

文章编号: 1672-8785(2020)11-0033-11

基于红边光谱特征和 XGBoost 算法的 冬小麦叶绿素浓度估算研究

郭宇龙^{1,2} 李岚涛^{1*} 陈伟强^{1,2} 崔佳琪^{1,2} 王宜伦^{1*}

(1. 河南农业大学资源与环境学院, 河南 郑州 450002;

2. 河南省土地整治与生态重建工程技术研究中心, 河南 郑州 450002)

摘 要: 研究了基于氮肥效应的冬小麦不同生育期的叶绿素浓度, 探讨了 XGBoost 算法在冬小麦叶绿素浓度估算中的适用性。利用该算法构建了冬小麦叶绿素浓度的高光谱估算模型, 并将其与偏最小二乘法(Partial Least Squares, PLS)以及人工神经网络(Neural Network, NN)算法进行了对比。结果表明: (1)冬小麦的叶绿素浓度随着氮肥用量的增加而逐渐升高。(2)基于一阶微分光谱(First-order Differential Reflectance, FDR)数据集的估算模型表现最好。通过对比建模数据集与验证数据集的决定系数 R^2 和相对分析误差(Residual Predictive Derivation, RPD)发现, XGBoost 算法的效果最佳。(3)通过波段重要性分析发现, XGBoost 算法的 8 个重要波段均在 738~753 nm 范围内。与 8 个常用的红边指数相比, 通过 XGBoost 算法筛选到的 8 个一阶微分光谱波段对叶绿素浓度的准确估算起到了更加重要的作用。该算法可以作为一种有效的高光谱信息挖掘手段来估算冬小麦的叶绿素浓度。

关键词: 冬小麦; 叶绿素浓度; 高光谱; XGBoost

中图分类号: S127 文献标志码: A DOI: 10.3969/j.issn.1672-8785.2020.11.006

Research on the Estimation of Winter Wheat Chlorophyll Content Based on Red Edge Spectral and XGBoost Algorithm

GUO Yu-long^{1,2}, LI Lan-tao^{1*}, CHEN Wei-qiang^{1,2}, CUI Jia-qi^{1,2}, WANG Yi-lun^{1*}

(1. College of Resources and Environment, Henan Agricultural University, Zhengzhou 450002, China;

2. Henan Engineering Research Center of Land Consolidation and
Ecological Restoration, Zhengzhou 450002, China)

收稿日期: 2020-08-22

基金项目: 国家自然科学基金项目(41701422); 国家重点研发计划课题(2017YFD0301106); 河南省重点研发与推广专项(科技攻关)(192102310251); 河南省高等学校重点科研项目(18B420004); 河南农业大学青年英才专项基金(30500427)

作者简介: 郭宇龙(1988-), 男, 河南长垣人, 讲师, 博士, 主要研究方向为高光谱遥感和多源影像融合。

* 通讯作者: E-mail: lilantao@henau.edu.cn; wangyilunrl@163.com

Abstract: The chlorophyll concentrations in different growth stages of winter wheat based on the effect of nitrogen fertilizer are studied, and the applicability of XGBoost algorithm in estimating the chlorophyll concentration of winter wheat is discussed. A hyperspectral estimation model for chlorophyll concentration in winter wheat is constructed using this algorithm which is compared with partial least squares and artificial neural network algorithms. The results show that: (1) The chlorophyll concentration of winter wheat increases gradually with the increase of nitrogen fertilizer. (2) The estimation model based on the first-order differential spectrum data set has the best performance. The XGBoost algorithm is found to work best by comparing R^2 and RPD of the modeling and verification data sets. (3) Through the band importance analysis, it is found that the 8 important bands of XGBoost algorithm are all within the range of 738~753 nm. Compared with the 8 commonly used red-edge parameters, the 8 first-order differential spectral bands screened by the XGBoost algorithm play a more important role in accurately estimating chlorophyll concentration. This algorithm can be used as an effective hyperspectral information mining method to estimate the chlorophyll concentration of winter wheat.

Key words: winter wheat; chlorophyll content; hyperspectral; XGBoost

0 引言

作物生长状态的快速、无损检测可为农田精准施肥和养分高效管理等提供重要的数据支撑,而遥感技术的快速发展使之成为可能。当前,利用地面高光谱数据来监测作物长势、生化组分等信息时主要有两种方法:一是经验型,即采用不同的光谱组合来构建敏感指数,然后将其与目标参数一起用于构建统计模型^[1-3]。其中比较经典的是利用植被光谱的红边特征,比如红边位置、红边振幅以及红边面积等。该算法的优势是模型构建方法简单,可解释性强,且易于推广使用。但如果模型参数完全基于实测数据,那么稳定性和外推性往往难以保证。二是机理型。最常用的是基于 PROSAIL 模型的反演方法^[4-6]。这类方法在理论上一旦建立,便具有普适性。它们可以应用于各种情况,而且可以同步获取多种生理参数的反演结果。然而在实际应用中,由于模型参数很多,在参数率定的过程中难免需要实测数据的辅助,而且精度难以保证。因此,目前构建的模型仍然以经验型光谱指数模型为主^[7-8]。

虽然利用高光谱数据波段窄这一特点可以构建精度较高的经验模型,但相邻波段间的高相关性表明高光谱数据存在大量的冗余信息。由于只能利用少数几个波段来构建植被指数模型,因此在寻找“最敏感”波段的过程中,自

然会忽略其他高相关波段所携带的有用信息,导致模型的精度和鲁棒性均有待提高。所以如何寻找高效、合理的光谱信息抽取方法,成为作物叶绿素浓度高光谱估算的一个重要问题。为了充分利用全波段的有用信息,PLS 算法^[9-11]发挥了重要作用。首先压缩高光谱数据,降低数据维度,并消除高光谱冗余信息。然后利用信息更加独立、质密的自变量进行回归分析,取得了较好的应用效果。也有学者将遗传算法(Genetic Algorithm, GA)与 PLS 算法相结合,形成 GA-PLS 算法。即先通过 GA 筛选优势波段,再进行 PLS 回归,从而更好地进行数据压缩与建模^[12]。但受限于 PLS 算法原理,只能利用高光谱波段与目标参数之间的线性关系,所以模型很难取得实质性突破。近年来,随着机器学习算法的快速发展,越来越多的学者开始利用该算法的强拟合能力来构建精度更高的估算方法^[13-14]。虽然通过常用的机器学习算法(如 NN 模型)能得到很好的拟合模型,但是模型的可解释性差。且一旦为了提高精度而增加模型的复杂度,很容易会出现过拟合现象,导致模型的可推广性较差。经过长足的发展,基于分类树的机器学习算法也开始得到了越来越多的关注。从开始的决策树,到随机森林,再到提升树,模型表现越来越好。由提升树发展而来的 XGBoost 算法由于在代价函

数中加入了正则项, 可以方便地控制模型的复杂度, 从而防止过拟合。该算法的易用性和高性能已经在多个定量遥感领域取得了优异的表现^[15-16]。同时, 在模型建立过程中生成的不同模型树重要性指数, 为 XGBoost 算法提供了良好的模型可解释性, 也为敏感光谱信息的探索提供了一种有力途径。

因此, 为了进一步提升冬小麦叶绿素浓度估算模型的精度, 并探讨非线性状态下冬小麦叶绿素浓度的敏感波段位置, 本文基于三个研究区 2017~2019 年冬小麦氮肥梯度大田试验, 利用 XGBoost 算法以及红边波段反射率光谱构建了冬小麦叶绿素浓度估算模型, 并将其与 PLS 及 NN 模型进行了对比; 同时综合红边特征信息, 探讨了冬小麦叶绿素浓度的高光谱敏

感波段分布。该研究为叶绿素浓度的高效、精确诊断以及后续模型构建提供了参考。

1 材料与方法

1.1 试验设计与方法

2017~2019 年分别在河南省的鹤壁市、原阳县和温县共开展了 3 项冬小麦氮肥梯度田间试验, 涉及不同的小麦品种、试验年份、生态区域和氮素水平(见表 1)。此外, 为了探究冬小麦生育期间冠层的高光谱差异效应, 并利用该技术精确监测地上部生物量和确定光谱监测的有效生育期, 所有氮肥田间试验均在分蘖期、拔节期、孕穗期和灌浆期开展。各试验处理均重复 3 次, 并随机区组排列。除氮肥外, 分别按 120 kg/hm² (P₂O₅) 和 90 kg/hm² (K₂O)

表 1 冬小麦试验季的基本信息情况

年份	地点	品种	土壤状况	氮素水平	测试时期
2017~2018	鹤壁	郑麦 369	有机质: 19.7 g/kg pH 值: 7.4 全氮: 1.14 g/kg 速效磷: 0.029 g/kg 速效钾: 0.152 g/kg	N0 (0), N75 (75), N150 (150), N225 (225), N300 (300)	分蘖期 (2018/01/12)
					拔节期 (2018/03/19)
					抽穗期 (2018/04/18)
					灌浆期 (2018/05/13)
2017~2018	原阳	平安 11 号	有机质: 17.9 g/kg pH 值: 7.3 全氮: 1.02 g/kg 速效磷: 0.029 g/kg 速效钾: 0.145 g/kg	N0 (0), N90 (90), N180 (180), N270 (270), N360 (360)	分蘖期 (2018/01/10)
					拔节期 (2018/03/14)
					抽穗期 (2018/04/15)
					灌浆期 (2018/05/08)
2018~2019	温县	平安 11 号	有机质: 18.4 g/kg pH 值: 7.1 全氮: 1.06 g/kg 速效磷: 0.033 g/kg 速效钾: 0.159 g/kg	N0 (0), N60 (60), N120 (120), N180 (180), N240 (240), N300 (300)	分蘖期 (2018/12/15)
					拔节期 (2019/03/16)
					抽穗期 (2019/04/23)
					灌浆期 (2019/05/21)

施入磷肥和钾肥。供试肥料品种分别为控释尿素(含 44% 的 N)、过磷酸钙(含 12% 的 P_2O_5)和氯化钾(含 60% 的 K_2O)。在小麦播种前,将所有肥料均作为基肥一次性施入,以免多次追施肥料对小麦生长发育及冠层高光谱测试连续性造成影响。

1.2 数据测量方法

1.2.1 冬小麦冠层的高光谱测定

在上述各生育期,采用美国 ASD 公司生产的 HandHeld 2 手持式地物光谱仪分别测试了冬小麦冠层的高光谱反射率。各小区选取 6 处具有代表性的小麦测试样方,每处采集 10 条光谱曲线,并将其平均值作为该小区的冠层高光谱测试值。将光谱仪探头置于距小麦冠层约 1.0 m 处,并在天气晴朗且太阳高度角变幅较小的 11:00~14:00 时段进行测试。各小区测试前后均采用 30 cm×30 cm 的 $BaSO_4$ 型标准白板进行光谱校正,以降低仪器自身和大气环境的噪声干扰。HandHeld 2 光谱仪的波段范围为 325~1075 nm,波长准确度为 1 nm,光谱分辨率小于 3.0 nm@700 nm,视场角为 25°。为提高光谱监测的准确度,去除了信噪比较低的 325~399 nm 和 951~1075 nm 波段。本文选用 400~950 nm 波段来开展冬小麦地上部生物量的高光谱定量反演与精度分析。

1.2.2 叶绿素浓度测定

冠层光谱测试结束后,采集各小区的冬小麦叶片样品并测试其叶绿素浓度(单位为 mg/g)。每次取小麦植株主茎的第三片完全展开叶 20 片,并将其带回实验室洗净、擦干。然后称取 0.5 g 左右的叶片并将其装入 50 mL 比色管中,接着加入丙酮-无水乙醇混合浸提液定容至 50 mL。将其置于黑暗处遮光浸提约 48 h,直到叶片完全发白为止。采用 722 型紫外分光光度计于 663 nm 和 645 nm 处测量吸光度,并计算叶绿素 a、叶绿素 b 以及总叶绿素(叶绿素 a 与 b)浓度。

1.3 数据处理与分析

1.3.1 建模数据集与验证数据集的划分

综合试验年份、地理位置以及小麦品种等信息,并考虑到模型自变量与因变量的变化范围,将各个生育期、各个氮梯度的数据随机分为 3 份。其中 2 份(样本数为 140)用于建模,1 份(样本数为 70)用于验证。

1.3.2 模型构建与应用

本文采用 PLS、NN 和 XGBoost 三种回归方法对红边波段反射率以及叶绿素浓度进行拟合。其中,PLS 是一种多元线性统计回归方法,其主要优势是通过压缩数据来消除自变量之间的自相关,从而充分利用高光谱数据的波段信息来获取稳定、高精度的拟合效果。而且其模型的可解释性强^[17],在作物生理参数估算、病虫害与胁迫诊断中发挥着重要作用^[18-21]。NN 是一种常用的监督学习算法。它通过模拟神经元传导信息的过程来对不同的输入信号进行加权处理,再通过误差反馈最终训练得到合适的网络。由于通常能得到比传统线性及非线性拟合更高的精度,NN 也得到了广泛的应用^[22-25]。XGBoost 是一种基于提升树的学习算法^[26]。在进行回归拟合时,首先构建回归树,然后利用提升算法不断修正拟合效果。由于提升树建立在上一步模型残差的基础上,因此可以不断提高拟合精度。此外,通过用简单的得分判别进行剪枝以及在惩罚函数中直接加入正则项等措施,使 XGBoost 在保证精度的同时可以很好地控制过拟合。最后,XGBoost 也保留了树模型的高可解释性。通过选择自变量的数量生成重要性值,便于模型分析,因此它在定量遥感领域受到了越来越多的关注^[27-30]。

除原始反射率光谱 R 外,本文同时计算了一阶微分光谱和连续统去除光谱(Continuum-removed Reflectance, CR),并将其作为模型的自变量进行输入。为了与直接采用原始光谱以及将变换光谱作为输入的方式进行对比,同时选取了一系列红边指数来进行模型比较。它们包括红边位置(Red Edge Position, REP)、红边振幅(FDR_{REP})、红边面积(Sum of the FDR, SDR)、红边左侧面积(Sum of the FDR with the left of

the peak, LSDR)、红边右侧面积(Sum of the FDR with the right of the peak, RSDR)、红边左右面积比值(Ratio of the LSDR and the RSDR, RISDR)、红边左右面积差值(Differential of the LSDR and the RSDR, DISDR)以及归一化红边面积差值(Normalized Differential of the LSDR and the RSDR, NDSDR)。其中, REP 为一阶微分光谱最大值所对应的波长, 红边振幅为 REP 所对应的一阶微分值。SDR、LSDR、RSDR、RISDR、DISDR 和 NDSDR 的表达式如下^[31]:

$$SDR = \int_{680}^{760} \frac{dR\lambda}{d\lambda} d\lambda \quad (1)$$

$$LSDR = \int_{680}^{REP} \frac{dR\lambda}{d\lambda} d\lambda \quad (2)$$

$$RSDR = \int_{REP}^{760} \frac{dR\lambda}{d\lambda} d\lambda \quad (3)$$

$$RISDR = LSDR / RSDR \quad (4)$$

$$DISDR = LSDR - RSDR \quad (5)$$

$$NDSDR = DISDR / SDR \quad (6)$$

式中, REP 代表红边位置(nm), λ 代表波长。

用决定系数 R^2 表示模型估算结果对实测结果的解释程度。 R^2 越高, 表明估算结果对实测结果的解释程度越高。用均方根误差(Root Mean Square Error, RMSE)和 RPD 表征模型的精确程度。RMSE 表示模型的误差大小, 其值越小表明误差越低。RPD 为样本标准差与 RMSE 的比值, 用于表征模型的鲁棒性。RPD < 1.4 表示鲁棒性低; 1.4 < RPD < 2.0 表示模型的鲁棒性可用; RPD > 2.0 表示模型的鲁棒性较好, 预测精度高。

光谱数据的基本处理与分析在 Excel 2016 软件中进行; PLS、NN 和 XGBoost 回归算法在 Python 3.6 软件中进行。通过反复尝试, 最终确定 PLS 模型使用 10 个主成分分量来完成拟合; NN 模型采用 1 个隐含层和 20 个隐藏节点来进行拟合; XGBoost 模型使用树模型作为基分类器, 将最大模型深度设置为 4, 其回归树个数为 400。

2 结果与讨论

2.1 施氮对冬小麦不同生育期叶绿素浓度的影响

施氮对冬小麦各个生育期的叶绿素浓度有显著影响(见图 1)。随着施氮量的增加, 各试验区的小麦叶绿素浓度均明显提高。经过施氮处理后, 2017~2018 年鹤壁点分蘖期、拔节期、抽穗期和灌浆期的叶绿素浓度平均值分别比 N0 时提高了 14%、18%、18% 和 25%; 2017~2018 年原阳点四个时期的增幅分别为 22%、24%、24% 和 24%; 2018~2019 年温县点分蘖期、越冬期、拔节期、抽穗期和灌浆期的增幅分别为 21%、43%、35%、42% 和 35%。

2.2 模型精度分析

表 2 列出了 PLS、NN 和 XGBoost 三种模型的精度对比情况。从建模数据集来看, 三个模型在原始光谱、一阶微分光谱以及连续统去除光谱三个数据集中均有不错表现, R^2 都能达到 0.8 以上, 说明三个模型得到的估算结果均能较好地对实测叶绿素浓度进行解释。从整体上看, 一阶微分光谱的建模精度最高, 连续统去除光谱次之, 原始光谱误差最大, 表明光谱本身的系统性偏差对模型构建属于噪声误差, 有必要将其剔除。从 RMSE 和 RPD 两个指标来看, PLS 和 NN 模型接近, 鲁棒性可用; 而 XGBoost 模型的鲁棒性较好。

从三个光谱数据集来看, 一阶微分光谱的整体拟合精度优于连续统去除光谱, 而原始光谱数据相对较差, 说明这两种光谱处理方法对于突出高光谱有效信息和抑制噪声来说是真实有效的。验证数据集的精度与建模数据集接近, 但有所下降。PLS 模型的精度仍在可接受范围内, 鲁棒性可用; NN 模型的鲁棒性在可用和较好范围内; 而 XGBoost 模型仍整体处在鲁棒性较好的区间内。与 PLS 相比, XGBoost 模型验证数据集的 R^2 在原始光谱、一阶微分光谱以及连续统去除光谱数据集中分别提高了 14.9%、29.9% 和 12%; 与 NN 模型相比, 分别提高了 2.9%、16% 和 1.1%。而三个数据集的 RMSE 分别比 PLS 模型降低了 17.7%、18.5% 和 17.8%, 且分别比 NN 模型降低了

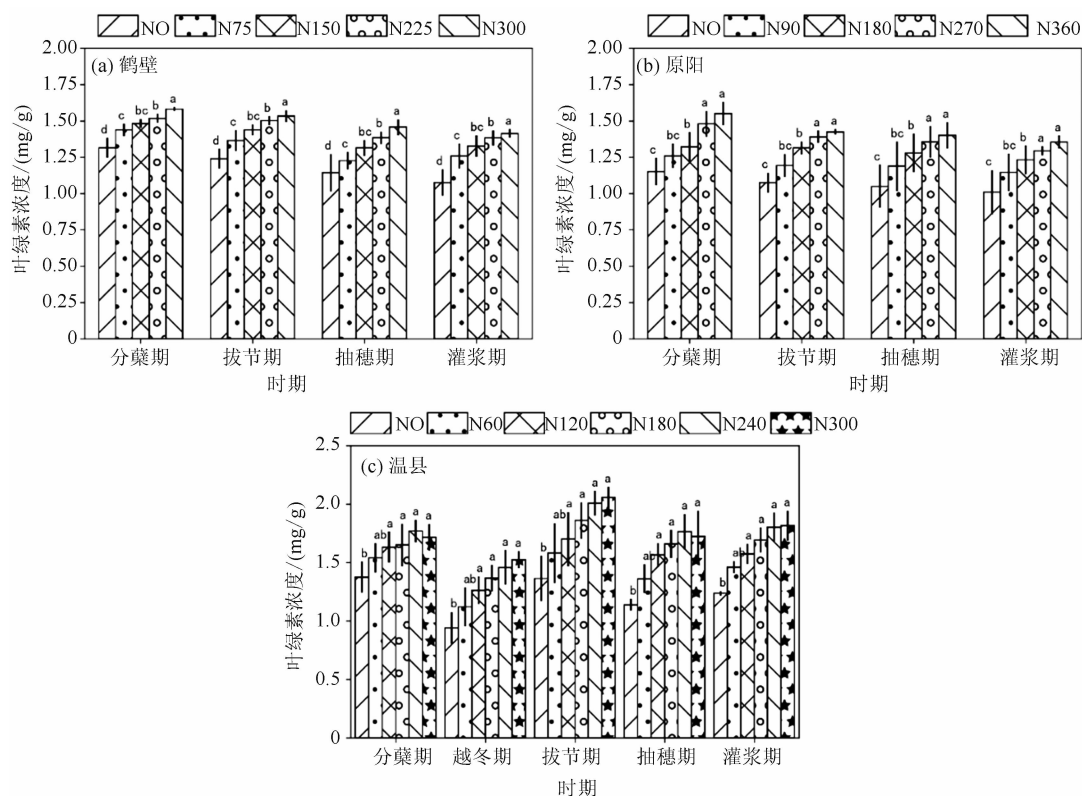


图 1 基于氮肥效应的冬小麦不同生育期叶绿素浓度分析

表 2 模型精度对比

		PLS			NN			XGBoost		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
建模数据集	R	0.818	0.149	1.664	0.833	0.143	1.736	0.913	0.107	2.317
	FDR	0.837	0.149	1.667	0.872	0.142	1.744	0.930	0.107	2.327
	CR	0.863	0.149	1.671	0.864	0.142	1.753	0.893	0.105	2.370
验证数据集	R	0.649	0.151	1.676	0.724	0.127	1.986	0.745	0.124	2.036
	FDR	0.609	0.150	1.679	0.683	0.126	1.996	0.792	0.123	2.061
	CR	0.629	0.149	1.699	0.697	0.126	2.001	0.705	0.122	2.066

2.5%、3.1%和3.2%。

本文对原始光谱、一阶微分光谱以及连续统去除光谱三个数据集的整体模型偏差(实测与预测叶绿素浓度的差值)直方图进行了统计(见图2)。可以看出,与PLS和NN模型相比,XGBoost模型主要有两点表现更加优异:(1)在横坐标的原点左右,XGBoost模型的直方图最集中,说明XGBoost模型的低误差点在整体中所占的比例最大,而NN和PLS模型的误差直方图峰值有所偏移。(2)在直方图尾部,PLS和NN模型经常出现异常的高值区(即图2

中的黑色圆形标记区域)。这些区域表示PLS和NN模型中存在一定量的偏差相对较大的点,而XGBoost模型中此类点的个数较少,进一步说明该模型在冬小麦叶片叶绿素浓度估算中具有较高的精度和鲁棒性。在三组数据中,一阶微分光谱数据的误差直方图(见图2(b))最为集中,与表2中的整体精度评估结果一致。

综上所述,XGBoost模型在一阶微分光谱数据集中的表现最优。下面结合冬小麦的红边光谱特征以及XGBoost模型的波段重要性分

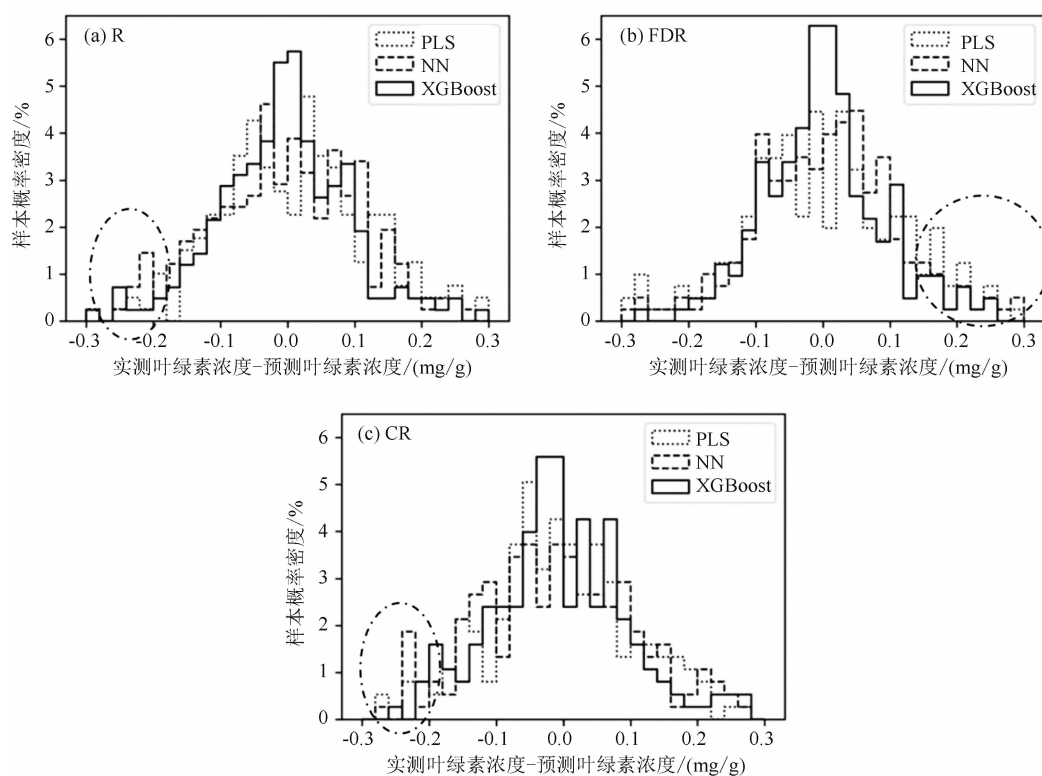


图 2 模型偏差概率密度直方图: (a)原始光谱数据集; (b)一阶微分光谱数据集; (c)连续统去处光谱数据集

析, 进一步研究由一阶微分光谱数据集构建的 XGBoost 估算模型。

2.3 波段重要性分析

在梯度提升树创建完成后, XGBoost 可以直接得到每个变量的重要性。重要性衡量的就是变量在提升树构建中的价值。对于叶绿素浓度光谱估算模型而言, 一个波段用于构建决策树的次数越多, 其重要性就越高。XGBoost 算法得到的波段重要性如图 3 所示。可以看出, 重要性较高的波段集中在 730~755 nm 之间, 与相关性较高的区间较好地重叠。但在相关性最高的 730 nm 左右, 重要性并不高。通过对比 XGBoost 和 PLS 算法的精度结果可以发现, XGBoost 算法在训练过程中不仅充分利用了高光谱与叶绿素浓度之间的线性相关性, 而且还挖掘和利用了红边波段与叶绿素浓度的非线性映射关系。

通过统计分析筛选出重要性大于整体均值加方差的波段, 并将其作为信息集中波段。可以发现, 信息集中波段分别为 738 nm、739

nm、740 nm、741 nm、745 nm、746 nm、752 nm 以及 753 nm。而在以前的大量研究中, 会用到相近波段的反射率信息。例如, MERIS 地表叶绿素指数^[32-33]和改进的叶绿素吸收比值指数^[33]都用到了 750 nm 波段; 红边对称指数^[34]用到了 755 nm 波段; 文献[35]中的研究发现 750 nm 左右的反射率接近近红外波段平台区, 可以很好地反映叶片健康程度; 文献[36]中的研究发现 735 nm 波段对高叶绿素浓度叶片比较敏感; Volgelmann J 等人^[37]利用 740 nm 波段构建的比值植被指数在叶片和冠层尺度上都取得了较好的应用效果^[34]; 欧空局发射的哨兵 2 号卫星搭载的多光谱载荷也将 740 nm 作为一个反映植被红边信息的重要波段^[38]。这些研究表明 XGBoost 算法具有较好的应用潜力。

2.4 红边特征及分析

在以前的研究中, 红边特征参数作为一个重要指标, 常常参与作物参数估算并取得了优异的模型效果。为了深入探讨红边特征与 XG-

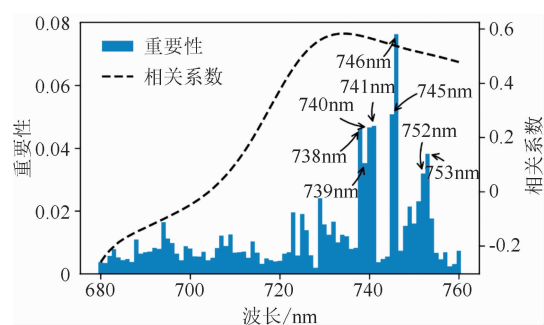


图 3 重要性直方图与红边-叶绿素浓度相关性曲线

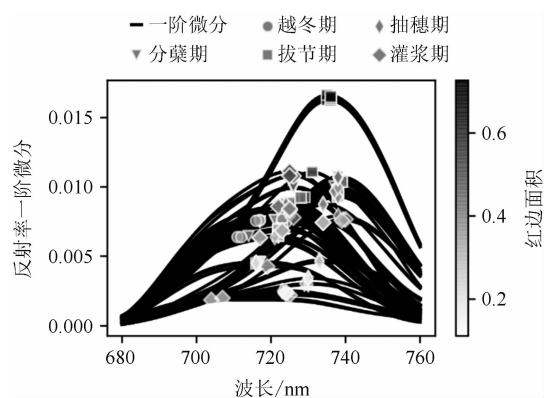


图 4 红边波段的一阶微分光谱及红边指数

Boost 一阶微分光谱模型的关联, 绘制了建模数据集中所有红边光谱的一阶微分曲线(见图 4), 并突出显示了红边位置的点。其中, 红边点的横坐标代表红边位置, 纵坐标代表红边振幅。采用红边点亮度来表示红边面积。亮度越高代表面积越小, 反之亦然。

从冬小麦的不同生育期来看, 分蘖期和越冬期小麦处在营养生长的早期阶段, 叶片覆盖度低, 光能吸收能力较弱。因此, 红边光谱信息较弱, 红边位置更偏向短波方向, 红边振幅和红边面积也相对较小。而在拔节期和抽穗期, 小麦生长迅速, 群体不断扩大, 营养生长旺盛, 叶绿素浓度迅速增加, 吸收增强, 因此冬小麦光谱的红边位置偏向更长的波段, 红边振幅和红边面积也更大。抽穗期以后, 冬小麦开始从营养生长转为生殖生长, 冠层下部叶片开始衰老死亡, 叶绿素减少, 覆盖度降低。反映在光谱上就是灌浆期的“红移”现象停止, 红边点开始回归较短波段, 红边振幅和红边面积也相应地有所减小。在冬小麦的不同生育

期, 红边指数体现出一定的规律性, 这也是红边参数模型的构建机理。

从整体上看, 红边特征在 680~760 nm 波段可以分为三个区间: (1) 680~700 nm 的第一个稳定区间。在该区间内, 反射率光谱开始进入红边区域, 整体加速抬升, 但尚未达到红边位置。(2) 700~740 nm 的红边点集中区间。在该区间内, 光谱陆续达到红边位置, 一阶微分值先后到达峰值, 随后开始下降。但由于不同样本之间的差异, 峰值位置和高度均有所不同, 一阶微分曲线形态复杂。(3) 740~760 nm 的第二个稳定区间。在该区间内, 光谱开始走出红边位置, 变化趋于稳定, 一阶微分值持续下降。从红边点在波长—一阶微分空间(见图 4)中的分布可以看出, 红边位置集中在红边特征的第二个区间。而最集中的区域在 720~740 nm 范围内, 与图 3 中的高相关性区间高度重合。红边位置的几何意义是指相邻波段间变化最大的光谱位置, 而且它与冬小麦的生育期及营养状况之间也有密切联系, 导致不同营养状况及生育期下的红边振幅和红边面积两个要素在这个区段内存在明显差异。二者共同决定了这部分是叶绿素浓度的高敏感波段。

然而值得注意的是, XGBoost 算法得到的高重要性波段都集中在 737 nm 右侧(见图 3)。除少量点重叠之外, 大多数都避开了红边特征的第二区间。这表明与第二区间内密集的红边特征参数相比, 第三区间内的一阶微分光谱与叶绿素浓度之间存在更加深入的非线性关联。从模型验证结果来看, XGBoost 算法可以较好地挖掘这部分非线性关联, 并能获得精度更高、鲁棒性更好的冬小麦叶绿素估算模型。

2.5 基于 XGBoost 有效波段和红边参数的冬小麦叶绿素预测精度分析

结合图 3 和图 4, 可以得到一个初步结论: 对 XGBoost 模型而言, 红边位置集中的波段对模型的贡献并不是最大的。而在实际应用中, 红边指数通常不直接以某个波段的反射率作为模型自变量输入, 而是计算一些形状参

数, 并构建其与目标因变量之间的统计回归模型。为了进一步对红边参数和筛选到的重要性集中波段进行对比, 并探清 XGBoost 模型构建的非线性映射中光谱形状特征和特征波段光谱值的贡献情况, 将所有的 8 个红边指数以及筛选到的 8 个波段的一阶微分值共同作为自变量, 利用建模数据集以及 XGBoost 模型进行了拟合, 并对结果进行了对比。其中 8 个波段光谱反射率的一阶微分值代表光谱值信息, 而 8 个红边指数则代表光谱形状特征。

拟合结果显示, 模型的 R^2 值为 0.940, RMSE 值为 0.107, RPD 值为 2.327, 即得到了较高的精度和鲁棒性, 且模型构建的映射关系稳定可信。其中 16 个自变量的重要性指标如图 5 所示。直观来看, 在 8 个红边指数中, REP、 FDR_{REP} 、LSDR 以及 RISDR 等 4 个变量的重要性相对更高, 但普遍低于 8 个一阶微分光谱波段值的重要性。同样将所有重要性参数的均值加方差作为阈值, 对高重要性自变量进行筛选, 发现 745 nm、746 nm 以及 753 nm 三个波段的反射率一阶微分值与其余变量相比具有较高的重要性。

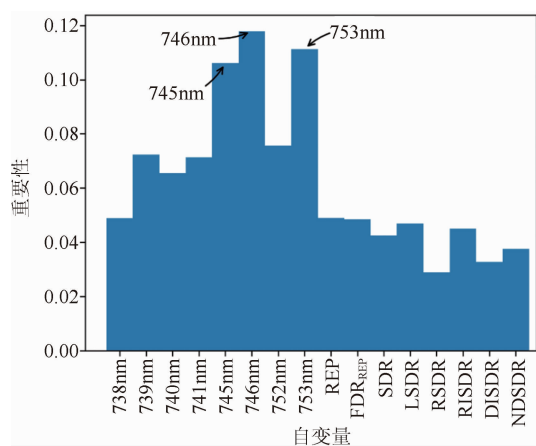


图 5 筛选的 8 条一阶微分光谱以及 8 个红边指数的重要性

上述分析表明, 虽然细化程度高的光谱红边指数在更多的线性模型应用中得到了较高的精度, 但与更加复杂的光谱红边指数相比, 直接使用特定波段的一阶微分光谱, 更有利于 XGBoost 模型对叶绿素浓度的拟合。这说明

XGBoost 模型具有更高的拟合精度和鲁棒性, 而且模型本身的复杂程度可以得到有效控制, 有利于实现推广应用。

3 结论

本文以 2017~2018 年鹤壁与原阳试验点, 以及 2018~2019 年温县试验点采集到的 210 组冬小麦叶绿素浓度和红边光谱为基础, 对比分析了 PLS、NN 以及 XGBoost 三种回归模型。结果表明:

(1) 冬小麦叶片的叶绿素浓度受氮素营养影响显著。随着氮肥用量的增加, 叶绿素浓度呈现逐步升高的趋势。

(2) 从 R^2 、RMSE 和 RPD 等 3 个精度评价指标来看, XGBoost 模型在自变量结实程度、绝对误差和模型鲁棒性方面均取得了最佳效果。

(3) XGBoost 模型得到的重要性波段集中在 740~750 nm 左右, 与目前主流的红边特征波段位置没有发生大面积重叠, 表明该模型可以在有效利用线性相关性的基础上深入挖掘光谱特征与叶绿素浓度之间的非线性关系。

(4) 将 8 个红边指数与 8 个高重要性波段同时代入 XGBoost 回归模型, 发现高重要性波段集中在 745 nm、746 nm 以及 753 nm 的光谱一阶微分值处。这表明与当前大量使用的光谱红边指数相比, 一阶微分光谱本身包含的信息更容易被 XGBoost 模型获取, 并可实现鲁棒性更高、更易于推广的冬小麦叶绿素浓度估算模型。

参考文献

- [1] Huang W, Lamb D W, Niu Z, et al. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging [J]. *Precision Agriculture*, 2007, **8** (4-5): 187-197.
- [2] Feng W, Yao X, Zhu Y, et al. Monitoring leaf nitrogen status with hyperspectral reflectance in wheat [J]. *European Journal of Agronomy*, 2008, **28**(3): 394-404.

- [3] Montesinos-López O A, Montesinos-López A, Crossa J, et al. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data [J]. *Plant methods*, 2017, **13**(1): 4.
- [4] Berger K, Atzberger C, Danner M, et al. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study [J]. *Remote Sensing*, 2018, **10**(1): 85–111.
- [5] Huang J, Ma H, Sedano F, et al. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model [J]. *European Journal of Agronomy*, 2019, **102**: 1–13.
- [6] Berger K, Atzberger C, Danner M, et al. Model-based optimization of spectral sampling for the retrieval of crop variables with the PROSAIL model [J]. *Remote Sensing*, 2018, **10**(12): 2063–2077.
- [7] Jay S, Maupas F, Bendoula R, et al. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping [J]. *Field Crops Research*, 2017, **210**: 33–46.
- [8] Li H, Liu G, Liu Q, et al. Retrieval of winter wheat leaf area index from Chinese GF-1 satellite data using the PROSAIL model [J]. *Sensors*, 2018, **18**(4): 1120–1138.
- [9] Caporaso N, Whitworth M B, Fisk I D. Protein content prediction in single wheat kernels using hyperspectral imaging [J]. *Food chemistry*, 2018, **240**: 32–42.
- [10] 尼加提, 卡斯木, 师庆东, 等. 基于高光谱特征和偏最小二乘法的春小麦叶绿素含量估算 [J]. *农业工程学报*, 2017, **33**(22): 208–216.
- [11] 付元元, 王纪华, 杨贵军, 等. 应用波段深度分析和偏最小二乘回归的冬小麦生物量高光谱估算 [J]. *光谱学与光谱分析*, 2013, **33**(5): 1315–1319.
- [12] Nunez J, Otazu X, Fors O, et al. Multiresolution-based image fusion with additive wavelet decomposition [J]. *IEEE Transactions on Geoscience and Remote sensing*, 1999, **37**(3): 1204–1211.
- [13] 王丽爱, 马昌, 周旭东, 等. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算 [J]. *农业机械学报*, 2015, **46**(1): 259–265.
- [14] Ashourloo D, Aghighi H, Matkan A A, et al. An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, **9**(9): 4344–4351.
- [15] Jia Y, Jin S, Savi P, et al. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation [J]. *Remote Sensing*, 2019, **11**(14): 1655–1680.
- [16] Jin Q, Fan X, Liu J, et al. Estimating tropical cyclone intensity in the South China Sea using the XGBoost model and FengYun satellite images [J]. *Atmosphere*, 2020, **11**(4): 423–445.
- [17] Geladi P, Kowalski B R. Partial least-squares regression: A tutorial [J]. *Analytica chimica acta*, 1986, **185**: 1–17.
- [18] McGoverin C M, Engelbrecht P, Geladi P, et al. Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics [J]. *Analytical and Bioanalytical Chemistry*, 2011, **401**(7): 2283–2289.
- [19] Shahin M A, Symons S J. Detection of fusarium damage in Canadian wheat using visible/near-infrared hyperspectral imaging [J]. *Journal of Food Measurement & Characterization*, 2012, **6**(1–4): 3–11.
- [20] Li L, Wang S, Ren T, et al. Evaluating models of leaf phosphorus content of winter oilseed rape based on hyperspectral data [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, **32**(14): 209–218.
- [21] Abdel-Rahman E M, Mutanga O, Odindi J, et al. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data [J]. *Computers and Electronics in Agriculture*, 2014, **106**: 11–19.

- [22] Liu H, Zhu H, Wang P. Quantitative modelling for leaf nitrogen content of winter wheat using UAV-based hyperspectral data [J]. *International Journal of Remote Sensing*, 2017, **38**(8-10): 2117-2134.
- [23] Li Z, Nie C, Wei C, et al. Comparison of four chemometric techniques for estimating leaf nitrogen concentrations in winter wheat (*Triticum Aestivum*) based on hyperspectral features [J]. *Journal of Applied Spectroscopy*, 2016, **83**(2): 240-247.
- [24] 沈文颖, 李映雪, 冯伟, 等. 基于因子分析-BP神经网络的小麦叶片白粉病反演模型 [J]. *农业工程学报*, 2015, **31**(22): 183-190.
- [25] 杨贵军, 赵春江, 邢著荣, 等. 基于 PROBA/CHRIS 遥感数据和 PROSAIL 模型的春小麦 LAI 反演 [J]. *农业工程学报*, 2011, **27**(10): 88-94.
- [26] He T, Xie C, Liu Q, et al. Evaluation and comparison of random forest and A-LSTM networks for large-scale winter wheat identification [J]. *Remote Sensing*, 2019, **11**(14): 1665-1686.
- [27] Zamani J, Cao C, Ni X, et al. PM2.5 Prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data [J]. *Atmosphere*, 2019, **10**(7): 373-392.
- [28] Su H, Yang X, Lu W, et al. Estimating subsurface thermohaline structure of the global ocean using surface remote sensing observations [J]. *Remote Sensing*, 2019, **11**(13): 1598-1620.
- [29] Li R, Cui L, Fu H, et al. Estimating high-resolution PM1 concentration from Himawari-8 combining extreme gradient boosting-geographically and temporally weighted regression (XGBoost-GT-WR) [J]. *Atmospheric Environment*, 2020, **229**: 117434.
- [30] 兰玉彬, 朱梓豪, 邓小玲, 等. 基于无人机高光谱遥感的柑橘黄龙病植株的监测与分类 [J]. *农业工程学报*, 2019, **35**(3): 92-100.
- [31] Feng W, Guo B B, Wang Z J, et al. Measuring leaf nitrogen concentration in winter wheat using double-peak spectral reflection remote sensing data [J]. *Field Crops Research*, 2014, **159**: 43-52.
- [32] Dash J, Curran P. The MERIS terrestrial chlorophyll index [J]. *International Journal of Remote Sensing*, 2004, **25**(23): 5403-5413.
- [33] Wu C, Niu Z, Tang Q, et al. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation [J]. *Agricultural and forest meteorology*, 2008, **148**(8-9): 1230-1241.
- [34] Chang H J, Yong C T, Xia Y, et al. Estimating leaf chlorophyll content using red edge parameters [J]. *Pedosphere*, 2010, **20**(5): 633-644.
- [35] Ecartot M, Compan F, Roumet P. Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer [J]. *Field Crops Research*, 2013, **140**: 44-50.
- [36] Boochs F, Kupfer G, Dockter K, et al. Shape of the red edge as vitality indicator for plants [J]. *Remote sensing*, 1990, **11**(10): 1741-1753.
- [37] Vogelmann J, Rock B, Moss D. Red edge spectral measurements from sugar maple leaves [J]. *Title Remote Sensing*, 1993, **14**(8): 1563-1575.
- [38] Clevers J G, Gitelson A A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3 [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2013, **23**: 344-351.