

郭 松, 常庆瑞, 赵泽英, 等. 基于高光谱的不同生育期玉米花青素含量估测[J]. 江苏农业学报, 2024, 40(2): 303-311.  
doi:10.3969/j.issn.1000-4440.2024.02.012

## 基于高光谱的不同生育期玉米花青素含量估测

郭 松<sup>1</sup>, 常庆瑞<sup>2</sup>, 赵泽英<sup>1</sup>, 李莉婕<sup>1</sup>, 童倩倩<sup>1</sup>

(1. 贵州省农业科技信息研究所, 贵州 贵阳 550006; 2. 西北农林科技大学资源环境学院, 陕西 杨凌 712100)

**摘要:** 花青素(Anthocyanin)是玉米体内的重要色素,对花青素含量的便捷、无损估测对监测玉米长势具有重要意义。利用关中地区拔节期、大喇叭口期、抽雄期以及乳熟期玉米冠层叶片 Anth 值及高光谱数据建立多个单因素模型和多因素模型。结果表明,不同生育期玉米叶片原始光谱特征总体一致、局部不同。变换光谱的特征波段与 Anth 值相关性优于原始光谱,其中一阶导数光谱特征波段最优。连续投影算法(SPA)降维能力较好,筛选出的建模参数在2~27个。最优单因素模型与多元性线性回归模型精度均为抽雄期最优,拔节期和大喇叭口期次之,乳熟期最差。所有模型中,抽雄期基于一阶导数光谱的麻雀搜索算法-极限学习机回归(SSA-ELMR)模型精度最佳,该模型建模与验证  $R^2$  分别为 0.847、0.895,相应  $nRMSE$  为 6.44% 和 7.21%。本研究结果表明抽雄期是玉米叶片花青素含量反演的最佳时期,极限学习机能进一步提升传统模型精度。

**关键词:** 玉米; 花青素; 光谱变换; 支持向量回归; 极限学习机回归

**中图分类号:** S127 **文献标识码:** A **文章编号:** 1000-4440(2024)02-0303-09

## Estimation of anthocyanin content in maize at different growth stages based on hyperspectral technology

GUO Song<sup>1</sup>, CHANG Qing-rui<sup>2</sup>, ZHAO Ze-ying<sup>1</sup>, LI Li-jie<sup>1</sup>, TONG Qian-qian<sup>1</sup>

(1. Guizhou Agricultural Science and Technology Information Institute, Guiyang 550006, China; 2. College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China)

**Abstract:** Anthocyanin is an important pigment in maize. Convenient and non-destructive estimation of anthocyanin content is of great significance for monitoring maize growth. Based on the Anth values and hyperspectral data of maize canopy leaves at jointing stage, flare opening stage, tasseling stage and milk-ripe stage in Guanzhong area, multiple single-factor models and multi-factor models were established. The results showed that the original spectral characteristics of maize leaves at different growth stages were generally consistent, but locally different. The correlations between the characteristic bands and Anth values of the transform spectrum were better than those of the primary spectrum, and the characteristic bands of the first derivative spectrum was the best. The successive projections algorithm (SPA) had a good dimension reduction effect, and the selected modeling parameters ranged from two to 27. The accuracy of the optimal single factor model and multiple linear regression model was the best at tasseling stage, followed by jointing stage and flare opening stage, and the worst at milk-ripe stage. Among all the models, the sparrow search algorithm-extreme learning machine regression (SSA-ELMR) model based on the first derivative spectrum at the tasseling stage was the optimal model in this study. The  $R^2$  values for modeling and verification of this model were 0.847 and 0.895, respectively, and the corresponding  $nRMSE$  values were 6.44% and 7.21%. The results of this study indicate that the tasseling period was the optimal period for inverting the anthocyanin content in maize leaves, and the extreme learning machine could further improve the accuracy of traditional models.

收稿日期: 2022-12-06

基金项目: 贵州省农业科学院青年科技基金项目 [黔农科院科技创新(2022)14号]

作者简介: 郭 松(1996-), 男, 贵州贵阳人, 硕士, 研究实习员, 主要从事农业遥感研究。(E-mail) 1185716519@qq.com

通讯作者: 赵泽英, (E-mail) 623619534@qq.com

**Key words:** maize; anthocyanin; spectral transformation; support vector regression; extreme learning machine regression

花青素(Anthocyanidin),又称花色素,是植物体内的重要色素之一,主要存在于植物的茎、叶和种子中。对于植物而言,花青素能有效调节叶片的光合作用环境,促进光合作用过程,进而提高植物初级生产力<sup>[1]</sup>;当植物受到干旱和低温等环境胁迫时,花青素的抗氧化性以及渗透调节作用还能帮助植物快速修复受损组织,抵御恶劣环境<sup>[2-3]</sup>。目前花青素及其衍生产物被广泛应用于抗肿瘤药、肠道调节以及纺织物的染色<sup>[4-5]</sup>。快速识别植物体内花青素的绝对/相对含量可以监测植物的长势,传统的花青素含量检测方法如吸收光谱法<sup>[6]</sup>、高效液相色谱法以及pH示差法<sup>[7]</sup>等虽然能准确测定植物花青素含量,但效率低下,对待测样品有着不可逆的损伤,且不同方法之间互通性较差。

近年来,高光谱遥感技术的迅速发展为植物花青素含量检测提供了新思路,国内外学者在花青素的高光谱定量反演方面已经做了一定的研究。Steele等<sup>[8]</sup>基于葡萄叶片的反射率构建了4个基本的植被指数以估测其花青素含量,最优模型的验证均方根误差为2~3 nmol/cm<sup>2</sup>;Gitelson等<sup>[9]</sup>提出了一种基于绿光、红光以及“红边”数据整合的植物花青素含量估测算法,建立的估测模型平均绝对误差为4.5%~6.1%;Choi等<sup>[10]</sup>结合芥菜的高光谱影像和偏最小二乘算法,构建花青素含量估测模型,模型的最高验证决定系数达0.83;刘秀英等<sup>[11]</sup>根据偏最小二乘回归的特性,使用不同的植被指数估测开花期牡丹叶片花青素含量,其估测模型的建模与验证决定系数分别达0.87和0.81;田明璐等<sup>[12]</sup>通过两两波段组合的方式确立了苹果叶片花青素含量估测模型的建模参数,模型的最高建模决定系数为0.82。综上所述,当下有关高光谱估测植物花青素含量的研究已经较为成熟,但是仍不足之处。目前的大部分研究是针对植物的单生育期,且模型的建立多是基于构造的植被指数,从多生育期及其光谱反射率本身出发去建立花青素反演模型的研究还较少。

关中地区作为中国玉米的重要种植区,玉米的稳定高产有利于促进中国经济的稳定发展,故本研究以该地区玉米为研究对象,通过玉米叶片的原始及变换光谱提取建模参数,基于普通回归和机器学习回归算法构建多种模型,以期确立一个鲁棒性和泛化性兼备

的玉米叶片花青素含量估测模型,为中国农业的信息化提供新思路。

## 1 材料与方法

### 1.1 种植区概况与实验设计

种植区位于关中平原咸阳市梁山镇(108.117 8°E,34.608 9°N)(图1),北接渭北旱塬,南邻秦巴山区,平均海拔997 m,属暖温带大陆性季风气候,四季分明,气候温和,全年无霜期在200 d以上,年平均气温10.4℃,年平均降水570 mm,多集中于夏秋两季,作物一年一熟或两年三熟。土壤类型为红油土,适于耕作。

此次田间试验开始于2021年4月21日,划分小区36个,每个小区长8 m,宽7 m,每个小区设两个采样点,为保证数据的代表性和模型的稳健性,给每个小区施用不同的氮、磷、钾3种肥料以保证小区间玉米长势出现明显差异(图1)。其中,N<sub>0</sub>~N<sub>5</sub>氮肥(纯氮)施用梯度为:0 kg/hm<sup>2</sup>、49.95 kg/hm<sup>2</sup>、100.05 kg/hm<sup>2</sup>、150.00 kg/hm<sup>2</sup>、199.95 kg/hm<sup>2</sup>、250.05 kg/hm<sup>2</sup>;P<sub>0</sub>~P<sub>5</sub>磷肥(五氧化二磷)施用梯度分别为:0 kg/hm<sup>2</sup>、30 kg/hm<sup>2</sup>、60 kg/hm<sup>2</sup>、90 kg/hm<sup>2</sup>、120 kg/hm<sup>2</sup>、150 kg/hm<sup>2</sup>;钾肥(氧化钾)K<sub>0</sub>~K<sub>5</sub>施用梯度分别为:0 kg/hm<sup>2</sup>、25.05 kg/hm<sup>2</sup>、49.95 kg/hm<sup>2</sup>、75.00 kg/hm<sup>2</sup>、100.05 kg/hm<sup>2</sup>、124.95 kg/hm<sup>2</sup>,种植前全部肥料施入土壤,其余与当地田间管理一致。

### 1.2 项目测定与方法

不同生育期的玉米表现特征不同<sup>[13]</sup>,为探究不同生育期玉米叶片花青素含量反演模型建立的可能性,分别选择玉米的拔节期(2021年6月10日)、大喇叭口期(2021年6月29日)、抽雄期(2021年7月31日)以及乳熟期(2021年8月30日)采集数据。

1.2.1 叶片花青素含量测定 玉米叶片的花青素含量通过DualExSCIENTIFIC+植物多酚-叶绿素测量计(法国Force-A公司产品)获取,该仪器测定结果中的Anth值表示叶片花青素相对含量<sup>[14]</sup>。测定时,在每个样点选择3张正常生长的玉米叶片,避开叶脉,于叶片的上、中、下3个位置分别测量3次,取单个样点所有测量值的平均值作为该样点的Anth值。

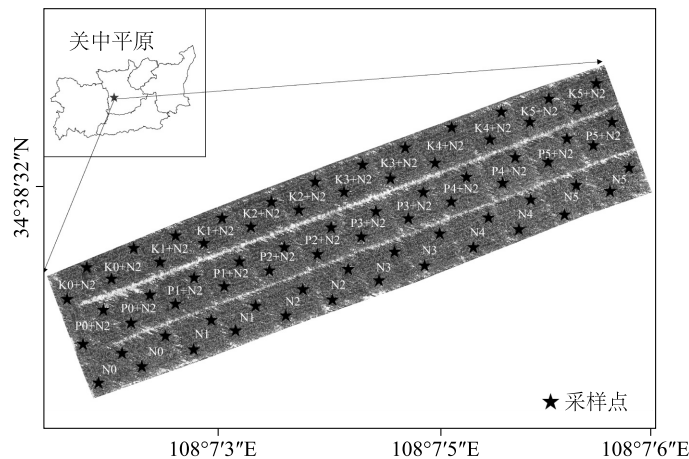


图1 研究区玉米试验田位置

Fig.1 Location of maize experiment field in the study area

1.2.2 冠层叶片高光谱反射率测定 通过 SVC HR-1024i(北京东方佳气科技有限公司产品)获取叶片高光谱反射率,该仪器采集的光谱为380~2 500 nm,当波长小于1 000 nm 时,光谱分辨率为3.5 nm,测定时,传感器探头正对待测叶片,光谱视角角设定为15°,当玉米植株较矮时探头距离待测叶片50 cm,当植株较高时,探头距离待测叶片15 cm,选择晴朗无风天的11:00~15:00进行测定,该时段天气稳定,太阳辐射总量较大,且为保证数据的科学性与可靠性,每0.5 h 进行1次白板校正,测定的叶片及位置与 Anth 值测定一致,于每片叶子的上、中、下3个位置分别获取两条高光谱曲线,取单个样点所有光谱曲线的平均值作为该样点的最终冠层光谱曲线。

### 1.3 数据处理

由于植被的研究光谱波长为400~900 nm<sup>[15]</sup>,因此对不同生育期各样点光谱曲线进行重采样,采样间隔1 nm,采样光谱波长为450~850 nm,同时为减弱光谱采样过程中的随机噪声,使用 Savitzky-Golay 二阶平滑滤波处理不同生育期各样点原始光谱(Primary spectrum,PS),并对原始光谱进行数学变换以放大原始光谱的局部特征<sup>[16]</sup>,得到基于原始光谱的连续统去除光谱(Continuum removal spectrum,CRS)<sup>[17]</sup>和一阶导数光谱(First derivative spectrum,FDS)。

将不同生育期各样点 Anth 值按降序排序,以3:1的比例进行分层抽样,各生育期均得到54个建模样本和18个验证样本(表1)。从生育期看,玉米叶片花青素含量先上升、后下降,于抽雄期达到峰值;从各统计特征看,各生育期划分的建模集与验证

集差别不大,说明划分方式较为科学,各建模集与验证集变异系数均小于10%。

表1 不同生育期 Anth 值统计特征

Table 1 Statistical characteristics of Anth values at different growth stages

生育期	样本个数	最大值	最小值	平均值	变异系数(%)
拔节期	54	0.143 7	0.064 0	0.104 6	6.276
	18	0.145 7	0.082 0	0.106 6	6.513
大喇叭口期	54	0.110 3	0.053 0	0.079 8	5.616
	18	0.113 7	0.054 7	0.081 3	5.584
抽雄期	54	0.140 7	0.070 7	0.096 8	6.244
	18	0.146 3	0.077 7	0.099 1	5.676
乳熟期	54	0.127 7	0.044 7	0.075 5	4.392
	18	0.132 3	0.050 0	0.078 4	4.047

### 1.4 模型建立

特征波段是与待反演对象相关性最高的波段<sup>[18]</sup>,本研究基于不同生育期各样点原始光谱、一阶导数光谱以及连续统去除光谱提取特征波段用以单因素模型构建,通过连续投影算法(Successive projections algorithm,SPA)降维以建立多因素模型;其中单因素模型包括一元线性函数、幂函数、指数函数、对数函数以及二次多项式函数,多因素模型包括多元线性回归(Multiple linear regression,MLR)模型、麻雀搜索算法(Sparrow search algorithm,SSA)优化的支持向量回归(Support vector regression,SVR)模型以及极限学习机回归(Extreme learning machine regression,ELMR)模型。对比各类模型的精度,筛

选反演玉米冠层叶片 Anth 值的最优模型。

单一模型评价指标具有片面性,故本研究采用决定系数( $R^2$ )和归一化均方根误差( $nRMSE$ )评价模型精度,其中决定系数范围为 $[0,1]$ ,当 $R^2$ 越接近 1,表示模型自变量对因变量的解释性越好,模型预测能力越佳;归一化均方根误差范围为 $[0,+\infty)$ , $nRMSE$ 越小,表示预测值与实测值越接近,即模型的预测能力越好<sup>[19]</sup>。

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

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}}{\bar{y}_i} \times 100\% \quad (2)$$

式中 $y_i$ 、 $\hat{y}_i$ 、 $\bar{y}_i$ 分别表示样点的实测值、预测值以及建模集或验证集实测值的平均值。

## 2 结果与分析

### 2.1 玉米冠层叶片高光谱特征

将同一生育期内的不同类型光谱曲线各点取平

均值作该生育期的代表光谱曲线(图 2)。“红边”是指红光区外叶绿素吸收减少部位到近红外高反射率部位,健康植物的光谱响应陡然增加的这一窄条带区。从光谱类型看,不同光谱类型的各生育期光谱曲线整体上特征一致,局部有差异;从原始光谱看,不同生育期玉米冠层叶片高光谱曲线总体上可见光反射率低,而近红外反射率高,二者由“红边”过渡,叶片所含色素对光能的吸收所造成可见光处反射率低,叶片内部细胞结构对近红外光多次反射导致近红外反射率高,其中以抽雄期叶片光谱特征最为典型,说明该时期的玉米冠层叶片差异较为明显;从连续统去除光谱看,不同生育期连续统去除光谱曲线均在 500 nm 和 675 nm 处形成较低的波谷,在 550 nm 处形成较高波峰,连续统去除变换后突出显示了原始光谱曲线中的“吸收谷”与“反射峰”;从一阶导数光谱看,不同生育期一阶导数光谱曲线均在 525 nm 和 725 nm 处形成较高的波峰,在 575 nm 处形成较低的波谷,其中 525 nm 和 575 nm 分别对应原始光谱“绿反射峰”的左右两端,725 nm 则对应“红边”,可见一阶导数变换光谱放大了原始光谱中反射率突变的位置。

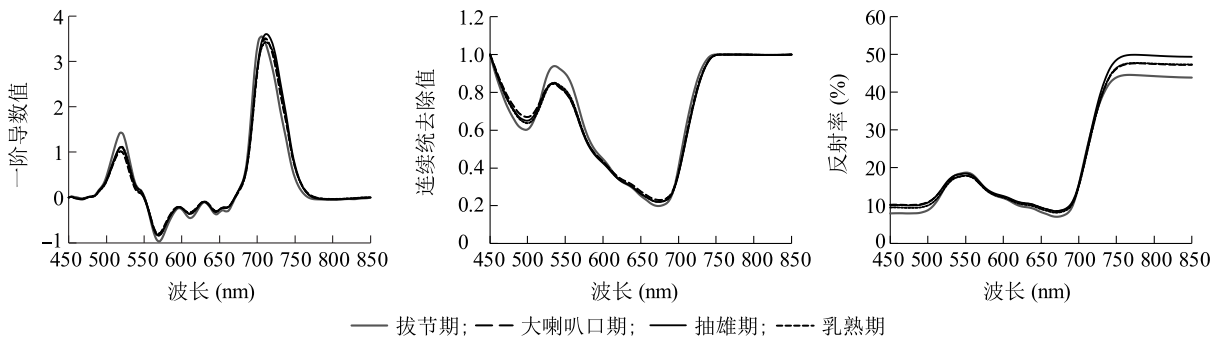


图 2 玉米叶片不同类型高光谱特征

Fig.2 Hyperspectral characteristics of maize leaves

### 2.2 不同类型光谱及特征波段与 Anth 值的相关性

2.2.1 不同类型光谱与 Anth 值的相关性 通过查阅相关系数检验表可知,样本数量为 72 时,两个变量相关系数绝对值大于等于 0.299 7 即表示二者极显著相关( $P < 0.01$ )。图 3 为不同类型光谱值与 Anth 值相关性图,可见原始光谱中大喇叭口期、抽雄期以及乳熟期通过显著性检验( $P < 0.01$ )的波段主要集中在可见光区域,其中抽雄期相关性最高,乳熟期相关性最差,拔节期通过显著性检验( $P < 0.01$ )的波段在可见光与近红外区域均有分布;一阶导数

光谱中各生育期通过显著性检验( $P < 0.01$ )的波段主要分布在可见光区域,局部分布在近红外区域,可见,相较原始光谱而言,一阶导数变换能在一定程度上提高近红外区域光谱值与 Anth 值的相关性;连续统去除光谱中不同生育期通过显著性检验( $P < 0.01$ )的波段同样集中于可见光区域,而在近红外区域,仅大喇叭口期和抽雄期通过显著性检验( $P < 0.01$ )的波段数量较多。综上所述,光谱变换在提高光谱与 Anth 值相关性方面主要作用于近红外区域。



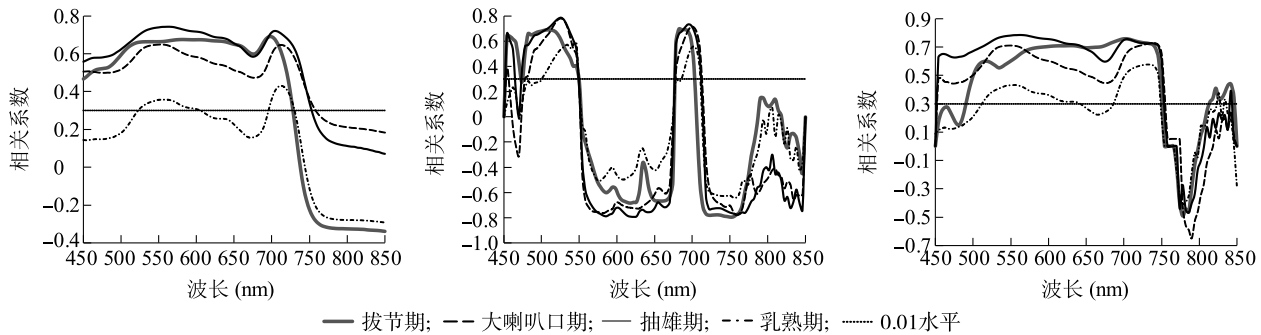


图3 不同类型光谱与 Anth 值的相关性

Fig.3 Correlation between different types of spectra and Anth values

**2.2.2 特征波段与 Anth 值的相关性** 从特征波段与 Anth 值相关性表(表2)中可看出,所有特征波段与 Anth 值的相关性均达到极显著水平( $P < 0.01$ ),波段位置主要位于红光(550 nm)与近红外(850 nm)之间,其中一阶导数光谱与连续统去除光谱的特征波段与 Anth 值的相关性要优于原始光谱特征波段,且一阶导数变换光谱的特征波段与 Anth 值全部为负相关,相应的原始光谱和连续统去除光谱特征波段与 Anth 值均为正相关;各生育期均为一阶导数光谱特征波段最优,连续统去除光谱特征波段次之,原始光谱特征波段最差,所有特征波段中与 Anth 值相关性最高的为抽雄期的一阶导数光谱特征波段(622 nm),其相关系数达-0.795。

表2 特征波段与 Anth 值的相关性

Table 2 Correlation between feature bands and Anth values

生育期	光谱类型	特征波段 (nm)	相关系数
拔节期	PS	698	0.691 **
	FDS	750	-0.767 **
	CRS	722	0.759 **
大喇叭口期	PS	554	0.650 **
	FDS	758	-0.780 **
	CRS	738	0.728 **
抽雄期	PS	562	0.744 **
	FDS	622	-0.795 **
	CRS	594	0.786 **
乳熟期	PS	714	0.430 **
	FDS	758	-0.640 **
	CRS	550	0.575 **

PS:原始光谱;FDS:一阶导数光谱;CRS:连续统去除光谱。\*\*表示极显著相关( $P < 0.01$ )。

### 2.3 单因素模型构建及精度评价

各生育期不同类型光谱下构建的单因素模型见表3,构建的单因素模型包括线性和非线性模型。从不同生育期看,除乳熟期外,拔节期、大喇叭口期以及抽雄期均为变换光谱下的单因素模型优于原始光谱下的单因素模型,其中一阶导数光谱优于连续统去除变换光谱;从模型精度看,单因素模型建模和验证决定系数( $R^2$ )为0.206~0.787,归一化均方根误差( $nRMSE$ )为8.42%~23.50%,模型预测能力普遍较差。所有模型中精度最好的是抽雄期的基于一阶导数光谱的单因素模型,其建模和验证 $R^2$ 分别为0.590和0.787,相应的 $nRMSE$ 为10.27%、8.42%,当考虑单因素模型反演玉米冠层叶片 Anth 值时,应首先使用此模型。

### 2.4 多因素模型构建及精度评价

**2.4.1 基于连续投影算法的多因素模型建模参数提取** 高光谱数据维数较多,未加处理便直接应用会导致计算成本大、建模精度差,且模型规律会被淹没在无用数据中,因此,在建模之前对高光谱数据中的有用信息进行提取十分重要。目前,高光谱降维方法主要分为特征提取和特征选择,其中特征提取通常会将原高光谱信息映射到其他空间,通过相应判定条件从新空间去提取目标参量,例如主成分分析以及奇异值分解等,该方法会改变原有数据的物理意义,使得模型参数可解释性变差,而特征选择是从原光谱数据中选择符合特征的子集,既没有改变原有数据,也保证了有用信息的纯度,例如递归特征消除算法(Recursive feature elimination algorithm, RFEA)和连续投影算法(SPA)等。

连续投影算法(SPA)是一种无监督式的前向变量

选择算法,其原理主要分为 3 步,首先从单个变量开始,确定选入变量和未选入变量,然后每次加入新的变量,计算选入和未选入变量的投影映射,最后确定最大投影即可。该方法既保证了选入变量与因变量的相关性,也确保了选入变量间的自相关性最低<sup>[20]</sup>。本研究将各生育期不同类型光谱中通过 0.01 相关性检验的

光谱值作为 SPA 的输入变量,Anth 值为因变量,以均方根误差最小为限制条件选择多因素模型建模参数。结果表明,各生育期不同类型光谱多因素模型建模参数为2~27 个,降维比在 73.269%以上,通过 SPA 降维选择,过滤了大部分次级信息(表 4)。

表 3 不同光谱类型单因素模型

Table 3 Single factor models based on different types of spectra

生育期	光谱类型	模型表达	建模精度		验证精度	
			$R^2$	$nRMSE$ (%)	$R^2$	$nRMSE$ (%)
拔节期	PS	$Y=0.052\ 1e^{0.050\ 2x}$	0.517	11.09	0.465	11.87
	FDS	$Y=-0.115\ 9x+0.154\ 6$	0.632	9.67	0.652	9.25
	CRS	$Y=0.009\ 5e^{2.895\ 7x}$	0.575	10.40	0.509	10.85
大喇叭口期	PS	$Y=0.011\ 1x^{0.683\ 1}$	0.414	13.65	0.442	13.78
	FDS	$Y=0.131\ 4e^{-1.683\ 0x}$	0.656	10.46	0.570	12.45
	CRS	$Y=0.582\ 1\ln x+0.106\ 6$	0.573	11.64	0.437	14.23
抽雄期	PS	$Y=0.000\ 1x^2+0.001\ 4x+0.036\ 3$	0.528	12.18	0.650	12.39
	FDS	$Y=0.047\ 8e^{-3.574\ 0x}$	0.590	10.27	0.787	8.42
	CRS	$Y=0.025\ 5e^{2.956\ 5x}$	0.580	10.40	0.626	11.18
乳熟期	PS	$Y=0.036\ 2e^{0.026\ 8x}$	0.211	20.37	0.141	23.50
	FDS	$Y=-0.034\ 0\ln x+0.031\ 1$	0.458	18.42	0.331	23.06
	CRS	$Y=0.031\ 5e^{1.045\ 4x}$	0.206	23.46	0.141	20.44

PS、FDS、CRS 见表 2。 $x$  表示光谱值, $Y$  表示预测的 Anth 值。 $R^2$ :决定系数; $nRMSE$ :归一化均方根误差。

表 4 SPA 降维结果

Table 4 Dimension reduction results based on successive projections algorithm (SPA)

生育期	光谱类型	建模参数		降维比(%)
		波长(nm)	个数	
拔节期	PS	558,650,690,722,850	5	95.052
	FDS	542,546,634,750	4	96.044
	CRS	494,706,774,782,846	5	95.048
大喇叭口期	PS	454,690,714	3	97.027
	FDS	470,478,630,710,718,802,830	7	93.072
	CRS	730,782	2	98.024
抽雄期	PS	450,498,522,570,626,650,662	7	93.072
	FDS	542,678,710,718,798	5	95.051
	CRS	454,458,462,466,474,478,482,486,490,506,530,570,626,650,666,674,686,702,714,730,746,750,774,782,786,790,794	27	73.269
乳熟期	PS	526,530,546,550,562,566,570,578,586,594,602,698,706,714,726,730	16	84.162
	FDS	506,546,678,710,718	5	95.053
	CRS	518,722	2	98.024

PS、FDS、CRS 见表 2。

2.4.2 多元线性回归模型构建及精度比较 以表 4 中的变量为自变量,相应 Anth 值作为响应变量,构建基于 SPA 的多元线性回归(MLR)模型,选择各生育

期下的最优 MLR 模型作为代表模型。仅乳熟期的最优 MLR 模型是基于原始光谱构建,其余生育期均基于一阶导数光谱和连续统去除光谱构建。与单因素

模型相比,MLR 模型精度整体较高,其建模与验证  $R^2$  分别提升 0.030~0.275、0.071~0.281,其中抽雄期基于连续统去除光谱的 MLR 模型是所有 MLR 模型中的最优模型,相应的建模与验证  $R^2$  分别为 0.842 和 0.870, $nRMSE$  为 9.37%、7.67%(表 5)。

2.4.3 机器学习回归模型建立及精度比较 机器学习算法本质上是一个学习策略,该策略赋予计算机模仿人类学习和反馈的能力并据此解决实际问题<sup>[21]</sup>。已有研究结果表明,相较于经典回归,机器学习算法在处理非线性回归方面具有不可替代的优越性<sup>[22-25]</sup>。

支持向量回归(SVR)是一种在高维空间中将超平面作为拟合曲线的机器学习回归算法,该算法不依靠所有数据来确立超平面,且保证了所有的数据点与该平面的距离最短;极限学习机回归(ELMR)是近年兴起的一种基于前馈神经网络的机器学习算法,完整的神经元由输入层、输出层和隐含层组成,其中输入层到隐含层的权重是随机产生,与普通的神经网络算法相比,极限学习机在学习速率和模型适用性方面表现更好<sup>[26]</sup>。本研究拟采用以上两种机器学习算法提升 MLR 模型中的最优模型精度。

表 5 各生育期的最优多元线性回归(MLR)模型

Table 5 Optimal multiple linear regression (MLR) models for each growth stage

生育期	光谱类型	系数	建模精度		验证精度	
			$R^2$	$nRMSE$ (%)	$R^2$	$nRMSE$ (%)
拔节期	FDS	0.180、-0.231、0.294、0.092、-0.135	0.662	9.27	0.735	8.02
大喇叭口期	FDS	0.053、-0.394、-0.152、-0.472、0.025、-0.029、0.161、0.278	0.692	9.89	0.641	10.94
抽雄期	CRS	638.377、1.121、0.348、-0.956、0.280、2.520、-4.397、2.131、-1.691、-0.087、0.240、0.126、0.850、-2.190、-0.642、8.910、-9.052、5.218、-1.706、0.390、-0.069、8.126、-14.439、-642.921、-25.344、85.090、-26.146、-23.054	0.842	9.37	0.870	7.67
乳熟期	PS	0.187、-0.154、0.135、-0.411、0.395、0.205、-0.376、0.327、-0.369、0.576、-0.228、0.156、0.150、-0.242、0.252、-0.266、0.152	0.733	11.76	0.612	15.57

PS、FDS、CRS 见表 2。 $R^2$ :决定系数; $nRMSE$ :归一化均方根误差。

为节约计算资源,采用麻雀搜索算法对 ELMR 的自适应度参数  $F$  和 SVR 的惩罚系数  $C$  进行迭代寻优, $F$  与  $C$  的迭代区间分别为  $[0,1]$ 、 $[0,10]$ <sup>[27]</sup>,SVR 中选择 epsilon-SVR 以及 RBF 分别作为向量机类型和核函数,ELMR 中使用 sigmoid 函数作为激活函数。机器学习回归模型参数及精度见表 6。与 MLR 模型相比,ELMR 的建模精度与验证精度分别提高了 0.005 和 0.025,SVR 建模精度提升了

0.021,验证精度略有下降,可见相较于经典的多元线性回归,ELMR 模型能全面提升模型质量,而 SVR 模型却在泛化性方面表现欠佳,存在过拟合现象,故此次研究中估测玉米冠层叶片 Anth 值的最优模型为抽雄期连续统去除光谱的 ELMR 模型,该模型建模与验证  $R^2$  分别为 0.847、0.895,相应的  $nRMSE$  为 6.44%和 7.21%,均处于较低水平。

表 6 机器学习回归模型参数及精度

Table 6 The parameters and accuracy of machine learning regression models

模型类型	生育期	模型参数	建模精度		验证精度	
			$R^2$	$nRMSE$ (%)	$R^2$	$nRMSE$ (%)
SSA-SVR	抽雄期	$P=30;I=1\ 000;C=0.313\ 3$	0.863	6.16	0.849	8.19
SSA-ELMR	抽雄期	$P=30;I=1\ 000;F=0.102\ 7$	0.847	6.44	0.895	7.21

SSA-SVR:由麻雀搜索算法(SSA)优化的支持向量回归(SVR)模型;SSA-ELMR:由麻雀搜索算法(SSA)优化的极限学习机回归(ELMR)模型; $P$ :麻雀种群数量; $I$ :最大迭代次数; $F$ :自适应度参数; $C$ :惩罚系数; $R^2$ :决定系数; $nRMSE$ :归一化均方根误差。

### 3 讨论

不同生育期的玉米生长特性不同<sup>[28]</sup>,该特点在

高光谱曲线上有所体现,不同生育期玉米冠层叶片光谱特征的差异性决定了分别对不同生育期构建 Anth 值反演模型的必要性,在玉米生长发育的早期

(拔节期),叶片对可见光波段的反射较多,对近红外波段的反射较少,随着生育期向抽雄期靠近,近红外波段反射率逐渐提高,可见光波段反射率逐渐降低,于抽雄期近红外波段反射率达到最大值,该时期玉米叶片对光能的利用效率趋近峰值,乳熟期叶片细胞老化,但色素含量较丰富,故叶片光谱特征在近红外区域逐步向拔节期靠近,该结果与袁媛等<sup>[29]</sup>的研究结果一致。光谱数据采集时由于传感器与叶片存在一定距离,所以该数据不仅包含了玉米冠层叶片的信息,还包含了小部分土壤背景和茎秆信息。故据此建立的各类模型是否适用于玉米纯冠层叶片条件下高光谱数据还有待验证,后期的研究应致力于对比纯叶片和非纯叶片条件下玉米叶片 Anth 值敏感波段,提取主要敏感波段的同时去除噪声波段,简化模型的输入参数和提升其精度,为将来的天空一体化模型建立及推广提供新思路。

不同生育期玉米叶片原始光谱与 Anth 值极显著相关( $P < 0.01$ )的波段主要集中在可见光波段,且均为正相关,该结果与落莉莉等<sup>[30]</sup>的研究结果不一致,原因在于花青素对玉米叶片的调节作用不同于叶绿素,有学者<sup>[31]</sup>发现,当玉米叶片患有红叶病时,花青素含量会异常偏高,而叶绿素含量偏高却表示玉米的生长状况较好,故未来的研究应着力于花青素对玉米叶片光谱响应贡献的研究;光谱变换是一种消除噪声、突出特征的数据处理方式,本研究的一阶导数变换和连续统去除变换均提升了特征波段与 Anth 值的相关性,但光谱变换种类较少,可将更多的变换方式应用到玉米叶片花青素含量反演研究,可进一步扩展到基于变换光谱的反演植被指数构建研究。

本研究基于 SPA 选择的多因素模型建模参数在 2~27 个,与前人的筛选个数不同<sup>[32-33]</sup>,一方面是因为反演的对象不同,另一方面则是由于数据采集的环境不同,基于 SPA 降维算法在模型可解释性方面的优越性,也有学者<sup>[34]</sup>提出改进型 SPA 算法,可以有效提升光谱定量模型的稳定性。构建的多个不同生育期玉米叶片 Anth 值反演模型中,多因素模型优于单因素模型,机器学习回归模型优于传统多元线性回归模型,该结论与前人<sup>[22,35]</sup>的研究结果一致。多因素模型建模参数较多,提供的信息较为全面,故模型精度较高,机器学习算法是由多个弱模型组成的强模型,所以模型质量要好于传统多元线性

回归模型。但是机器学习回归模型的建立往往需要较大的计算成本,因此,近年来灰狼算法、蚁群算法以及狮群算法等群体智能算法<sup>[36]</sup>被逐渐应用到机器学习算法的参数寻优中,本研究结合麻雀搜索算法与机器学习算法,构建了精度较高的玉米叶片 Anth 值估测模型。

本研究建立的最优模型为抽雄期基于一阶导数光谱 SSA-ELMR 模型,抽雄期作为玉米快速发育的时期,该时期玉米的生长情况直接关系到后期玉米籽粒的质量好坏,通过该 SSA-ELMR 模型可快速监测到抽雄期玉米叶片 Anth 值是否超过临界值<sup>[31]</sup>。但是受仪器、地区以及样本数的限制,本研究所建立的最优模型是否可以应用到其他年份以及其他地区的玉米叶片花青素含量监测还有待检验,此外,未来应进行多年份、多地区的试验以进一步改进模型的鲁棒性和泛化性,且模型的参数要尽量向高光谱遥感卫星参数靠拢,不断扩大模型适用性,保证在各个尺度上均能监测玉米叶片花青素含量。

## 4 结 论

本研究以玉米为研究对象,基于不同生育期叶片的原始光谱及其变换光谱构建玉米叶片 Anth 值反演模型,得到如下结论:(1)不同生育期的玉米叶片各类型光谱特征总体一致、局部不同;原始光谱中可见光反射率较低而近红外反射率较高,二者通过“红边”过渡。(2)一阶导数变换能放大原始光谱中突变的部位,连续统去除变换能突出原始光谱中反射峰的两端;变换光谱特征波段与 Anth 值的相关性高于原始光谱。(3)多因素模型优于单因素模型,且机器学习算法可进一步提升多因素模型的精度。所有模型中,抽雄期基于一阶导数光谱的 SSA-ELMR 模型为最优模型,其建模与验证  $R^2$  分别为 0.847、0.895,相应的  $nRMSE$  为 6.44% 和 7.21%。

## 参考文献:

- [1] CLOSE D C, BEADLE C L. The ecophysiology of foliar anthocyanin[J]. The Botanical Review, 2003, 69(2): 149-161.
- [2] LINDA-CHALKER C. Environmental significance of anthocyanins in plant stress responses[J]. Photochemistry and Photobiology, 1999, 70(1): 1-9.
- [3] GOULD K S, MCKELVIE J, MARKHAM K R. Do anthocyanins function as antioxidants in leaves? Imaging of  $H_2O_2$  in red and green leaves after mechanical injury[J]. Plant, 2010, 25(10):



- 1261-1269.
- [4] 陈敏. 食品化学[M]. 北京:中国林业出版社,2008.
- [5] 吴金滢,胡迪,孙新. 原花青素抗衰老作用的研究进展[J]. 吉林医药学院学报,2022,43(6):452-454.
- [6] GITELSON A A, CHIVKUNOVA O B, MERZLYAK M N. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves[J]. American Journal of Botany, 2009, 96(10):1861-1868.
- [7] 刘芳芳,向成钢,张宏,等. 高效液相色谱法与pH示差法测定紫萼豌豆花青素含量的比较[J]. 食品科技,2022,47(1):298-303.
- [8] STEELE M R, GITELSON A A, RUNDQUIST D C, et al. Nondestructive estimation of anthocyanin content in grapevine leaves[J]. American Journal of Enology and Viticulture, 2009, 60(1):87-92.
- [9] GITELSON A, SOLOVCHENKO A. Generic algorithms for estimating foliar pigment content[J]. Geophysical Research Letters, 2017, 44(18):9293-9298.
- [10] CHOI J H, PARK S H, JUNG D H, et al. Hyperspectral imaging-based multiple predicting models for functional component contents in Brassica juncea[J]. Agriculture-Basel, 2022, 12:1515.
- [11] 刘秀英,申健,常庆瑞,等. 基于可见/近红外光谱的牡丹叶片花青素含量预测[J]. 农业机械学报,2015,46(9):319-324.
- [12] 田明璐,班松涛,常庆瑞,等. 高光谱影像的苹果花叶病叶片花青素定量反演[J]. 光谱学与光谱分析, 2017, 37(10):3187-3192.
- [13] LEI W, WANG P X, LI L, et al. Developing an integrated indicator for monitoring maize growth condition using remotely sensed vegetation temperature condition index and leaf area index[J]. Computers and Electronics in Agriculture, 2018, 152:340-349.
- [14] 王伟东,常庆瑞,王玉娜. 冬小麦叶片花青素相对含量高光谱监测[J]. 麦类作物学报,2020,40(6):754-761.
- [15] 杨可明,郭达志,陈云浩. 高光谱植被遥感数据光谱特征分析[J]. 计算机工程与应用,2006,42(31):213-215.
- [16] 张贤龙,张飞,张海威,等. 基于光谱变换的高光谱指数土壤盐分反演模型优选[J]. 农业工程学报,2018,34(1):110-117.
- [17] 李粉玲,常庆瑞. 基于连续统去除法的冬小麦叶片全氮含量估算[J]. 农业机械学报,2017,48(7):174-179.
- [18] 陈澜,常庆瑞,高一帆,等. 猕猴桃叶片叶绿素含量高光谱估算模型研究[J]. 西北农林科技大学学报(自然科学版), 2020, 48(6):79-89, 98.
- [19] 张佑铭. 基于多源遥感数据的猕猴桃园 LAI 反演研究[D]. 杨凌:西北农林科技大学,2022.
- [20] SOARES S F C, GOMES A A, FILHO A R G, et al. The successive projections algorithm[J]. Trac-Trends in Analytical Chemistry, 2013, 42:84-98.
- [21] VERRELST J, MUNOZ J, ALONSO L, et al. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3[J]. Remote Sensing of Environment, 2012, 118:127-139.
- [22] LI W, CHANG Q R, LI F L, et al. Effects of growth stage development on paddy rice leaf area index prediction models[J]. Remote Sensing, 2019, 11(3):361.
- [23] LEE H, WANG J F, LEBLON B. Using linear regression, random forests, and support vector machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn[J]. Remote Sensing, 2020, 12(13):2071.
- [24] 由明明,常庆瑞,田明璐,等. 基于随机森林回归的油菜叶片 SPAD 值遥感估算[J]. 干旱地区农业研究, 2019, 37(1):74-81.
- [25] 余蛟洋,常庆瑞,由明明,等. 基于高光谱和 BP 神经网络模型苹果叶片 SPAD 值遥感估算[J]. 西北林学院学报, 2018, 33(2):156-165.
- [26] HUANG G, HUANG G B, SONG S, et al. Trends in extreme learning machines: a review[J]. Neural Networks, 2015, 61:6132-6148.
- [27] 薛建凯. 一种新型的群智能优化技术的研究与应用[D]. 上海:东华大学,2020.
- [28] 刘俊,孟庆岩,葛小三,等. 基于 BP 神经网络的夏玉米多生育期叶面积指数反演研究[J]. 遥感技术与应用, 2020, 35(1):174-184.
- [29] 袁媛,瑚波,武兴厚,等. 基于植被指数的夏玉米不同生育期叶绿素含量遥感估算[J]. 中国农学通报, 2015, 31(15):254-259.
- [30] 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.
- [31] LUO L L, CHANG Q R, WANG Q, et al. Identification and severity monitoring of maize dwarf mosaic virus infection based on hyperspectral measurements[J]. Remote Sensing, 2021, 13(22):4560.
- [32] 王玉娜,李粉玲,王伟东,等. 基于连续投影算法和光谱变换的冬小麦生物量高光谱遥感估算[J]. 麦类作物学报, 2020, 40(11):1389-1398.
- [33] 白丽敏,李粉玲,常庆瑞,等. 结合 SPA 和 PLS 法提高冬小麦冠层全氮高光谱估算的精确度[J]. 植物营养与肥料学报, 2018, 24(5):1178-1184.
- [34] 郝勇,孙旭东,王豪. 基于改进连续投影算法的光谱定量模型优化[J]. 江苏大学学报(自然科学版), 2013, 34(1):49-53.
- [35] 张敏,郭涛,刘轲,等. 基于高光谱植被指数的大米 LAI 遥感估算[J/OL]. 西南农业学报, 2022, 35(11):1-10.
- [36] 高岳林,杨钦文,王晓峰,等. 新型群体智能优化算法综述[J]. 郑州大学学报(工学版), 2022, 43(3):21-30.

(责任编辑:成纾寒)