

高光谱早稻生理指标可跨期预测性的初步研究

祝海竣¹, 付虹雨^{1,2}, 王学华^{1*}, 崔国贤^{1,2*}, 石爱龙¹, 薛卫纯³

1. 湖南农业大学农学院, 湖南 长沙 410128
2. 湖南农业大学苎麻研究所, 湖南 长沙 410128
3. 益阳市赫山区农业农村局, 湖南 益阳 413000

摘要 叶绿素含量和叶片氮含量是作物生长状况的重要指标, 对其实时精准的监测有助于田间生产管理以及作物品质产量的提高。当前, 高光谱技术和经验回归方法被广泛应用于构建作物生化参数预测模型。但是, 有关同一生命活动周期内, 作物跨期预测叶片生化参数的研究还存在空白。以超级早稻为研究对象, 分别获取了蘖盛期、孕穗期、齐穗期、灌浆期和成熟期 5 个时期 120 组叶片高光谱数据、叶绿素以及叶片氮含量(LNC), 采用 python 3.6 编程, scikit-learn(0.22.1)用来构建模型和验证评估, 通过网格搜索(Grid-Search)和五折交叉验证(5-fold cross validation)在训练集中确定偏最小二乘回归(PLSR)、随机森林(RF)、支持向量回归(SVR)3 种不同算法的最佳模型参数, 结合数据的跨期处理, 分别建立针对不同生育期的水稻叶片氮含量跨期预测模型和叶绿素跨期预测模型。此外, 评估跨期预测模型与传统非跨期预测模型, 独立数据对模型进行检验。结果表明, (1) 基于高光谱的早稻叶绿素、氮素的非跨期预测模型中, PLSR 模型预测效果最佳, 分别为叶绿素($R^2=0.84$, $RMSE=1.85$)、氮素($R^2=0.85$, $RMSE=0.11$)。(2) 基于 SVR 的早稻叶绿素跨期预测模型预测效果最佳, 分别为跨分蘖期($R^2=0.54$)、跨孕穗期($R^2=0.36$)、跨齐穗期($R^2=0.30$)、跨灌浆期($R^2=0.55$)、跨成熟期($R^2=0.74$)。该结果为利用高光谱数据构建超级稻叶绿素含量预测模型提供新的理论参考, 为水稻叶绿素含量的动态监测提供了模型依据。(3) 早稻叶片氮含量跨期预测模型的拟合度很差, 叶片氮含量预测具有不可跨期性。与非跨期预测模型相比, 跨期预测模型虽然精度有所下降, 但能有效克服经验模型普适性差的缺陷, 有利于在同一生命活动周期内, 实现作物不同生育期的生理指标预测, 对实际生产管理具有重要意义。研究发现, 作物生理指标存在可跨期预测性, 这一概念为作物表型、作物内部品质以及产量的预测研究提供了新的思路。

关键词 高光谱; 早稻; 氮含量预测; 叶绿素; 可跨期性

中图分类号: S126 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)01-0170-06

引言

叶绿素和氮素是两个非常重要的表征作物营养和生长状况的指标^[1]。实时、精准监测作物叶绿素含量和氮素状况对作物科学栽培、田间肥料精准管理、提高作物产量和品质等均具有重要意义^[1-2]。

当前, 光谱反演是常用的估算作物叶片氮含量和叶绿素含量的方法^[2-3], Zhang 等^[4]结合高光谱技术和多元线性回归方法建立了实时监测叶片氮状况的定量模型, 结果表明实测值与估算值高度拟合。但传统线性回归方法在模型拟合度

方面存在缺陷, 模型精度往往不高。随着机器学习技术的发展, RF, PLSR 和 SVM 等算法被广泛用于作物生化参数反演研究, 成为传统逐步回归方法的很好替代。如 PLSR 融合了 MLR, PCR 和 CCA 3 种分析方法的优点, 具有很好的可解释性和预测功能; SVR 能够考虑高维数据和噪声模式, 进而提高预测精度; RF 通过对训练数据的自举采样以及随机选择变量使得决策更加快速、稳健^[5]。这些算法对于作物生化参数预测具有各自的优势, 然而, 在众多算法中, 最佳算法选择仍具有挑战。因此, 需要进一步工作来测试不同算法反演作物叶片氮含量和叶绿素含量的性能。

作物在不同生育期的不同光谱特性导致生化参数预测模

收稿日期: 2020-12-04, 修订日期: 2021-03-19

基金项目: 国家重点研发计划项目(2017YFD0301501, 2018YFD0301005, 2018YFD0201106), 国家自然科学基金项目(31471543)资助

作者简介: 祝海竣, 1997 年生, 湖南农业大学农学院博士研究生 e-mail: 921706423@qq.com

* 通讯作者 e-mail: 13873160151@163.com; 627274845@qq.com

型的拟合效果存在差异^[6]。如 Li 等^[7]认为不同生育期光谱数据对作物氮含量预测能力不同。杨福芹等^[8]研究表明不同生育期高光谱数据对冬小麦氮素营养状况的监测效果存在较大的差异。这些研究都表明作物叶片光谱反射率对不同生长时期的生化参数变化极其敏感, 作物生化参数反演模型的普适性问题存在争议。针对这一问题, Katherine 等^[9]根据两年光谱数据以及叶片氮含量(leaf N concentration, LNC)数据, 探讨了 PLSR 模型的跨年应用能力, 并认为, 经验模型受作物类型和地域环境的影响, 存在年际间普适性不高的缺陷, 因此往往需要进行调整甚至重新建模。已有研究探究了模型的年际间普适性^[10], 但针对模型不同生育期间普适性的研究还存在空白。作物在不同生育期具有的光谱特性是否存在相互联系, 如果有联系, 是否可以依靠同一生命活动周期的跨期数据实现作物生化参数的预测?

与以往研究单一地从不同生育期、不同算法等角度建立基于高光谱作物叶绿素和叶片氮含量预测模型不同, 本工作对水稻叶片光谱检测的研究更为系统, 将多种时期、多种建模方法和跨期预测方法融于一体, 分析水稻叶绿素和叶片氮含量与光谱反射率的定量关系, 包括: (1) 比较 PLSR, RF 和 SVR 3 种算法用于构建早稻叶片氮含量和叶绿素含量预测模型的预测性能。(2) 评估跨期预测模型预测水稻叶绿素含量与叶片氮含量的效果。

1 实验部分

1.1 试验材料与试验地概况

试验于 2020 年 3 月至 2020 年 7 月在益阳市赫山区笔架山乡中塘村(112°50'83"E, 28°49'61"N)进行。该地属于亚热带季风气候, 降水充沛, 光热条件良好, 适宜水稻生长。供试早稻品种为株两优 819 和株两优 39。采用随机区组实验, 每个品种设置 4 个施氮梯度, 分别为 D1(不施氮)、D2(90 kg·hm⁻²)、D3(150 kg·hm⁻²)、D4(210 kg·hm⁻²)。每个水平按小区设计 3 次重复, 共 24 个小区, 每小区面积 20 m²。试验材料采用大田育秧, 于 3 月 20 日播种, 4 月 20 号移栽。种植株行距为 16.5 cm×20 cm, 每穴 2 粒谷苗。

1.2 数据采集

1.2.1 叶片高光谱数据采集

于早稻分蘖盛期(5 月 22 日)、孕穗期(6 月 3 日)、齐穗期(6 月 17 日)、灌浆期(6 月 27 日)和成熟期(7 月 11 日)5 个时期, 使用美国 ASD 公司生产的 FieldSpec 3 地物光谱仪测量早稻叶片高光谱数据, 光谱范围为 350~2 500 nm。测量选择在当地时间 11:00—13:00, 天气晴朗、无风或者风速较小的情况下进行。每个试验小区选取 5 株具有代表性、均匀无病虫害危害的样本, 测量其最上部 2 片全展叶, 共获取 120 组光谱数据。测量时使用仪器内置光源, 并且每隔 10 min 进行一次参考白板校正。

1.2.2 叶片含氮量测定

于高光谱数据采集当天取样, 从每个试验小区采样点获

取早稻叶片, 分别装入自封袋中并标注小区名称和编号。去除叶片表面灰尘后, 将叶样放入 105 °C 烘箱中杀青 30 min, 并在 80 °C 下烘干至恒重。使用凯氏定氮法测定叶片氮含量。

1.2.3 叶绿素含量测定

采用日本 Minolta Camera 公司生产的 SPAD-502 便携式叶绿素仪测量叶片叶绿素含量, 样本株与光谱测定的一致, 每小区叶绿素测量 5 株, 每株测超级早稻的 2 片最上部的全展叶, 每片叶的基部、中部和尖部各测一点并求其平均值即为该叶的 SPAD 值。

1.3 模型构建

1.3.1 跨期预测模型

将 120 组光谱数据按分蘖期、孕穗期、齐穗期、灌浆期和成熟期分为 5 组, 采用 PLSR, RF 和 SVR 3 种方法构建早稻叶片氮含量和叶绿素跨期预测模型。跨期预测模型是指在构建某一特定时期(X 时期)预测模型的过程中, 跨过 X 时期数据, 仅以其他时期数据为模型输入。以分蘖期叶片氮含量跨期预测模型为例, 以孕穗期、齐穗期、灌浆期和成熟期光谱数据为输入, 实测早稻叶片氮含量为输出, 建立叶片氮含量跨分蘖期预测模型, 然后以分蘖期数据作为验证集。

1.3.2 非跨期预测模型

试验共采集全生育期光谱数据 120 组, 将数据混合并按照 4:1 的比例随机划分为训练集和测试集。以光谱数据为输入, 实测早稻叶片氮含量或叶绿素含量为输出构建早稻叶片含氮量和叶绿素的非跨期预测模型。

1.3.3 建模环境及模型评价指标

采用 python 3.6 进行编程, scikit-learn(0.22.1)构建模型和验证评估。通过网格搜索和五折交叉验证在训练集中确定各算法的最佳模型参数, 然后分别以最佳参数通过训练集进行训练, 将训练好的各算法在测试集中评估模型表现。选用 R² 和 RMSE 作为评估指标, R² 越大, RMSE 越小, 预测效果越好。RMSE 计算公式如式(1)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (1)$$

式(1)中, y_i 为实测值, \hat{y}_i 为预测值, n 为样本数。

2 结果与讨论

2.1 早稻叶绿素含量跨期预测模型分析

各时期的早稻叶绿素含量跨期预测模型结果见图 1。5 个跨期预测模型效果存在差异, 其中跨成熟期预测模型效果最佳, R² 达到 0.74, RMSE 为 1.77, 具体表现为跨成熟期 > 跨灌浆期 > 跨分蘖期 > 跨齐穗期 > 跨孕穗期。可见, 在同一生命活动周期内, 依靠水稻生长早期光谱以及叶绿素数据, 能够较准确预测成熟期数据。此外, 不同算法对跨期预测模型精度也有影响。PLSR 算法在跨齐穗期预测模型中效果最好, 而在另外 4 个跨期预测模型中, SVR 算法更加具有优势。

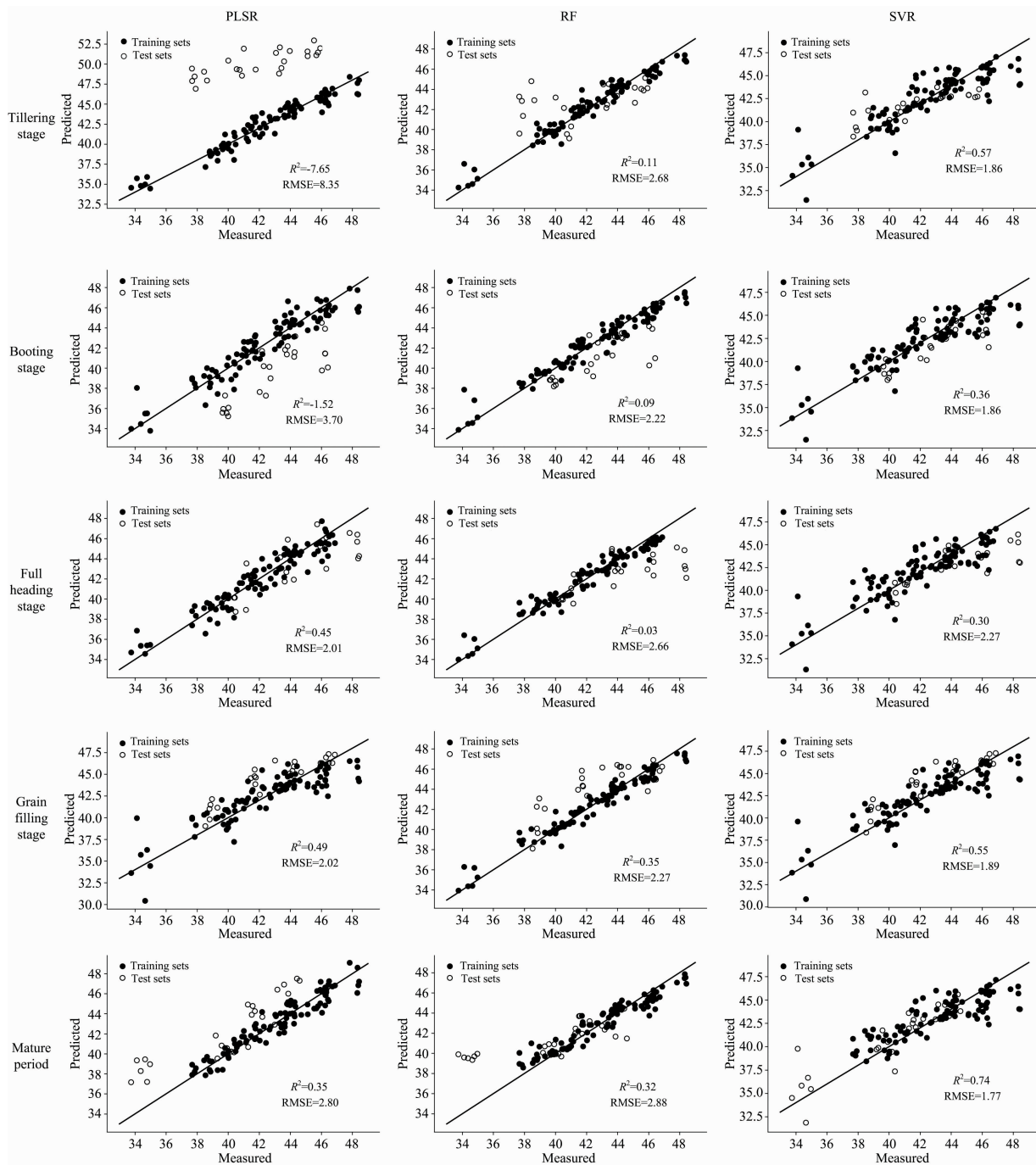


图 1 早稻叶绿素含量跨期预测模型分析

Fig. 1 Analysis of intertemporal prediction model of chlorophyll content in early rice

2.2 早稻叶片氮含量跨期预测模型分析

图 2 为早稻叶片氮含量跨期预测模型的精度分析。结果显示针对各时期的早稻叶片氮含量跨期预测模型的拟合效果都很差。除基于 PLSR 算法建立的齐穗期叶片氮含量预测模型 R^2 为 0.09、基于 PLSR 算法建立的跨灌浆期叶片氮含量预测模型 R^2 为 0.41 以及基于 SVR 算法建立的跨灌浆期叶片氮含量预测模型 R^2 为 0.23 外，其他模型的 R^2 都为负值。这表明早稻叶片氮含量具有不可跨期预测性，各生育期早稻

叶片光谱特性仅能反映同时期叶片含氮量，依据统计关系建立的早稻叶片氮含量预测模型普适性不高。

2.3 水稻叶片氮含量和叶绿素含量非跨期预测模型

基于以上结果，我们发现水稻叶片氮含量跨期预测模型不具有预测能力。为进一步探究其原因，构建了早稻叶片含氮量和叶绿素的非跨期预测模型。如图 3 所示，水稻叶绿素非跨期预测模型的精度高，3 种不同算法的性能表现为：PLSR>SVR>RF。其中 PLSR 算法最具有优势， R^2 为

0.84, RMSE 为 1.25。水稻叶片含氮量非跨期预测模型也具有较高的预测能力, 其中 PLSR 模型预测效果最好 ($R^2 = 0.85$, $RMSE=0.11$), 其次为 SVR 模型和 RF 模型。SVR 的

R^2 为 0.81, $MRSE$ 为 0.12; DF 的 R^2 为 0.56, $RMSE$ 为 0.19。因此, 跨期预测模型与非跨期预测模型之间的模型差异由跨期因素导致。

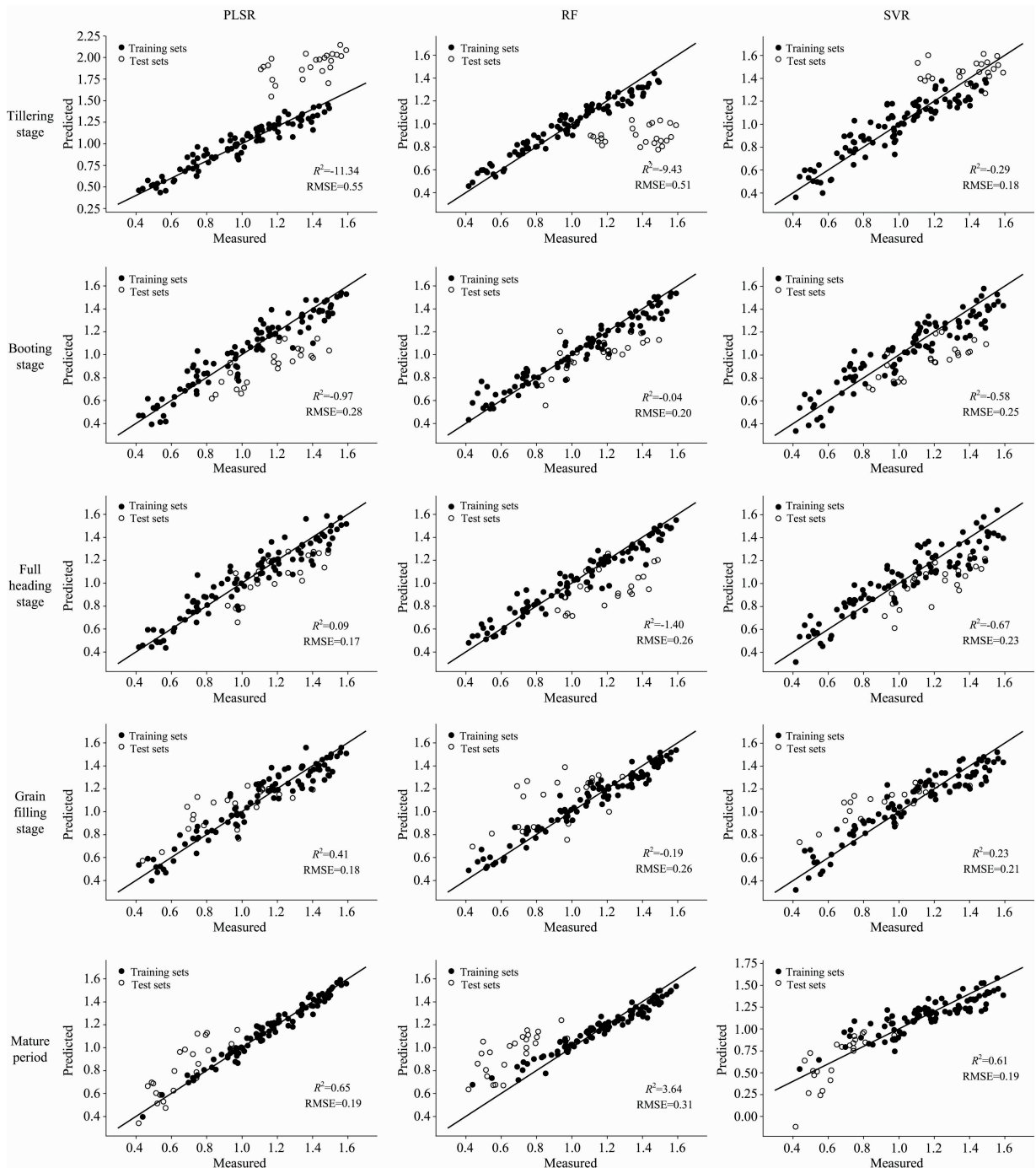


图 2 水稻叶片氮含量跨期预测模型分析

Fig. 2 Analysis of inter-temporal prediction model of rice leaf nitrogen content

2.4 不同模型构建算法的比较分析

随着机器学习在各类算法上的突破, 大量线性和非线性计算方法被用于基于复杂光谱的作物生化参数反演建模。比较了 PLSR 线性回归和 RF 和 SVR 两种非线性回归方法的预测性能。PLSR 线性回归在非跨期预测模型中比 RF 和

SVR 两种非线性回归的预测性能更好, 这表明早稻叶片光谱反射特征主要呈线性相关, PLSR 算法在包含线性和简单数据的数据集中是可行的, 这与 Ghasemi 等^[11]的研究结果一致。在跨期预测模型中, SVR 算法明显具有优势。在 Shao 等^[12]、Yang 等^[13]的研究中, SVR 也表现出比线性回归更

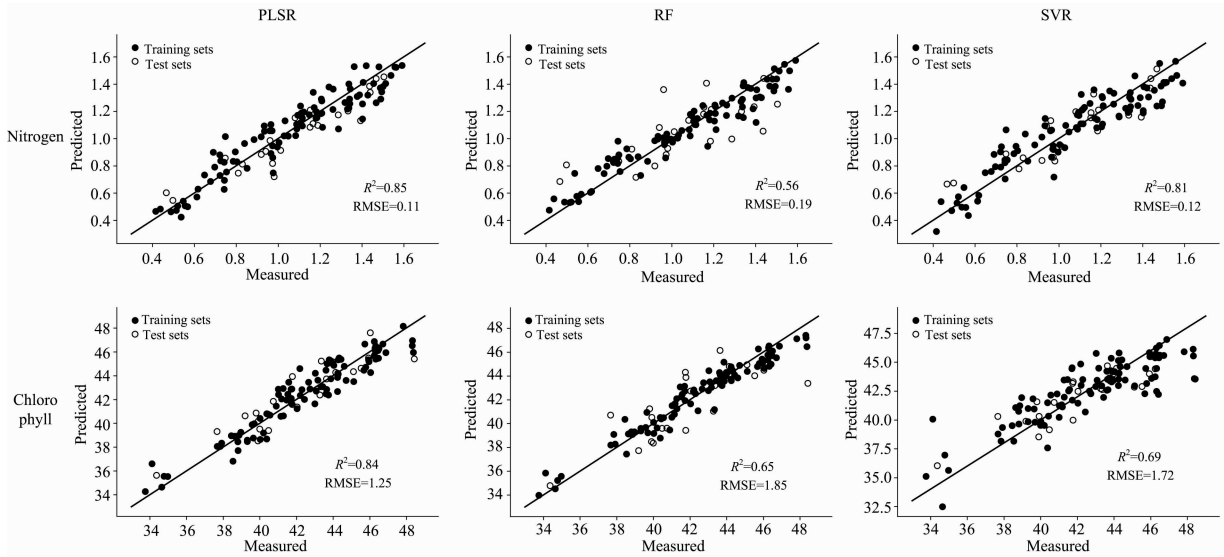


图 3 水稻叶片氮含量和叶绿素含量非跨期预测模型

Fig. 3 Non-intertemporal prediction model of rice leaf nitrogen content and chlorophyll content

好的预测性能。相比于 PLSR 线性回归模型,非线性 SVR 模型能更深度地挖掘光谱参数与水稻生化成分参数之间的统计关系。

2.5 跨期预测模型与非跨期预测模型比较分析

我们实现了水稻叶片氮含量和叶绿素的非跨期预测,表明基于光谱的作物叶片生化成分定量分析是可行的。

一些研究讨论了经验模型的普适性,如 Li 等^[7]认为用经验模型普适性差且反演作物生化成分的内在机理存在争议。对于这一问题的解释是,受不同环境影响,经验模型在不同年际的反演效果上必然存在差异,因此必须谨慎使用经验模型来推断不同年际间作物各项生化成分状况。为探讨模型在不同生育期的普适性,利用跨期数据进行了同一生命活动周期的水稻叶片氮含量和叶绿素进行预测研究。结果显示,对比非跨期预测模型,模型构建未包括同期数据时,模型精度降低或失去预测能力。其中,针对不同生育期的早稻叶绿素含量跨期预测模型精度明显下降;而针对不同生育期的水稻叶片氮含量跨期模型的 R^2 大多为负值,根据生育期前期数据模型无法预测后期氮素状况。对模型效果差异的解释是由跨期因素导致。同期数据受生长环境等因素影响具有一致性,而在跨期预测模型中,完全避免了同期数据的干扰,从而导致跨期预测模型精度下降以及无法预测。

2.6 作物生理生化指标预测的可跨期性

对早稻叶片氮含量和叶绿素进行跨期预测,结果表明,叶绿素含量表现出可跨期预测潜力,而叶片氮含量并不可跨期预测。这可能是由于叶绿素在早稻全生育期中仅与光合作用相关,影响因素较少,便于全生育期预测。而氮素参与了氨基酸合成、激素调控等多种代谢途径,这些途径在作物不

同生育期差异较大而在相同生育期差异较小,影响因素多,导致其跨期预测准确性较差,这也解释了氮素可以进行非跨期预测原因。其中的内在机理还需要进一步研究。本研究提出了作物生理指标的“可跨期预测性”,这一特性反映了作物在不同生育期下某一生理指标之间存在的内在联系,有助于实际生产过程中的早期预测,也为作物表型性状以急内部品质的预测研究提供了新思路。目前,对作物生理指标预测的可跨期性研究还并不深入,本研究仅仅对水稻的叶片氮含量以及叶绿素含量进行预测,并且试验年限较短,下一步研究应进一步探讨不同作物、不同生育期、不同生理指标的可跨期预测性。

3 结论

分析了不同施氮梯度下早稻叶片含氮量、叶绿素含量的季节变化以及采用叶片光谱反射率反演生化参数的潜力。得到以下结论:

(1)光谱可以用于早稻叶绿素含量以及叶片氮素的估算。在非跨期预测研究中,PLSR 模型对早稻叶绿素含量和叶片氮含量的预测效果比 RF 模型和 SVR 模型更好。

(2)相对非跨期预测模型,跨期预测模型精度较差,早稻叶绿素含量跨期预测模型精度明显下降。而在叶片氮含量预测过程中,根据生育期前期数据模型无法预测后期氮素状况。

(3)作物生理生化指标存在“可跨期预测性”,并且跨期预测的准确度受生育期选择影响较大,有必要进一步探讨不同作物、不同生理生化指标的“可跨期预测性”。

References

- [1] Yao X, Huang Y, Shang G, et al. *Remote Sensing*, 2015, 7: 14939.
- [2] LIN Yi, LIU Si-yuan, YAN Lei, et al(林沂, 刘思远, 晏磊, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(6): 1956.
- [3] Hu H, Bai Y L, Yang L P, et al. *Plant Nutrition and Fertilizer Science*, 2009, 15(6): 1317.
- [4] Zhang Manyin, Li Mengjie, Liu Weiwei, et al. *Spectroscopy Letters*, 2019, 52(9): 483.
- [5] Qiu Shanshan, Wang Jun, Gao Liping. *Journal of Agricultural and Food Chemistry*, 2014, 62(27): 6426.
- [6] ZHANG Yin-jie, WANG Lei, BAI You-lu, et al(张银杰, 王磊, 白由路, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2019, 39(9): 2829.
- [7] Li Zhenhai, Jin Xiuliang, Yang Guijun, et al. *Remote Sensing*, 2018, 10(9): 1.
- [8] YANG Fu-qin, FENG Hai-kuan, LI Zhen-hai, et al(杨福芹, 冯海宽, 李振海, 等). *Journal of Shanxi Agricultural University • Nature Science Edition(山西农业大学学报 • 自然科学版)*, 2020, 40(5): 99.
- [9] Katherine Meacham-Hensold, Christopher M Montes, Jin Wu, et al. *Remote Sensing of Environment*, 2019, 231: 111176.
- [10] LIANG Xue-ying, FAN Feng-lei(梁雪映, 樊风雷). *Journal of Agricultural Science and Technology(中国农业科技导报)*, 2019, 21(1): 51.
- [11] Ghasemi J B, Tavakoli H. *Analytical Methods*, 2013, 5: 1863.
- [12] Shao Y, Zhao C, Bao Y, et al. *Food & Bioprocess Technology*, 2012, 5(1): 100.
- [13] Yang J, Du L, Gong W, et al. *PLOS ONE*, 2018, 13(1), 1.

Preliminary Study on the Intertemporal Predictability of the Physiological Index of Early Rice Based on Hyperspectral

ZHU Hai-jun¹, FU Hong-yu^{1,2}, WANG Xue-hua^{1*}, CUI Guo-xian^{1,2*}, SHI Ai-long¹, XUE Wei-chun³

1. Agronomy College, Hunan Agricultural University, Changsha 410128, China

2. Ramie Research Institute, Hunan Agricultural University, Changsha 410128, China

3. Agriculture and Rural Bureau of Heshan District, Yiyang 413000, China

Abstract Chlorophyll and leaf nitrogen content (LNC) are important indicators of crop growth. Real-time and accurate monitoring of Chlorophyll and LNC is helpful for field management and improvement of crop quality and yield. Currently, hyperspectral technology and empirical regression methods are widely used to construct crop biochemical parameter prediction models. However, there are still gaps in the research on predicting leaf biochemical parameters across periods. In this study, 120 leaf hyperspectral data, chlorophyll and LNC were obtained at five stages of early rice. Three different algorithms, namely, Partial Least Squares Regression (PLSR), Random Forest (RF), and Support Vector Regression (SVR) were used to establish inter-temporal prediction models for Chlorophyll and LNC in different growth stages. In addition, the intertemporal prediction model and the traditional non-intertemporal prediction model were evaluated. The results showed that the early rice chlorophyll inter-temporal prediction SVR model had the best prediction effect, with the inter-tillering prediction model having R^2 of 0.54, the inter-booting prediction model has R^2 of 0.36, the inter-heading prediction model has R^2 of 0.30, inter-grouting prediction model has R^2 of 0.55, and inter-mature prediction model has R^2 of 0.74. The intertemporal prediction model of early rice leaf LNC was poor, and the LNC cannot be intertemporally predicted. Compared with the non-intertemporal prediction model, although the accuracy of the intertemporal prediction model has decreased, it can effectively overcome the shortcomings of poor universality of the empirical model, which is beneficial to realize the physiological characteristics of different growth stages of crops in the same life cycle. At the same time, we found that crop physiological indicators have inter-temporal predictability. This concept provides new ideas for the prediction of crop phenotype, internal crop quality and yield.

Keywords Hyperspectral; Early rice; Leaf nitrogen content; Chlorophyll; Intertemporal prediction

* Corresponding authors

(Received Dec. 4, 2020; accepted Mar. 19, 2021)