

考虑水分光谱吸收特征的水稻叶片 SPAD 预测模型

于滋洋¹, 王翔¹, 孟祥添¹, 张新乐^{1*}, 武丹茜¹, 刘焕军^{1,2}, 张忠臣^{3*}

1. 东北农业大学资源与环境学院, 黑龙江 哈尔滨 150030

2. 中国科学院东北地理与农业生态研究所, 吉林 长春 130012

3. 东北农业大学农学院, 黑龙江 哈尔滨 150030

摘要 叶绿素是植被光合作用的重要色素, 传统实验室方法测定叶绿素含量需破坏性取样且操作复杂。通过构建高精度 SPAD 光谱估算模型, 可以实现对水稻叶片叶绿素含量的实时无损监测。以黑龙江省不同施氮水平下水稻为研究对象, 采用 SVC HR768i 型光谱辐射仪共获取移栽后、分蘖期、拔节期、孕穗期、抽穗期共五个关键时期水稻叶片反射光谱数据。光谱探测范围 350~2 500 nm。利用自带光源型手持叶片光谱探测器直接测定叶片光谱, 光源为内置卤素灯。采用 SPAD-502 型手持式叶绿素仪同步测定水稻叶片的 SPAD 值。叶片水分是植物光合作用的基本原料, 也间接影响着叶绿素含量。叶片含水量降低则会影响植物的光合作用, 导致其叶绿素含量随之降低。因此将叶绿素敏感波段与水分吸收范围结合作为 SPAD 估算的输入量。随机森林模型是一个基于多个分类树的算法。算法在采样的过程中包括两个完全随机的过程, 一是有放回抽样, 可能会得到重复的样本, 二是选取自变量是随机的。因此本文对叶片光谱反射率进行去包络线(CR)处理, 综合考虑可见光近红外波段提取水稻叶片反射光谱特征参数和植被指数, 综合分析光谱指标与 SPAD 相关关系, 采用随机森林算法构建不同输入量的 SPAD 高光谱估算模型。结果表明: (1) 水稻叶片 SPAD 与光谱反射率的相关系数在叶绿素敏感波段红波段范围(600~690 nm)、红边范围(720~760 nm)、水分吸收波段范围(1 400~1 490 和 1 900~1 980 nm)均为 0.75 以上; (2) 在光谱参数与 SPAD 的相关分析中, NDVI, DP2 与水稻叶片 SPAD 值相关性最好, 相关系数为 0.811 和 0.808; (3) 以结合水分光谱信息后的 $CR_{(v_1, v_2, v_3, v_4)}$ 为自变量所建立的随机森林模型精度最高, R^2 为 0.715, RMSE 为 2.646, 可作为水稻叶片叶绿素预测模型。研究结果揭示了不同品种水稻的光谱响应机制, 提供了水稻叶片 SPAD 值高精度反演的技术方法, 为监测与调控东北地区水稻正常生育进程提供技术支持。

关键词 高光谱; 随机森林; 叶绿素相对含量; 去包络线; 水稻

中图分类号: S511 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2528-05

引言

叶绿素含量是作物生长的重要指标, 是评价作物光合作用能力的重要依据, 因此可以通过监测作物叶绿素含量间接对作物生长条件进行调控。分光光度计法, 荧光分析法, 活体叶绿素仪法等被用于监测叶绿素含量^[1], 但这些方法操作复杂繁琐, 且具有一定的破坏性。

遥感可以为作物生理指标的定量化监测提供简单快捷、非破坏性的方法。国内外许多学者在在监测植被生理指标方

面做了大量研究。针对叶绿素反演的研究也有很多, 有学者提出一种估算作物叶片叶绿素含量和冠层叶绿素含量反演方法并建立混合反演模型^[2]; 以红边参数为自变量同样可以对 SPAD 进行良好的估测, 并且相对于红边位置以红边峰度为自变量可以提高估算精度^[3]。Gitelson 等提出 CI_{RE} 和 $MTCI$ 指数在叶片尺度的叶绿素监测精度较高^[4]。以上研究均局限在可见光波段范围结合红边参数计算植被指数或者光谱参数作为自变量建立估算模型。植被指数是通过光谱反射率间接计算得到, 包络线去除处理可以增强吸收特征, 直接表征植被的光谱反射率信息^[5]。水分是植物光合作用的基本原料,

收稿日期: 2019-01-08, 修订日期: 2019-05-11

基金项目: 国家重点研发计划(2016YFD0300604-4), 国家自然科学基金项目(41671438), 中国科学院东北地理与农业生态研究所“引进优秀人才”项目资助

作者简介: 于滋洋, 1995 年生, 东北农业大学资源与环境学院硕士研究生 e-mail: ziyangyu@yeah.net

* 通讯联系人 e-mail: xinlezhang@yeah.net; zzcneau@neau.edu.cn

叶片含水量降低则会影响植物正常的光合作用，导致其叶绿素含量随之降低，水分含量也间接影响着叶绿素含量^[6-7]。因此，尝试将叶绿素敏感波段与水分吸收波段结合估算 SPAD 含量。

随着人工智能时代的发展，一些机器学习算法逐渐在各个研究领域广泛应用。随机森林模型是一个基于多个分类树的算法。通过对大量分类树的汇总提高了模型的预测精度，并且便于计算变量的非线性作用^[8]，拟合能力强且不会出现过度拟合问题。有学者选择最优光谱指数和确定最佳自变量个数的基础上结合随机森林算法可以构建高精度 LAI 反演模型^[9]。但利用随机森林回归算法结合去包络线构建叶片 SPAD 估算模型的研究尚不多见。

以水稻为研究对象，综合分析水稻不同生育期光谱参数与 SPAD 相关关系，利用可见光近红外叶绿素敏感波段与短波红外水分吸收波段光谱指数共同作为输入量，采用随机森林算法构建水稻叶片 SPAD 的遥感估算模型。旨在为提升水稻长势监测精度提供理论依据与技术支持。

1 实验部分

1.1 样品采集

2017 年 6 月—9 月在黑龙江省哈尔滨市呼兰区试验田 (N45.9°, E126.8°)，选取晚熟品种 188 和早熟品种龙稻 11 号两种，共获取移栽后、分蘖期、拔节期、孕穗期、抽穗期共五个关键时期数据。施肥处理采用移栽施一次基肥，分蘖期和抽穗期之前进行两次追肥的施肥方式。基肥，分蘖肥和穗肥施肥比例为 5:2:3，氮磷钾比例为 2:1:2。总共设置 6 个小区，氮肥处理设置 0, 90, 112.5, 135, 157.5 和 180 kg·hm⁻² 等 6 个肥力水平，分别用 N1, N2, N3, N4, N5 和 N6 表示。施肥品种分别为二铵、尿素和硫酸钾，二铵的含磷量为 46%，含氮量为 17%，尿素的含氮量为 46%，硫酸钾的含钾量为 50%。每个施肥水平设置 4 次重复，随机区组排列，小区面积为 72 m²。5 月 24 日播种，田间管理按照大田种植方式进行。

1.2 数据采集与处理

水稻叶片反射光谱数据获取采用美国 SVC HR768i 型光谱辐射仪，光谱探测范围 350~2 500 nm，光谱分辨率为 3.5, 8.5 和 6.5 nm。利用自带光源型手持叶片光谱探测器直接测定叶片光谱，光源为内置卤素灯。采用 SPAD-502 型手持式叶绿素仪同步测定水稻叶片的 SPAD 值，具体步骤见文献^[10]。

1.3 特征参数选择

表 1 为不同特征参数的计算方法。提取的参数共分为两类，分别为基于原始反射率计算的植被指数和基于包络线去除提取的去包络线参数。

图 1 为去包络线后的反射光谱曲线，可提取的吸收特征主要包括：4 个吸收谷 (V1, V2, V3, V4) 的吸收深度 (DP) 和吸收谷面积 (A)。依据上述吸收特征参数提取指标包括 V1 (440~550 nm)，V2 (550~760 nm)，V3 (1 300~1 700 nm)，V4 (1 850~2 250 nm)，这 4 个吸收谷的深度 (DP₁~DP₄) 和

吸收谷面积 (A₁~A₄)。

表 1 光谱指数计算公式

Table 1 Calculation formula of spectral index

类型	光谱指数	定义及计算方法	文献
植被指数	NDVI	$(R_{800} - R_{670}) / (R_{800} + R_{670})$	[11]
	MTCI	$(R_{NIR} - R_{705}) / (R_{705} - R_{660})$	[4]
	CI _{RE}	$(R_{NIR} / R_{RE}) - 1$	[4]
	MCARI	$(R_{700} - R_{670}) - 0.2(R_{700} - R_{550}) \times (R_{700} / R_{670})$	[12]
去包络线参数	DP	$1 - CR_n$	[13]
	A	$\sum_n (1 - CR_n)$	[13]

注：N 为吸收谷中间的波长个数，CR_n 为第 n 个波长对应的去包络线值

Note: N is the number of wavelengths in the middle of the absorption valley, CR_n is the value of the continuum removal corresponding to the n-th wavelength

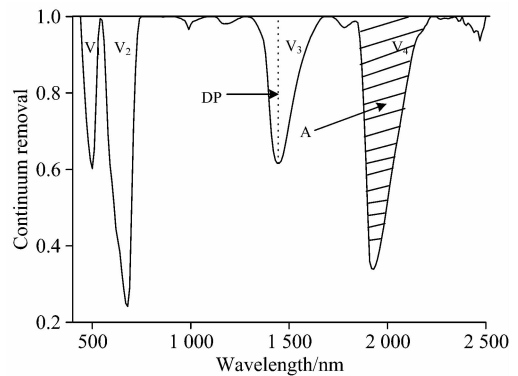


图 1 水稻叶片去包络线反射光谱

Fig. 1 Rice leaf reflectance continuum-removed spectra

1.4 模型构建与精度验证

随机森林是由多个决策树模型组成的预测模型。在采样的过程中包括两个完全随机的过程：(1)有放回抽样，可能会得到重复的样本；(2)选取自变量是随机的。然后对采样之后的数据使用完全分裂的方式建立决策树，再根据投票的形式，选取最优的模型进行预测。随机森林模型在 R 软件中实现。

将不同方法计算的参数作为自变量建立随机森林回归模型。将参数提取范围分为两类，分别是 440~1 000 和 440~2 500 nm。以 440~1 000 nm 反射率、去包络线、去包络线前两个吸收谷参数作为自变量记为 $R_{(440\sim 1\ 000\ nm)}$ ， $CR_{(440\sim 1\ 000\ nm)}$ ， $CR_{(V1, V2)}$ 。以 440~2 500 nm 反射率、去包络线、去包络线四个吸收谷参数作为自变量记为 $R_{(440\sim 2\ 500\ nm)}$ ， $CR_{(440\sim 2\ 500\ nm)}$ ， $CR_{(V1, V2, V3, V4)}$ 。将 120 个样本中的 72 个作为训练样本，48 个作为验证样本，以决定系数 (R^2)、均方根误差 (RMSE) 作为指标评价指标。 R^2 越大，RMSE 越小，模型越稳定。

2 结果与讨论

2.1 不同生长期水稻叶片光谱特征分析

图 2 为品种 188 水稻 N3 施肥处理的不同生育期叶片反射率曲线与去包络线。叶片反射率主要受叶片内色素、细胞结构及叶片含水量三方面影响。在可见光波段(350~720 nm)叶片反射率受叶绿素影响最大。随水稻植株生长,叶片对红光、蓝光吸收增强,使绿波段反射率逐渐突出。在近红外波段(720~1 300 nm),由于水稻叶片的细胞壁和细胞空隙间折射率影响,导致反射率较高。在短波红外内(1 400~2 500 nm),水稻反射率主要受叶片含水量影响。经过去包络线处理后,光谱吸收特征更加明显。不同品种水稻,在不同生育时期,去包络线曲线特征大致相似。在 440~2 500 nm 范围内都有 4 个吸收谷,随着生育进程的发展吸收谷的深度和面积不断变大,基于以上特征提取吸收特征参数。

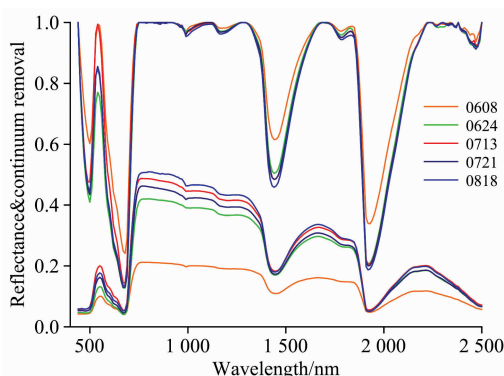


图 2 水稻不同生育期叶片光谱反射率曲线与去包络线
Fig. 2 Spectral reflectance curve and continuum removal line of rice leaves at different growth stages

2.2 相关性分析

2.2.1 SPAD 与不同数学变换光谱数据的相关分析

光谱反射率与水稻叶片 SPAD 值在整个波长范围内均存在正相关关系,在 440~680 nm 范围内相关系数逐渐减小,至红边处(690~760 nm)逐渐变大,在近红外(760~1 800 nm)内相关系数稳定在 0.65 以上。1 900~1 930 nm 范围的水分吸收带处相关系数最小。

去包络线与 SPAD 值相关关系在整个波长范围内基本呈负相关关系。在红光范围(620~680 nm),水分吸收带(1 400~1 450, 1 900~1 930 nm)等强吸收范围内的相关系数在 0.75 以上,在波长 660 nm 处相关系数最大为 0.81。

2.2.2 SPAD 与不同光谱指数相关分析

将不同时期测量的水稻叶片 SPAD 值与不同方法计算的光谱指数分别进行相关分析,结果如表 2 所示。选取与 SPAD 极显著相关的 4 种植被指数。NDVI、RVI 与 SPAD 相关性最好,相关系数在 0.78 以上,而 CI_{RE} 、MTCI 与 SPAD 相关系数在 0.6 以上,且相关性依次递减。

在去包络线参数与 SPAD 相关分析中发现,可见光处吸收谷深度与吸收谷面积与 SPAD 均呈极显著正相关且相关

系数较大, DP_2 与 SPAD 相关系数最大为 0.808。水分吸收处深度与吸收谷面积与 SPAD 同样呈极显著相关,相关系数小于可见光波段光谱指数。综合来看,吸收谷深度与 SPAD 相关性高于对应的吸收谷面积。对不同方法计算的参数与 SPAD 相关性进行比较可以发现,所有去包络线参数相关性接近 NDVI 且好于其余植被指数。

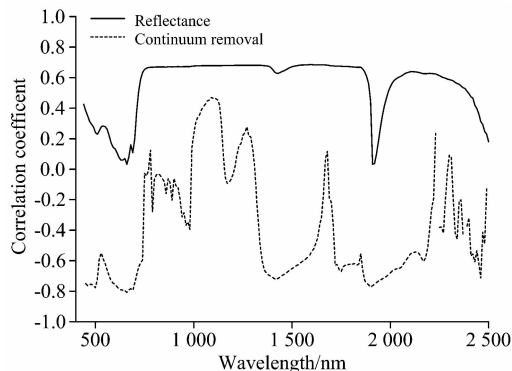


图 3 光谱反射率、去包络线与 SPAD 相关系数
Fig. 3 Spectral reflectivity, continuum removal and SPAD correlation coefficient

表 2 光谱指数与水稻叶片 SPAD 相关系数(样本数 $n=120$)

Table 2 Correlation coefficient between spectral index and SPAD in rice leaves				
类型	光谱指数	相关系数	光谱指数	相关系数
植被指数	NDVI	0.811**	CI_{RE}	0.663**
	MTCI	0.629**	MCARI	0.278**
去包络线参数	DP_1	0.778**	A_1	0.754**
	DP_2	0.808**	A_2	0.758**
	DP_3	0.705**	A_3	0.663**
	DP_4	0.750**	A_4	0.697**

注: *, ** 分别表示 0.05 和 0.01 水平上显著相关

Note: *, ** indicate that correlations are significant at confidence levels of 0.05 and 0.01, respectively

2.3 基于随机森林回归模型的 SPAD 估算

由表 3 可知,在 440~1 000 nm 范惜内,以 $CR_{(440\sim 1000\text{ nm})}$ 作为自变量建立模型估算精度最高,精度明显优于光谱反射率。当参数波段范围为 440~2 500 nm 内,模型精度显著提高。与 $CR_{(440\sim 1000\text{ nm})}$, $CR_{(V_1, V_2)}$ 相比,以结合水分吸收特征后的 $CR_{(440\sim 2500\text{ nm})}$, $CR_{(V_1, V_2, V_3, V_4)}$ 作为自变量建立的模型, R^2 分别提高 6% 和 39%。经对比分析可以得出, $CR_{(V_1, V_2, V_3, V_4)}$ 提取特征参数简便,估算精度高,可作为估算 SPAD 值得最优输入量。

对原始反射率进行包络线去除处理,可以增强水稻叶片吸收特征,使得不同叶绿素含量的光谱曲线差异显著。在植被生长过程中叶绿素在不同阶段变化速率不同,导致在生长末期叶绿素含量本身存在饱和现象。研究发现以去包络线参数为自变量时估算精度最高。说明去包络线参数与随机森林回归结合能较好地解决饱和现象及噪声问题,并且由于其自身的效率高及不易过度拟合等优点适合解决水稻叶片叶绿素

估算问题。

表 3 水稻叶片 SPAD 估算验证样本精度

Table 3 Accuracy of SPAD estimation for rice leaves

参数波段范围/nm	参数	R^2	RMSE
440~1 000	$R_{(440\sim 1\ 000\ \text{nm})}$	0.548	3.997
	$CR_{(440\sim 1\ 000\ \text{nm})}$	0.669	2.755
	$CR_{(V_1, V_2)}$	0.510	3.634
	植被指数	0.616	3.253
440~2 500	$R_{(440\sim 2\ 500\ \text{nm})}$	0.552	3.670
	$CR_{(440\sim 2\ 500\ \text{nm})}$	0.710	2.591
	$CR_{(V_1, V_2, V_3, V_4)}$	0.715	2.646

相关性分析发现, 经过去包络线处理后的光谱, 在整个波长范围内(350~2 500 nm)与 SPAD 相关性较高的波段范围在红光范围(620~680 nm)、红边(720~760 nm)、水分吸收范围(1 400~1 490, 1 900~1 980 nm)。水分吸收范围与 SPAD 相关性高是因为植物叶片含水量与叶绿素含量本身显著相关, 二者存在相互依存的关系, 二者缺一导致植物不能健康生长, 光合作用受到影响使叶片枯黄, 另一方含量随即减少。之前的研究大多集中在可见光范围提取参数对叶绿素含量进行估算。本工作将蓝光、红光与水分吸收范围结合作为 SPAD 估算的输入量建立随机森林模型。结果表明, 考虑水分吸收特征后的模型精度显著提高。

References

- [1] HUANG Chi-du, HU Xiao-song, LIAO Xiao-jun, et al(黄持都, 胡小松, 廖小军, 等). China Food Additives(中国食品添加剂), 2007, (3): 114.
- [2] Liang L, Qin Z, Zhao S, et al. International Journal of Remote Sensing, 2016, 37(13): 2923.
- [3] YAO Fu-qi, CAI Huan-jie, LI Ya-long, et al(姚付启, 蔡焕杰, 李亚龙, 等). China Rural Water and Hydropower(中国农村水利水电), 2015, (3): 84.
- [4] Gitelson A, Solovchenko A. Geophysical Research Letters, 2017, 44(18): 9293.
- [5] GUO Chao-fan, GUO Xiao-yu(郭超凡, 郭道宇). Acta Ecologica Sinica(生态学报), 2016, 36(20): 6538.
- [6] HU Song-ping, MEI Han-wei, ZOU Gui-hua, et al(胡颂平, 梅捍卫, 邹桂花, 等). Chinese Journal of Plant Ecology(植物生态学报), 2006, 30(3): 120.
- [7] Sanchez R A, Hall A J, Trapani N, et al. Photosynthesis Research, 1983, 4(1): 35.
- [8] LI Xin-hai(李欣海). Chinese Journal of Applied Entomology(应用昆虫学报), 2013, 50(4): 1190.
- [9] Liang L, Di L, Zhang L, et al. Remote Sensing of Environment, 2015, 165: 123.
- [10] LI Yuan-yuan, CHANG Qing-rui, LIU Xiu-ying, et al(李媛媛, 常庆瑞, 刘秀英, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(16): 135.
- [11] Sims D A, Gamon J A. Remote Sensing of Environment, 2002, 81(2-3): 337.
- [12] Daughtry C S T, Walthall C L, Kim M S, et al. Remote Sensing of Environment, 2000, 74(2): 229.
- [13] LIU Huan-jun, ZHANG Xiao-kang, ZHANG Xin-le, et al(刘焕军, 张小康, 张新乐, 等). Journal of Remote Sensing(遥感学报), 2017, 21(1): 105.

本研究区位于东北地区缺乏不同地区模型适宜性的比较并且在模型构建的方法是否可适用不同作物不同参数有待于进一步探究与验证, 以提升随机森林算法在农业遥感中的适用性。

3 结 论

以东北地区不同品种五个时期水稻叶片为研究对象, 构建了去包络线-光谱参数-随机森林 SPAD 值预测模型, 得出以下结论:

(1)水稻叶片光谱反射率与 SPAD 值在 760~1 800 nm 范围内呈极显著相关, 相关系数稳定在 0.65 以上; 去包络线参数与 SPAD 值在 620~680, 1 400~1 450 和 1 900~1 930 nm 范围内呈极显著负相关, 在 660 nm 处具有最大相关系数 0.81。

(2)在光谱参数中, NDVI, DP2 与水稻叶片 SPAD 值相关性最好, 相关系数为 0.811 和 0.808。去包络线参数与 SPAD 值相关性好于除 NDVI 外其余植被指数。

(3)以引入水分敏感波段后的去包络线光谱指数为自变量, 建立的 $CR_{(V_1, V_2, V_3, V_4)}$ 随机森林模型精度最高, R^2 为 0.715, RMSE 为 2.646。研究结果表明考虑水分信息能提高 SPAD 估算精度, 可作为水稻叶片叶绿素监测的重要方法。

SPAD Prediction Model of Rice Leaves Considering the Characteristics of Water Spectral Absorption

YU Zi-yang¹, WANG Xiang¹, MENG Xiang-tian¹, ZHANG Xin-le^{1*}, WU Dan-qian¹, LIU Huan-jun^{1,2}, ZHANG Zhong-chen^{3*}

1. College of Resources and Environment, Northeast Agricultural University, Harbin 150030, China

2. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130012, China

3. College of Agriculture, Northeast Agricultural University, Harbin 150030, China

Abstract Chlorophyll is an important pigment in vegetation photosynthesis, and the traditional laboratory method needs destructive sampling and complex operation. By constructing a high-precision SPAD spectral estimation model, the real-time non-destructive monitoring of chlorophyll content in rice leaves can be realized. In this paper, the data of five key stages of transplanting, tillering stage, jointing stage, booting stage and heading stage were obtained from rice under different nitrogen levels in Heilongjiang Province. The reflectance spectrum data of rice leaves were measured by SVC HR768i spectral radiometer with a spectral detection range of 350~2 500 nm. The spectrum of the blade was measured directly by the handheld blade spectrum detector with its own light source, which was built-in halogen lamp. The SPAD value of rice leaves was measured synchronously by SPAD-502 hand-held chlorophyll meter. Leaf water is the basic raw material of plant photosynthesis, and the decrease of leaf water content will affect the normal photosynthesis of plant, resulting in the decrease of chlorophyll content and the indirect effect of water content on chlorophyll content. Therefore, the chlorophyll sensitive band and the range of water absorption are combined as the input of SPAD. The Random Forest model is an algorithm based on multiple classification trees. In the process of sampling, the algorithm includes two completely random processes, of which one is that the sampling process is carried out with a return sampling, and the other is that the sample may be repeated, and the other is random when we select the independent variables. In this paper, the spectral reflectance of rice leaves is extracted by continuum removal (CR), and the characteristic parameters of reflectance spectrum and vegetation index of rice leaves are extracted by taking into account the visible and near infrared bands. The correlation between spectral indices and SPAD was analyzed and the SPAD hyperspectral estimation model with different inputs was constructed by the Random Forests. Results are: (1) The correlation coefficient between SPAD and spectral reflectance of rice leaves was above 0.75 in the range of chlorophyll sensitive band (600~690 nm), red edge region (720~760 nm) and water absorption band (1 400~1 490, 1 900~1 980 nm). (2) In the correlation analysis between spectral parameters and SPAD, the correlation between, NDVI, DP₂ and SPAD value of rice leaves was the best, and the correlation coefficients were 0.811 and 0.808; (3) The Random Forests model with CR_(V1, V2, V3, V4) combined with water spectral information had the highest accuracy and R² was 0.715, RMSE was 2.646, which could be used as a chlorophyll prediction model for rice leaves. The results revealed the spectral response mechanism of different varieties of rice, provided a high precision inversion method of SPAD values of rice leaves, and provided technical support for monitoring and regulating the normal growth process of rice in Northeast China.

Keywords Hyperspectral; Random Forests; Relative chlorophyll content; Continuum removal; Rice

(Received Jan. 8, 2019; accepted May 11, 2019)

* Corresponding authors