 113°25'48"E—113°26'24"E)开展水稻变量施肥的小区实验。试验基地内,2019年的试验品种为美香占2号(V1),插秧时间2019年8月8日,插秧密度为20 cm×20 cm。共设计15个小区采样测试;每个小区插秧规格为16穴×16穴。2020年的试验品种为美香占2号(V1)和五丰优615(V2),插秧时间为2020年8月8日,共30个小区,插秧密度为20 cm×20 cm,根据插秧规格,每个小区16穴×20穴。

2019年及2020年试验共设计5个氮素水平(N0, N1, N2, N3, N4),分别为0, 60, 120, 180和240 kg N·ha⁻¹,设3次重复;其中基肥、分蘖肥、穗肥的施用比例为5:2:3;磷、钾肥用量分别为54和144 kg·ha⁻¹。

2019年和2020年分别于分化期(2019-09-13, 2020-09-10)和抽穗期(2019-10-11, 2020-10-09)进行水稻地上生物量(above ground biomass, AGB)、叶面积指数(LAI)、土壤营养参数及其他水稻品质相关数据的田间采集,其中2019年每时期获取15个样本,2020年每时期获取30个样本。数据

获取情况见表1。

表1 试验数据获取

Table 1 Test data acquisition

参数	2019年		2020年	
	分化期	抽穗期	分化期	抽穗期
冠层光谱	√	√	√	√
长势参数(AGB, LAI)	√	√	√	√
土壤养分参数	未获取		√	√

1.2 光谱测定与预处理

水稻冠层高光谱数据采集使用美国ASD Filed Spec Pro 2500背挂式野外光谱仪,光谱范围为350~2500 nm。根据前人的研究,作物光谱的可见光与近红外范围已能够反映作物的生长状况,因此本次实验采用454~950 nm的冠层光谱数据,重采样后间隔为4 nm^[12-13]。测量时间为北京时间10:00—14:00,期间天气晴朗,在每一个采样点测量前后均用标准白板对冠层辐亮度数据进行校正。探头距离冠层高度约1 m,垂直向下,探头角度为25°;每个采样点取10次测量平均值作为该样方的冠层辐亮度值。同一年的试验中,记录采样点的位置,保证不同生育期同一小区在相同位置采集数据。对测定的冠层辐亮度和白板辐亮度利用式(1)计算目标的光谱反射率。

$$R = \frac{L_{\text{target}}}{L_{\text{board}}} \times R_{\text{board}} \quad (1)$$

式(1)中, R 为冠层反射率, L_{target} 为冠层辐亮度($\mu\text{W} \cdot \text{cm}^{-2} \cdot \text{nm}^{-1} \cdot \text{sr}^{-1}$), L_{board} 为白板辐亮度($\mu\text{W} \cdot \text{cm}^{-2} \cdot \text{nm}^{-1} \cdot \text{sr}^{-1}$), R_{board} 为白板反射率。

1.3 叶面积指数与生物量

测量光谱后,在小区内随机选择6穴水稻植株样本,去根并逐丛计数茎蘖数,分化期茎叶分离,抽穗期将茎叶和穗分离,测定叶面积2000 cm²(S)左右,将其烘干后记录重量(w_1),余叶也一并烘干称重(w_2),利用以式(2)计算叶面积指数(LAI)

$$\text{LAI} = \frac{(w_1 + w_2) \times S}{\gamma \times w_1} \times D \quad (2)$$

式(2)中, γ 为取样植株样品穴数, D 为种植密度,由小区内水稻种植穴数除以小区面积得到。

根据采样点种植密度和水稻样本的干重计算单位面积植株的地上生物量(above ground biomass, AGB, $\text{g} \cdot \text{m}^{-2}$),计算公式如式(3)

$$\text{AGB} = \frac{W_L + W_S + W_E}{\gamma} \times D \quad (3)$$

式(3)中, W_L , W_S 和 W_E 分别为水稻测试样本叶片、茎、穗的干重(g), γ 为取样植株样品穴数, D 为种植密度。

1.4 土壤养分测定及作物养分吸收量计算

2020年水稻抽穗期(2020年10月10日)及收获后(2020年11月25日)分别于不同氮肥处理小区获取田间0~20 cm土样,测试土壤理化性质,根据2020年试验小区施肥量、土壤残留养分含量计算了不同小区作物可吸收养分量,如式(4)所示

作物氮(磷、钾)肥养分吸收量=施肥氮(磷、钾)肥养分总量-土壤氮(磷、钾)养分残留量 (4)

其中施肥养分量由试验中所施用氮、磷、钾肥料中有效养分含量百分比折算,本试验中氮肥尿素中 N 含量 46%,钾肥氯化钾中 K 含量 60%,磷肥磷酸钙中 P 含量 12%;土壤养分残留量由土壤养分测试指标与耕层土质量相乘得到,本研究中耕层土质量按照耕层厚度 20 cm 计算,研究中未考虑肥料挥发及淋滤损失量。

1.5 产量测定

于成熟期逐小区实收 125 丛稻株(5 m²)测产,将稻谷风干,取 100 g 左右于 105 °C 下烘干 48 h,测定含水量,然后将稻谷转换成含水量为 14%的稻谷产量。

1.6 数据分析方法

偏最小二乘回归(partial least square regression, PLSR)通过最小化误差的平方和,寻找一组新的潜在变量来解释自变量 X 与因变量 Y 之间的统计关系,并且可以在建模过程中实现对数据的主成分分析、典型相关性分析和回归分析,是一种常见的对数据进行降维处理、解决数据多重共线性问题、简化建模过程的线性回归方式^[14-15]。贝叶斯岭回归(Bayesian ridge regression, BRR)是基于贝叶斯方法与最小二乘法的改进而提出的,通过对线性贝叶斯回归模型加入 L2 正则化,结合相关参数的先验信息形成先验分布并给出预估数值^[16]。支持向量回归(support vector regression, SVR)的基本思想是通过寻找最优划分超平面,忽略小于偏差 ϵ 的样本,对其他样本进行回归;偏差 ϵ 的引入是 SVR 区别于传统回归模型的地方,即以预测 y 值为中心,与真实 y 值之间存在一个宽度为 2ϵ 的区域,在此区域内,预测 y 值与真实 y 值的差别认为是 0。其回归模型为 $f(x) = w^T x + b$, w 和 b 为模型待确定参数^[17]。它能够较好的完成自变量与因变量之间的非线性回归。

1.7 模型验证方法

利用 PLSR, BRR 和 SVR 三种算法构建水稻产量估算模型时,采用 k-fold 交叉验证方法进行建模,即将样本集随机分为 k 组子数据集,轮流使用其中的 $k-1$ 份子数据集建

模,另一份作为验证, k 次建模后的均值为模型的精度。采用决定系数(R^2)和归一化均方根误差(normalized root mean square error, NRMSE)两个指标联合验证模型预测精度, R^2 越大,代表模型拟合度更高, NRMSE 越小,模型稳定性越好, NRMSE<10%表示模型具有较好的稳定性。

2 结果与讨论

2.1 水稻产量的关键影响因子分析

2.1.1 水稻产量与不同生育期 AGB、LAI 的相关性分析

2019 年与 2020 年合并后的水稻产量与 AGB 和 LAI 进行相关性分析可知, AGB, LAI 与产量在分化期相关系数分别为 0.809 和 0.678,抽穗期 AGB 和 LAI 与产量的相关性则是有所下降,分别为 0.635, 0.590,但都达到了 0.001 水平显著($r(0.001) = 0.474, n = 45$)。2020 年水稻产量与 AGB、LAI 进行相关性分析,分化期与抽穗期所有长势参量均达到了 0.001 水平显著($r(0.001) = 0.570, n = 30$),两时期比较,分化期相关系数高于抽穗期, AGB、LAI 与产量在分化期相关系数分别为 0.596 和 0.839,抽穗期相关系数分别为 0.586, 0.696。同时, LAI 与 AGB 同样具有较高的相关性, LAI 代表叶片尺度植株的生理状态, AGB 可表示植株地上部分整体的发育状况;由于水稻产量的形成是一个动态且复杂的生物学过程,在同一生育时期,不同的长势参数可能对水稻的产量形成具备不同的作用机理,故建模过程中同时加入 AGB 和 LAI 两个长势参量探究其对产量形成的影响。

2.1.2 水稻产量作物养分吸收量的相关性分析

于 2020 年 10 月 10 日,水稻分化肥施用 1 个月之后,以及 11 月 23 日,水稻成熟收获后,分别采集不同氮肥处理小区土壤数据,进行土壤养分测试,获取小区土壤碱解氮(mg·kg⁻¹)、速效钾(mg·kg⁻¹)及有效磷(mg·kg⁻¹)含量数据。根据小区施肥量数据及土壤养分残留含量数据,分别计算了作物养分吸收量,并分析了水稻抽穗期及成熟期养分吸收量与 AGB、LAI 及产量的相关关系,结果如表 2 所示。

表 2 2020 年 AGB、LAI 及产量与作物养分吸收量的相关性($n=30$)

Table 2 Correlation between biomass, LAI, yield and crop nutrient uptake in 2020

作物养分吸收量/ (kg·hm ⁻²)	相关系数 r				
	产量	AGB		LAI	
		分化期	抽穗期	分化期	抽穗期
至抽穗期作物吸收 N 量	0.713**	0.797**	0.623**	0.789**	0.731**
至抽穗期作物吸收 P 量	-0.086	-0.085	-0.181	-0.093	-0.098
至抽穗期作物吸收 K 量	0.526*	0.536*	0.425	0.591**	0.490*
至成熟期作物吸收 N 量	0.723**	0.595**	0.609**	0.818**	0.716**
至成熟期作物吸收 P 量	-0.242	-0.288	-0.415	-0.254	-0.378
至成熟期作物吸收 K 量	0.402	0.380	0.275	0.478*	0.341

注: $r(0.01) = 0.463, r(0.001) = 0.570$, * 表示通过 0.01 水平显著性检验, ** 表示通过 0.001 水平显著性检验

Note: $r(0.01) = 0.463, r(0.001) = 0.570$, * significant difference at 0.01 level; ** significant difference at 0.001 level

抽穗期作物养分吸收量与产量的相关系数分别为 0.713, -0.086 和 0.526;成熟期作物养分吸收量与产量的

相关系数分别为 0.723, -0.242 和 0.402,其中两时期作物吸收 N 与作物吸收 K 和产量具有较好的相关性,作物吸收 P

则与产量线性关系不显著，但不能排除是否存在其他关系；分化期与抽穗期的 AGB、LAI 和两时期作物吸收 N 均达到了极显著相关，最大相关系数为 0.818。由于成熟期作物吸收 N、P 和 K 量在水稻成熟后才可获得，难以实现水稻产量早期监测，故模型中土壤养分参数仅使用至抽穗期作物吸收 N、P 和 K 量。

2.1.3 冠层光谱与水稻长势参数的相关性分析

如图 1、图 2 所示，对水稻分化期与抽穗期冠层光谱与长势参数进行相关性分析，2019 年与 2020 年所有长势参数在两生育期相关系数有所不同，但均具有相似的变化趋势，近红外部分均保持在某一值持平，整体变化幅度很小且以正相关为主，可见光部分则以负相关为主。所有长势参数在 550 nm 附近出现相关性“低谷”，相关系数低于其他可见光部分。在可见光区域与近红外区域的交界处，光谱反射率受叶片内细胞间隙折射率不同的影响，反射率急剧增加，相关系数迅速由负转正，有明显的降低后再抬升的趋势。

图 1(a)和(b)是将 2019 年与 2020 年数据分生育期合并后的冠层光谱与长势参数的相关系数，LAI 与 AGB 在两时期具有相似的变化趋势，在可见光部分，抽穗期 LAI 与 AGB 同冠层光谱的相关性较差，均未超过 0.4，分化期明显优于抽穗期，尤其在 602~690 nm 之间，冠层光谱与 LAI 最大相关系数达到 -0.685(666 nm)，与 AGB 最大相关系数达到 -0.762(662 nm)，在近红外部分，则是抽穗期优于分化期且 LAI 与冠层光谱的相关性更高，LAI 与 AGB 的最大相关系数分别为 0.708(922 nm)、0.584(922 nm)。图 2(a)和(b)为 2020 年两生育期冠层光谱与长势参数的相关性，LAI 与冠层反射率相关性全波段抽穗期高于分化期，在近红外区域相差 0.1 左右，分化期最大相关系数 -0.623(666 nm)，抽穗期最大相关系数为 0.753(886 nm)，AGB 与冠层反射率相关性在可见光区域分化期高于抽穗期，而在近红外区域抽穗期更高，分化期最大相关系数为 -0.665(662 nm)，抽穗期最大相关系数为 0.616(878 nm)。

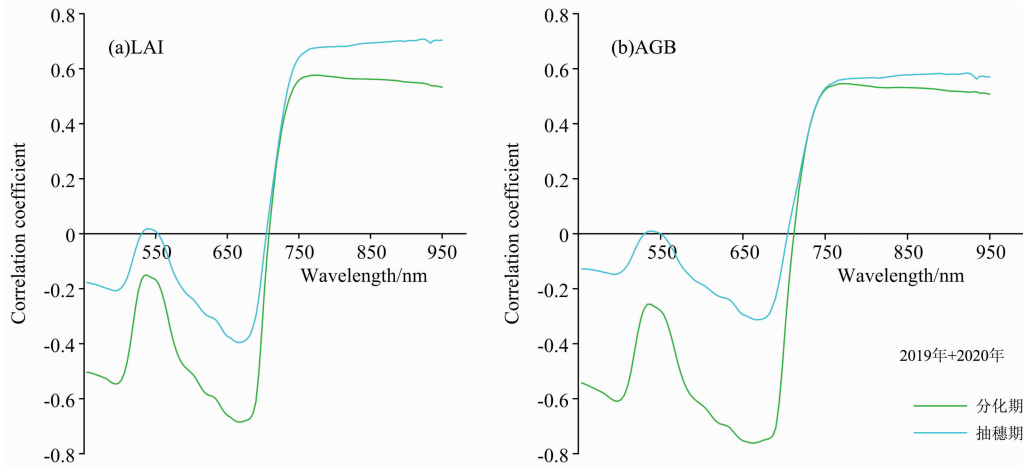


图 1 2019 年与 2020 年水稻不同生育期冠层光谱与长势参数相关性

Fig. 1 Correlation between canopy spectra and growth parameters at different growth stages of rice in 2019 and 2020

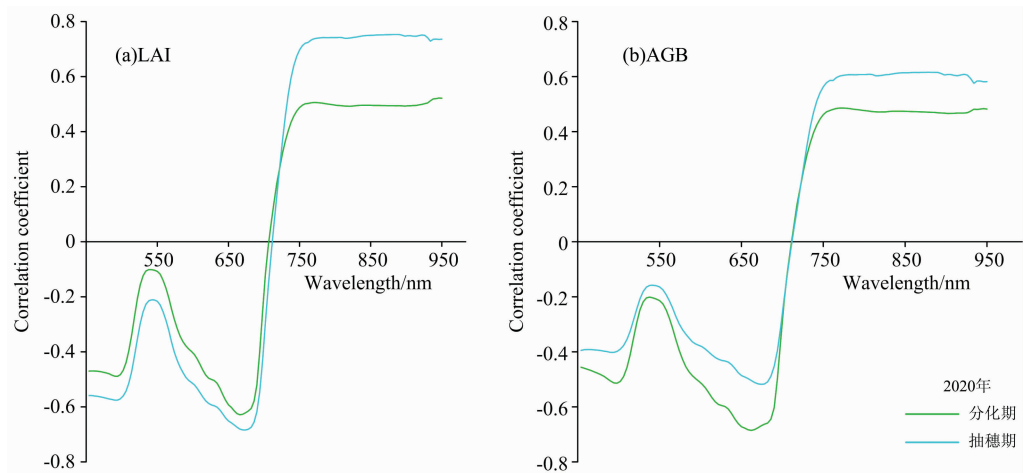


图 2 2020 年水稻不同生育期冠层光谱与长势参数相关性

Fig. 2 Correlation between canopy spectra and growth parameters at different growth stages of rice in 2020

LAI 与 AGB 在冠层光谱的大部分区域均表现出较强的相关性, 可以认为 LAI 与 AGB 参数能够被冠层光谱较好的表达, 而 LAI 与 AGB 和产量的相关性达到极显著水平, 故利用作物生长前期的光谱数据进行产量监测是可行的。

2.2 水稻产量模型构建与分析

分别使用 BRR, SVR 和 PLSR 构建水稻产量预测模型, 基于 2019 年与 2020 年合并后分化期 ($n=45$) 和抽穗期 ($n=45$) 数据, 构建基于全波段光谱、全波段光谱+AGB、LAI 的产量预测模型; 基于 2020 年分化期 ($n=30$) 和抽穗期 ($n=30$) 数据, 构建基于全波段光谱、全波段光谱+AGB、LAI、全波段光谱+AGB、LAI+作物养分吸收量的产量预测模型; 表 3 为三种方法使用 2019 年与 2020 年数据建模的决定系数 (R^2) 与归一化均方根差 (NRMSE), 表 4 为三种方法使用 2020 年数据建模的决定系数 (R^2) 与归一化均方根差 (NRMSE)。

2.2.1 基于光谱信息与农学参量的水稻产量预测

由表 3 可知, 以全波段光谱作为输入参数时, 三种方法所建立的产量预测模型, 分化期精度优于抽穗期, BRR 模型在分化期 R^2 达到了 0.897 1, NRMSE 为 6.76%, 较抽穗期 R^2 高 0.032 4, NRMSE 低 1%; SVR 模型在分化期 R^2 为 0.877 6, NRMSE 为 7.59%, 抽穗期 R^2 为 0.865 5, 两时期预测精度与模型稳定性接近, 而 PLSR 模型在分化期的 R^2 较抽穗期相差较大, 相差 0.199, NRMSE 相差 3.6%。以全波段光谱+AGB、LAI 作为输入参数时, BRR 模型与 SVR

模型抽穗期精度则优于分化期, 其中 BRR 模型在分化期 R^2 为 0.903 0, NRMSE 为 6.57%, 在抽穗期 R^2 为 0.915 2, NRMSE 为 6.24%, R^2 均达到 0.9 以上; SVR 模型在分化期较仅输入全波段的模型精度反而有所下降, R^2 降低 0.027 5, 抽穗期模型较仅输入全波段的模型精度有提升, 但提升不大; PLSR 模型则是在分化期表现较好, 但仍未超过 BRR 模型与 SVR 模型的预测精度。从整体来看, BRR 模型在仅使用全波段作为输入参量时已取得较好的预测精度, 加入长势参数时, 离群点变的更少, 所有点基本位于 0 误差线两侧, 对产量中的较大与较小值均有较好的预测效果。

表 3 基于 2019 与 2020 年数据的模型精度 ($n=45$)

Table 3 Model accuracies based on the data in 2019 and 2020

参数	模型	精度			
		分化期		抽穗期	
		R^2	NRMSE /%	R^2	NRMSE /%
全波段	BRR	0.897 1	6.76	0.864 7	7.76
	SVR	0.877 6	7.59	0.865 5	7.82
	PLSR	0.754 1	10.44	0.555 1	14.04
全波段+AGB, LAI	BRR	0.903 0	6.57	0.915 2	6.24
	SVR	0.850 1	8.22	0.870 9	8.01
	PLSR	0.752 0	10.48	0.560 7	13.95

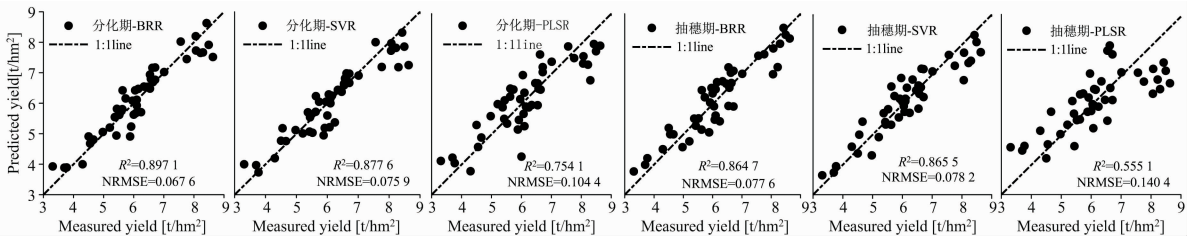


图 3 基于全波段光谱的产量预测模型

Fig. 3 Yield prediction model based on full band spectra

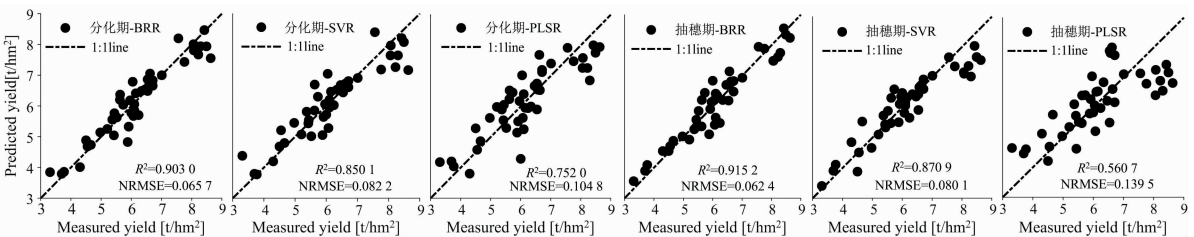


图 4 基于全波段光谱—长势参数的产量预测模型

Fig. 4 Yield prediction model based on full band spectrum-growth parameters

2.2.2 基于光谱信息-农学参量及作物养分吸收量的水稻产量预测

基于 2020 年分化期与抽穗期数据所建的产量预测模型, 以全波段光谱作为输入参数时, 三种模型在分化期的模型精度优于抽穗期模型精度; 以全波段光谱+AGB、LAI 作为输入参数时, 分化期 BRR 模型 R^2 提高至 0.925 0, 抽穗期

BRR 模型 R^2 提高至 0.905 7, SVR 模型与 PLSR 模型较仅输入全波段时未有较大提升, 但 SVR 模型仍具有良好的估测精度; 以全波段+AGB、LAI+作物养分吸收量作为输入参量时, BRR 模型在分化期 R^2 达到最大, 为 0.940 3, NRMSE 为 4.34%, 抽穗期模型 R^2 为 0.922 4, NRMSE 为 4.95%, 散点分布均匀, 在分化期模型估算精度更好, SVR 模

型则是在抽穗期模型估算精度更好。PLSR 模型在分化期与抽穗期表现均低于 BRR 模型与 SVR 模型，且离群点更明显。

表 4 基于 2020 年数据的模型构建精度 (n=30)

Table 4 Model construction accuracy based on the data in 2020

参数	模型	精度			
		分化期		抽穗期	
		R ²	NRMSE / %	R ²	NRMSE / %
全波段	BRR	0.902 2	5.54	0.849 3	6.89
	SVR	0.836 2	7.30	0.827 6	7.41
	PLSR	0.644 8	10.53	0.549 5	11.85
全波段+ AGB, LAI	BRR	0.925 0	4.86	0.905 7	5.45
	SVR	0.837 0	7.26	0.845 7	7.36
	PLSR	0.654 6	10.38	0.659 8	11.58
全波段+ AGB, LAI+ 作物养分吸收量	BRR	0.940 3	4.34	0.922 4	4.95
	SVR	0.845 7	7.03	0.889 0	6.39
	PLSR	0.659 8	10.30	0.575 3	11.51

3 结论

分别以 2019 年与 2020 年分化期与抽穗期水稻全波段光

谱、全波段光谱+作物长势参数、全波段光谱+作物长势参数+作物养分吸收量作为自变量，采用 BRR, SVR 和 PLSR 三种回归模型对水稻产量进行建模估计，选取最佳估算模型与最佳估算时期。实验结果表明，利用分化期与抽穗期冠层光谱监测当季水稻产量是可行的；三种方法中，BRR 模型能够更有效的对水稻产量进行估计，各生育期估算均达到最优，基于 2019 年与 2020 年数据的模型，最大 R² 为 0.915 2，基于 2020 年数据，最大 R² 为 0.940 3，在仅输入全波段光谱的时候也取得了较好的估算结果；加入水稻长势参数、作物氮素吸收量后，能够有效提高模型的拟合优度与稳定性，这是由于水稻产量与作物不同时期吸收和转化来自土壤和环境的养分存在密切关系，合理优化作物养分吸收能够有效提高作物产量。综合对比下，水稻产量的最佳估算时期是分化期；BRR 和 SVR 两种机器学习模型的估算精度在两种监测模型中均优于 PLSR，虽然 PLSR 选取了较少的参数，降低了模型复杂度，但是也丢失了部分用来估算水稻产量的信息波段，而两种机器学习模型利用所有信息进行回归建模，较大程度的保留了与水稻产量相关的信息，且简化了数据处理流程；在不同时期与不同输入参数下，基于先验分布的 BRR 模型能够更好的解决高光谱数据中的多重共线性问题，对于数据中存在的噪声能够有效抵抗，在具有上百个波段的高光谱数据的输入下，保持了模型的稳定性，有利于模型的推广。

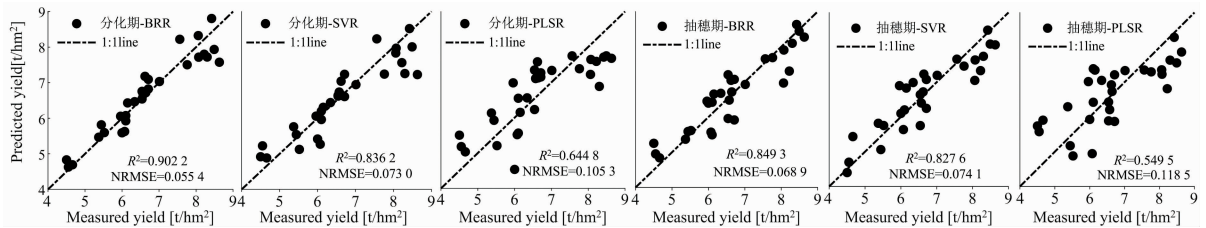


图 5 基于全波段光谱的产量预测模型

Fig. 5 Yield prediction model based on full band spectrum

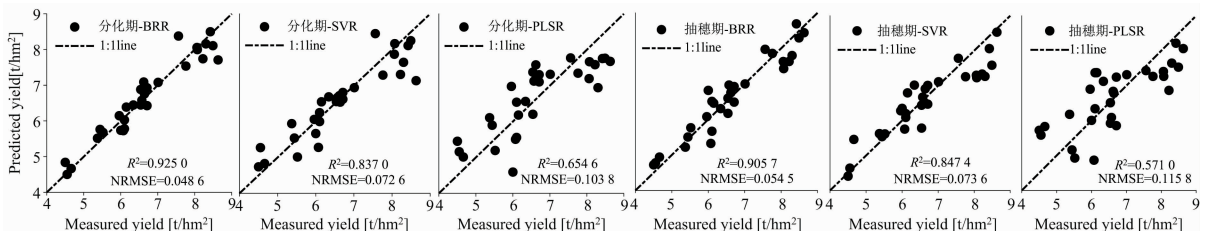


图 6 基于全波段光谱—长势参数的产量预测模型

Fig. 6 Yield prediction model based on full band spectrum-growth parameters

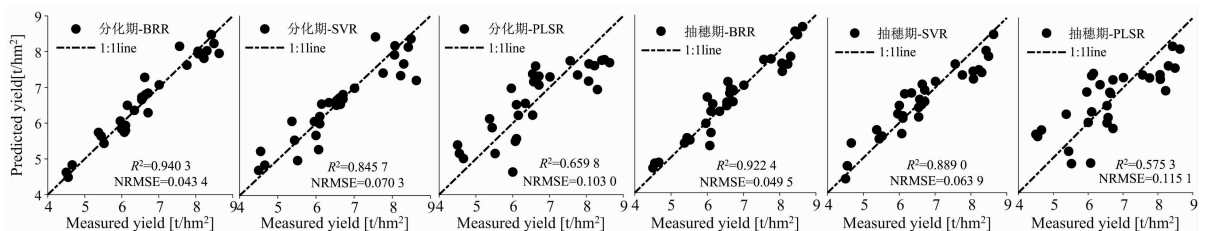


图 7 基于全波段光谱—长势参数—作物养分吸收量的产量预测模型

Fig. 7 Yield prediction model based on full band spectrum-growth parameters-crop nutrient absorption

本工作仅选用了 BRR, SVR 和 PLSR 三种方法构建水稻产量的模型, 未能选用更多的方法进行监测模型建立, 若增加其他方法, BRR 方法是否还能保持最优还有待研究。此外, 将不同参数作为自变量进行输入时只是简单拼接, 未考

虑不同参数的权重分布, 不同参数与产量是否存在最优映射的关系有待下一步研究。全波段建模方法需要算法具有良好的噪声抵抗能力, 进一步提高算法的噪声抵抗能力以及稳定性, 有利于获取更好的监测精度。

References

- [1] ZHANG Han, ZHAO Xiao-min, GUO Xi, et al(张 晗, 赵小敏, 郭 熙, 等). *Jiangsu Agricultural Science*(江苏农业科学), 2018, 46(12): 1.
- [2] JIANG Huan-yu, YING Yi-bin, XIE Li-juan(蒋焕煜, 应义斌, 谢丽娟). *Spectroscopy and Spectral Analysis*(光谱学与光谱分析), 2008, 28(6): 1300.
- [3] Liu X D, Sun Q H. *International Journal of Pest Management*, 2016, 62(3): 205.
- [4] XUE Li-hong, CAO Wei-xing, LUO Wei-hong(薛利红, 曹卫星, 罗卫红). *Journal of Remote Sensing*(遥感学报), 2005, (1): 100.
- [5] SONG Hong-yan, HU Ke-lin, PENG Xi(宋红燕, 胡克林, 彭 希). *Journal of China Agricultural University*(中国农业大学学报), 2016, 21(8): 27.
- [6] Kawamura K, Ikeura H, Phongchanmaixay S, et al. *Remote Sensing*, 2018, 10(8): 1249.
- [7] FENG Wei, ZHU Yan, TIAN Yong-chao, et al(冯 伟, 朱 艳, 田永超, 等). *Journal of Wheat Crops*(麦类作物学报), 2007, (6): 1076.
- [8] Bennett K P, Parrado-Hernández E. *Journal of Machine Learning Research*, 2006, 7: 1265.
- [9] Liakos K G, Busato P, Moshou D. *Sensors*, 2018, 18(8): 2674.
- [10] Bao Y, Mi C, Wu N, et al. *Applied Sciences*, 2019, 9(19): 4119.
- [11] ZHANG Fu-suo, CUI Zhen-ling, WANG Ji-qing, et al(张福锁, 崔振岭, 王激清, 等). *Botanical Bulletin*(植物学通报), 2007, (6): 687.
- [12] Curran P J, Dungan J L, Macler B A, et al. *Remote Sensing of Environment*, 1992, 39(2): 153.
- [13] ZHAO Xiao-min, SUN Xiao-xiang, WANG Fang-dong, et al(赵小敏, 孙小香, 王芳东, 等). *Journal of Jiangxi Agricultural University*(江西农业大学学报), 2019, 41(1): 1.
- [14] Davari M, Karimi S A, Bahrami H A, et al. *CATENA*, 2021, 197: 104987.
- [15] dos Santos F R, de Oliveira J F, Bona E, et al. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2021, 175: 106016.
- [16] Assaf A G, Tsionas M, Tasiopoulos A. *Tourism Management*, 2019, 71: 1.
- [17] DING Shi-fei, QI Bing-juan, TAN Hong-yan(丁世飞, 齐丙娟, 谭红艳). *Journal of the University of Electronic Science and Technology*(电子科技大学学报), 2011, 40(1): 2.

Comparison of Machine Learning Algorithms for Remote Sensing Monitoring of Rice Yields

JING Xia¹, ZHANG Jie^{1,2}, WANG Jiao-jiao², MING Shi-kang², FU You-qiang³, FENG Hai-kuan², SONG Xiao-yu^{2*}

1. School of Surveying and Mapping Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China

2. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100094, China

3. Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

Abstract Rice is China's major grain food crop, grown mainly in the Yangtze River valley and southern China and on the Yunnan-Guizhou Plateau. The use of hyperspectral remote sensing technology to monitor rice yield before it matures is important. It can promptly adjust cultivation management methods and guide reasonable fertilization, and accurately grasp rice yield information to help the government make decisions in advance. In this study, the nitrogen fertilizer gradient experiment was carried on 2019—2020 at Zhongluotan experimental base in Baiyun District, Guangzhou City. The rice canopy hyperspectral data, crop population growth parameters (plant above-ground biomass (AGB) and leaf area index (LAI)) and crop nitrogen nutrient intakes at the rice differentiation and heading stage were obtained. Three machine learning algorithms, Bayesian Ridge Regression (BRR), Support Vector Regression (SVR), and Partial Least Square Regression (PLSR), were used to establish yield estimate models based on the different data sources, including rice canopy spectrum data, spectrum data combined with

crop growth parameters, and spectral data, crop growth parameters, and crop nutrient intake data. The estimation accuracy of BRR, SVR and PLSR models, were evaluated and compared, then the best estimation model and optimal estimation growth period for rice yield were determined. The results showed that among the three methods, the BRR and SVR methods were more suitable for yield monitoring, with better performance in different periods and different parameter combinations ($R^2 > 0.82$, $\text{NRMSE} < 8.22\%$). Based on the 2019 and 2020 data, the best monitoring model for yield monitoring using full-band spectral information was the BRR model at the differentiation stage with R^2 of 0.90 and the SVR model at the tasseling stage with R^2 of 0.87. When full-band spectral information was used for yield monitoring, the best monitoring model for both periods was the BRR model with R^2 of 0.90 and 0.92, respectively, compared with the BRR and SVR models, the highest R^2 of the PLSR model was only 0.75 for different periods and different combinations of parameters; based on the 2020 data, when three different parameter combinations were used as inputs, the BRR model was the best in both periods, and the modeling accuracy was higher in the differentiation period than in the tasseling period (R^2 increased by at least 0.02 and NRMSE decreased by at least 0.61%), and when the input parameter combinations were full-band spectra with crop population growth parameters and crop nutrient uptake, the accuracy of the BRR model for yield estimation reached Through the experimental study, it was concluded that the optimal monitoring period for yield is the differentiation period. The optimal monitoring model is the BRR model. The study results can provide a reference for early remote sensing monitoring of rice yield.

Keywords Hyperspectral remote sensing; Rice yield estimation; Bayesian ridge regression; Support vector regression

(Received Sep. 30, 2021; accepted Feb. 25, 2022)

* Corresponding author