

裴浩杰,冯海宽,李长春,等. 基于多元线性回归和随机森林的苹果叶绿素含量高光谱估测方法比较[J]. 江苏农业科学,2018,46(17):224-230.  
doi:10.15889/j.issn.1002-1302.2018.17.060

# 基于多元线性回归和随机森林的苹果叶绿素含量高光谱估测方法比较

裴浩杰<sup>1,2,3,4</sup>, 冯海宽<sup>2,3,4</sup>, 李长春<sup>1</sup>, 李振海<sup>2,3,4</sup>, 杨贵军<sup>2,3,4</sup>, 王衍安<sup>5</sup>, 郭建华<sup>2,3,4</sup>

(1. 河南理工大学测绘与国土信息工程学院,河南焦作 454000; 2. 国家农业信息化工程技术研究中心,北京 100097; 3. 农业部农业信息技术重点实验室,北京 100097; 4. 北京市农业物联网工程技术研究中心,北京 100097; 5. 山东农业大学生命科学学院,山东泰安 271018)

**摘要:**叶绿素含量是果树营养胁迫和光合作用等生理状态的良好指示剂。为快速准确地估测苹果叶片叶绿素含量,利用采集的苹果叶片光谱和叶片叶绿素含量数据,通过分析原始光谱与叶绿素含量的相关性,筛选出 554、708、995 nm 3 个最佳敏感波段,构建基于原始光谱敏感波段的多元线性回归(multivariable linear regression,简称 MLR)模型和随机森林(random forest,简称 RF)模型,用于叶绿素含量估测;使用相关系数绝对值(absolute value of correlation coefficient,简称|*r*|)和 RF 的袋外数据(out of bag,简称 OOB)重要性分别对植被指数与叶片叶绿素含量的关联性进行分析,筛选植被指数,然后使用 MLR 和 RF 算法构建模型,依次增加植被指数的输入数,筛选出 10 个植被指数的 MLR 最优模型和 5 个植被指数的 RF 最优模型;比较上述 4 个模型的估测精度。基于原始光谱的 MLR 模型和 RF 模型以及基于植被指数的 MLR 最优模型和 RF 最优模型建模的  $R^2$  分别为 0.578、0.527、0.602、0.609, RMSE 分别为 8.240、8.728、8.004、7.930  $\mu\text{g}/\text{cm}^2$ , 4 个模型建模精度相近。在模型验证方面,4 个模型的  $R^2$  分别为 0.899、0.411、0.854、0.843, RMSE 分别为 8.297、14.455、11.242、11.034  $\mu\text{g}/\text{cm}^2$ 。基于原始光谱的 MLR 模型的叶绿素含量估测精度高于其他 3 个模型,能够精确地估测苹果叶片叶绿素含量。另外,基于植被指数的 MLR 模型和 RF 模型对苹果叶片叶绿素含量估测也具有一定的应用潜力。

**关键词:**苹果叶片;高光谱;叶绿素;MLR;RF

**中图分类号:** S127 **文献标志码:** A **文章编号:** 1002-1302(2018)17-0224-07

叶绿素含量是果树生长过程中重要的生化参数,对于果树光合能力、发育阶段以及营养状况有指示作用,其变化还是果树是否受环境胁迫或病害的指示剂<sup>[1-2]</sup>。运用高光谱提取果树叶片叶绿素含量信息,对于监测果树的生长状况和营养诊断具有十分重要的意义。近年来,国内外对果树遥感监测取得了一定的进展。Mark 等建立了红边光谱与葡萄叶绿素含量的模型,实现估测模型均方根误差(RMSE) < 30  $\text{mg}/\text{m}^2$ <sup>[3]</sup>。张瑶等使用苹果叶片原始光谱和小波滤波后的叶片光谱,运用支持向量机法和偏最小二乘法分别建立叶绿素估测模型,实现了叶绿素含量的精确估测<sup>[4]</sup>。李敏夏分别对苹果叶片原始光谱和一阶微分光谱与苹果叶片叶绿素含量的相关性进行分析,分别以光谱位置、植被指数和光谱面积作为变量建立叶片叶绿素回归模型,结果显示,蓝边位置构建的指数模型具有较高的估测精度<sup>[5]</sup>。李萍利用逐步回归分析方法,分别将香梨叶片全氮含量与原始光谱敏感波段和光谱

一阶微分敏感波段结合,建立叶片全氮含量估测模型,最终确定将光谱一阶微分参与构建的模型作为叶片全氮含量估测模型<sup>[6]</sup>。张蕾蕾使用苹果叶片原始光谱、一阶微分光谱,分别利用主成分分析方法和逐步回归分析方法建立叶绿素含量估测模型<sup>[7]</sup>。研究人员针对果树叶片叶绿素含量的估算进行了多种尝试,而在果树遥感监测方面,随机森林(random forest,简称 RF)方法的运用较少,而 RF 作为一种机器学习方法已开始应用在农业遥感领域。王丽爱等运用 RF 和人工神经网络,使用环境卫星数据构建植被指数,对冬小麦叶面积指数进行反演<sup>[8]</sup>。李粉玲等利用 RF 构建模型,运用高分一号卫星数据建立植被指数,对冬小麦土壤、作物分析仪器开发值(soil and plant analyzer development,简称 SPAD)进行估测<sup>[9]</sup>。王丽爱等运用 RF 回归算法构建小麦拔节期、孕穗期和开花期 SPAD 反演算法模型<sup>[10]</sup>。韩兆迎等利用支持向量机和 RF 构建苹果树冠叶面积指数的高光谱估测模型<sup>[11]</sup>。岳继博等通过将相关系数、灰色关联、袋外数据的重要性分别与 RF 结合构建模型,证实了 RF 对冬小麦生物量遥感估测的潜力<sup>[12]</sup>。

分析国内外的研究成果发现,大多数研究集中在对单一生长时期的果树进行营养诊断,使得测量结果对于整个果树生长周期的营养状况缺乏评价标准和指导作用<sup>[4]</sup>。因此,本研究连续 2 年针对苹果全生育期的叶片叶绿素含量进行估测。此外,由于目前多数研究只使用叶绿素相关的敏感波段或者只使用植被指数进行建模,并没有比较用敏感波段和植被指数进行建模的不同。在建模分析方法上,以上研究分别

收稿日期:2017-12-10

基金项目:国家自然科学基金(编号:41601346、41471285、41301475);国家高技术研究发展计划(编号:2011AA100703)。

作者简介:裴浩杰(1991—),男,河南郑州人,硕士研究生,主要从事农业定量遥感研究。E-mail:xmljphj@163.com。

通信作者:冯海宽,硕士,助理研究员,主要从事农业定量遥感研究, E-mail:fenghaikuan123@163.com;李长春,博士,副教授,主要从事农业遥感长势监测与评估研究工作, E-mail:lichangchun610@126.com。

较多地使用线性回归和非线性回归等建模方法,较少同时考虑线性建模和非线性建模方法。

综上所述,利用叶片光谱估测苹果叶片叶绿素含量的相关研究并不多,并且较少有研究将 RF 模型与光谱敏感波段结合起来构建模型用于苹果叶绿素含量估测。本研究整合苹果多生育期的数据,利用多元线性回归(multivariable linear regression,简称 MLR)和 RF,分别结合原始光谱敏感波段和植被指数,拟构建基于原始光谱的 MLR 模型和 RF 模型,以及基于植被指数的 MLR 模型和 RF 模型,分别比较 4 个模型的叶绿素含量估测精度,旨在将 RF 模型应用于苹果叶片叶绿素含量的估测,并筛选出估测苹果叶片叶绿素含量的最佳

模型,以期为实现苹果叶片叶绿素含量快速估测提供一种指导方法。

## 1 材料与方法

### 1.1 研究区概况

本试验于 2012、2013 年开展,试验区位于山东省肥城市潮泉镇下寨村,研究区区位见图 1。地理位置为  $116^{\circ}50'22''\text{E}$ ,  $36^{\circ}14'01''\text{N}$ 。试验区果园于 2001 年建园,苹果树的主要栽种品种为富士、嘎啦,行距为 5 m,株距为 3 m,树高约为 3 m,果树树干高约为 0.5 m,树形为纺锤形。

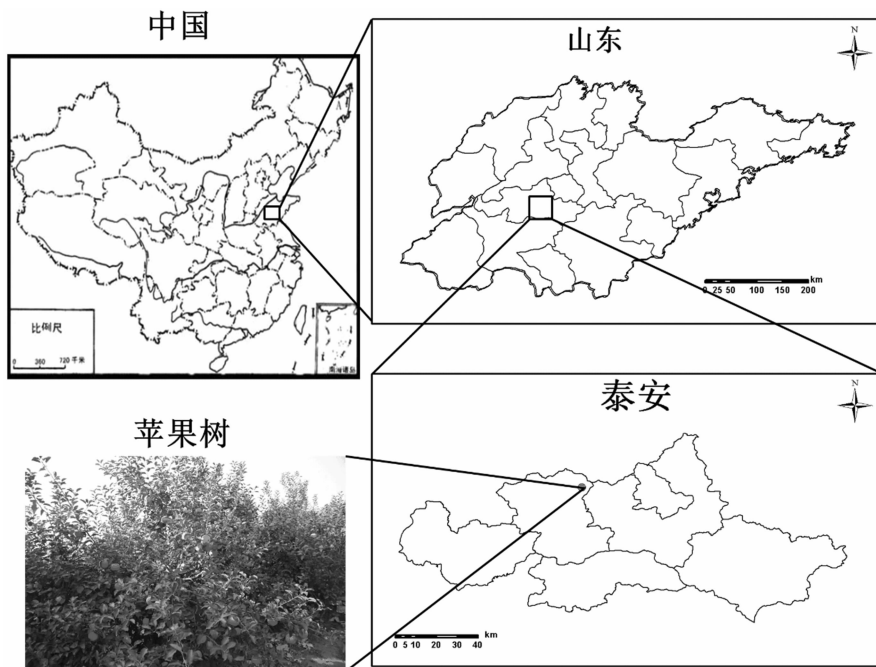


图1 研究区区位

### 1.2 叶片取样及光谱测定

叶片取样时,在树冠东、西、南、北 4 个方向、树冠外围中部、当年延长枝中部叶(带叶柄)各取 1 张叶片用于光谱测量,放入自封袋中。1 株树作为 1 个样本,2012 年共采集苹果树 299 株(其中富士 220 株,嘎啦 79 株),2013 年共采集苹果树 180 株(其中富士 135 株,嘎啦 45 株)。

叶片光谱采用美国 ASD(analytical spectral devices)公司的 Fieldspec FR2500 型野外光谱辐射谱仪和叶片夹(ASD leaf clip)进行测定。光谱仪相关参数如下:光谱仪波段范围为  $350 \sim 2\,500\text{ nm}$ ,间隔为  $1\text{ nm}$ ,其中光谱分辨率在  $350 \sim 1\,000\text{ nm}$  内的间隔为  $3\text{ nm}$ ,在  $1\,000 \sim 2\,500\text{ nm}$  内的间隔为  $10\text{ nm}$ ,在  $350 \sim 1\,000\text{ nm}$  内的采样间隔为  $1.4\text{ nm}$ ,在  $1\,000 \sim 2\,500\text{ nm}$  内的采样间隔为  $2\text{ nm}$ 。每张叶片先用黑板测定 4 个不同位置(叶脉两侧各 2 个,测试时注意避开叶脉,基本覆盖整张叶片),取 4 个反射率平均值作为该叶片的反射率。测量前用叶片夹自带的标准白板进行校正。

### 1.3 苹果叶片叶绿素含量的测定

在实验室中,用直径为  $0.6\text{ cm}$  的打孔器对测量光谱后的 4 张苹果叶片进行打孔处理,每张叶片避开叶脉打 4 个孔,与光谱测量的位置对应,基本覆盖整张叶片,样品质量控制在

$0.2\text{ g}$  左右。然后将样品放入 95% 无水乙醇溶液中,并在黑暗环境下静置  $24 \sim 48\text{ h}$ ,直到叶片变成白绿色,最后采用紫外分光光度计测定苹果叶片的叶绿素含量( $\mu\text{g}/\text{cm}^2$ )<sup>[13]</sup>。本研究中使用的数据采集时间见表 1,经测定得到的叶绿素含量数据统计特征见表 2。本研究利用 2012 年获取的苹果叶片叶绿素含量和叶片光谱反射率数据( $n=299$ )构建模型,使用 2013 年采集的数据( $n=180$ )验证 MLR 和 RF 模型的精度。

### 1.4 方法

1.4.1 植被指数选取 根据前人的研究结果,本研究选取与叶绿素关联性较好的 25 个植被指数,作为估算叶绿素含量的待选变量,详见表 3。

1.4.2 相关系数绝对值 相关系数(correlation coefficient,简称  $r$ )是用来描述 2 组数据之间关系密切程度的统计指标。 $|r|$  一般为小于 1 的数, $|r|$  越大,表示 2 组数据之间的关系越密切。在本研究中,使用  $|r|$  筛选与苹果叶片关系密切的植被指数,并将这些植被指数作为 MLR 模型的输入变量。

1.4.3 随机森林 RF 是由美国科学家 Breiman 于 2001 年提出的一种机器学习算法<sup>[35]</sup>。RF 基于 bootstrap 抽样方法,从原始样本中有放回地抽取多个样本,对每个 bootstrap 样本使用决策树建模,然后组合多株决策树进行预测,以投票的方

表 1 本试验测定项目的详细信息

时间	富士生育期	嘎啦生育期	富士样本量 (个)	嘎啦样本量 (个)	叶绿素含量 测定情况	光谱反射率 测定情况
2012-05-10	春梢旺长期	春梢旺长期	44	16	√	√
2012-07-03	春梢停长期	春梢停长期	44	15	√	√
2012-08-10	果实膨大期	果实成熟期	44	16	√	√
2012-09-20	果实膨大期	叶变色期	44	16	√	√
2012-10-18	果实成熟期	叶变色期	44	16	√	√
2012 年总计			220	79		
2013-04-20	开花盛期	开花盛期	27	9	√	√
2013-05-25	春梢旺长期	春梢旺长期	27	9	√	√
2013-07-18	秋梢旺长期	秋梢旺长期	27	9	√	√
2013-08-30	果实膨大期	果实成熟期	27	9	√	√
2013-10-25	果实成熟期	叶变色期	27	9	√	√
2013 年总计			135	45		
总和			355	124		

注:“√”表示为数据已测定。

表 2 苹果叶片的叶绿素含量统计特征

样本	样本数 (个)	叶绿素含量(μg/cm <sup>2</sup> )			
		最大值	最小值	均值	标准差
建模集	299	112.562	51.291	76.529	12.707
验证集	180	103.518	35.677	74.367	18.133
总样本	479	112.562	35.677	75.717	14.996

式来决定最后的预测结果。

Bagging 是 RF 理论中的一部分。假设样本集的样本容量为  $N$ , 每次有放回抽取的 Bootstrap 样本数量为  $n$ , 这一部分未被抽中的样本称为袋外数据<sup>[36]</sup>。这些未被选中的袋外数据可用于估计 RF 的单株决策树分类强度, 分类强度越大, RF 的泛化误差界就越小, 分类的准确性会越高, 预测结果就会越准确<sup>[37]</sup>。在本研究中, 使用 RF 中的袋外数据估计方法对现有植被指数与苹果叶片叶绿素含量进行重要性分析并进行排序。依次将靠前的植被指数作为决策树组合成 RF 进行建模和估测。经反复尝试, 决策树数量设置为 1 000 为最佳。

1.4.4 多元线性回归 多元线性回归是使用多个自变量对因变量进行解释的回归方法<sup>[37]</sup>。该回归方法要求自变量与因变量之间要有较显著的相关性。

1.5 统计分析

本研究估测和验证模型精度均选择决定系数( $R^2$ )、均方根误差(RMSE)作为评价标准。一般而言, RMSE 越小,  $R^2$  越大, 表明模型的精度越高。

2 结果与分析

2.1 叶片叶绿素含量与光谱的相关性分析

如图 2 所示, 在 525~581 nm 波段与 693~735 nm 波段范围内, 苹果叶片叶绿素含量与原始光谱呈显著负相关; 在 745~1 350 nm 波段范围内, 叶片叶绿素含量与原始光谱呈显著正相关。分别在绿光、红光、近红外波段选取与叶片叶绿素含量相关性最好的波段, 波长分别为 554、708、995 nm, 相关系数分别为 -0.46、0.66、0.61, 以此 3 个波段的反射率作为 MLR 模型、RF 模型的输入变量。

2.2 基于原始光谱的 MLR 模型和 RF 模型

利用原始光谱敏感波段 554、708、995 nm 的反射率

$R_{554\text{ nm}}$ 、 $R_{708\text{ nm}}$ 、 $R_{995\text{ nm}}$  构建 MLR 模型和 RF 模型, 使用 2013 年光谱和叶绿素含量数据验证 MLR 和 RF 模型精度。建模  $R^2$  分别为 0.578、0.527, RMSE 分别为 8.240、8.728 μg/cm<sup>2</sup>, 详见表 4。

2.3 植被指数的筛选

使用 lri 和 RF 的袋外数据(out of bag, 简称 OOB)重要性估计方法分别对植被指数与苹果叶片叶绿素含量的关联性进行计算。考虑到模型的可操作性和简单性, 本研究只考虑 2 种方法各自排序后的前 10 个植被指数, 详见表 5。通过 lri 和 OOB 估计 2 种方法对植被指数进行排序, 比较发现, 2 种方法的前 10 个植被指数中有 9 个相同, 分别为 NDVI2、MMTCI、MTCI、VOG2、DCNI I、VOG、SR1、MSR2 和 MNDVI2, 另外 1 个不同的植被指数, OOB 重要性估计的为 ARI, lri 为 DI, 说明 lri 与 OOB 重要性估计在植被指数与叶片叶绿素含量关联性分析方面的一致性很高, OOB 重要性估计的可信度很高。

2.4 植被指数的 MLR 最优回归模型

根据相关性分析结果, 选取前 10 个植被指数, 依次增加植被指数的输入数量作为自变量, 分别建立叶绿素含量估测模型。从表 6 可以看出, 随着植被指数数量的增加,  $R^2$  不断增大, RMSE 不断减小, 当 10 个植被指数作为 MLR 回归模型的自变量输入时,  $R^2$  达到最大值, RMSE 同时为最小值, 说明此时构建的模型在仅选择前 10 个植被指数的情况下为最优模型。

2.5 随机森林最优模型

根据 OOB 重要性估计选取前 10 个植被指数, 依次增加植被指数的输入数量, 分别建立叶绿素含量估测模型。如表 7 所示, 当输入植被指数数量从 1 个增加到 5 个时,  $R^2$  的总体趋势是增大的, RMSE 则逐渐减小。当植被指数数量为 5 个时,  $R^2$  达到最大值, 为 0.609, RMSE 最小, 为 7.930 μg/cm<sup>2</sup>。当植被指数数量为 6~10 个时,  $R^2$  从 0.597 上升到 0.606, RMSE 从 8.067 μg/cm<sup>2</sup> 下降到 7.966 μg/cm<sup>2</sup>, 但是  $R^2$  均小于 5 个植被指数的  $R^2$ , RMSE 也均大于 5 个植被指数的 RMSE。

因此可见, 在选择前 10 个植被指数的情况下, 选取前 5 个植被指数构建的 RF 模型为最优模型。

从表 6 和表 7 可以看出, 将 RF 模型的 5 个植被指数作为

表 3 叶绿素相关植被指数

植被指数	计算公式	参考文献
归一化叶绿素指数(normalized pigment chlorophyll index,简称 NPCI)	$(R_{680}-R_{430})/(R_{680}+R_{430})$	[14]
简单比值色素指数(simple ratio pigment index,简称 SRPI)	$R_{430}/R_{680}$	[15]
改进的叶绿素吸收反射率指数(modified chlorophyll absorption reflectance index,简称 MCARI)	$[(R_{700}-R_{670})-0.2(R_{700}-R_{550})](R_{700}/R_{670})$	[16]
改进的叶绿素吸收反射率指数 2(简称 MCARI2)	$1.2[2.5(R_{800}-R_{670})-1.3(R_{800}-R_{500})]$	[17]
变换的叶绿素吸收比值指数(transformed chlorophyll absorption ratio index,简称 TCARI)	$3[(R_{700}-R_{670})-0.2(R_{700}-R_{550})(R_{700}/R_{670})]$	[18]
MERIS 地面叶绿素指数(MERIS terrestrial chlorophyll index,简称 MTCI)	$(R_{754}-R_{709})/(R_{709}-R_{681})$	[19]
改进的简单植被指数 2(modified simple ratio 2,简称 MSR2)	$(R_{750}-R_{445})/(R_{705}-R_{445})$	[20]
改进的归一化差分植被指数 2(modified normalized difference vegetation index 2,简称 MNDVI2)	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$	[20]
归一化植被指数 2(normalized difference vegetation index 2,简称 NDVI2)	$(R_{750}-R_{705})/(R_{750}+R_{705})$	[21]
花青素反射率指数(anthocyanin reflectance index,简称 ARI)	$(1/R_{550})-(1/R_{700})$	[22]
复归一化植被指数(renormalized difference vegetation index,简称 RDVI)	$R_{800}-R_{670}/\sqrt{R_{800}+R_{670}}$	[23]
改良土壤调整指数(improved soil adjusted vegetation index,简称 MSAVI)	$0.5[2R_{800}+1-\sqrt{(2R_{800}+1)^2-8(R_{800}-R_{670})}]$	[24]
红边植被指数(red-edge position,简称 REP)	$700+40\frac{[(R_{670}+R_{780})/2]-R_{770}}{R_{740}-R_{700}}$	[25]
光谱多边形植被指数(spectral polygon vegetation index,简称 SPVI)	$0.4[3.7(R_{800}-R_{670})-1.2(R_{530}-R_{670})]$	[26]
简单比值植被指数 1(simple ratio,简称 SRI)	$R_{750}/R_{700}$	[27]
Vogelmann 指数(Vogelmann indices,简称 VOG)	$R_{740}/R_{720}$	[28]
Vogelmann 指数 2(简称 VOG2)	$(R_{734}-R_{747})/(R_{715}+R_{726})$	[29]
改进的叶绿素吸收积分指数(modified chlorophyll absorptions integral,简称 MCAI)	$A-\int_{R_{552}}^{R_{752}} f$	[30]
导数指数(derivative index,简称 DI)	$D_{730}/D_{706}$	[31]
叶绿素吸收比值指数(chlorophyll absorption atio index,简称 CARI)	$CARI=\frac{a\times R_{670}+R_{670}+b R_{700}}{\sqrt{a^2+1}\times R_{670}},a=\frac{R_{700}-R_{550}}{150},b=R_{550}\times a$	[32]
双峰冠层氮指数(double peak canopy nitrogen index,简称 DCNI)	$(R_{720}-R_{700})/(R_{700}-R_{670})/(R_{720}-R_{670}+0.03)$	[33]
双峰冠层氮指数 I(简称 DCNI I)	$[(R_{750}-R_{670}+0.09)(R_{750}-R_{700})]/(R_{700}-R_{670})$	[34]
改进的 MERIS 地面叶绿素指数(modified MERIS terrestrial chlorophyll index,简称 MMTCI)	$[(R_{750}-R_{680}+0.03)(R_{750}-R_{710})]/(R_{710}-R_{680})$	[34]
组合指数Ⅲ(combined index Ⅲ)	$[(R_{800}-R_{445})/(R_{800}-R_{680})]/(R_{800}/R_{670})$	[34]
组合指数Ⅳ(combined index Ⅳ)	$[(R_{550}-R_{450})/(R_{550}+R_{450})]/[(R_{800}-R_{670})/(R_{800}+R_{670})]$	[34]

注:R 表示反射率,下标代表具体波段(nm)。

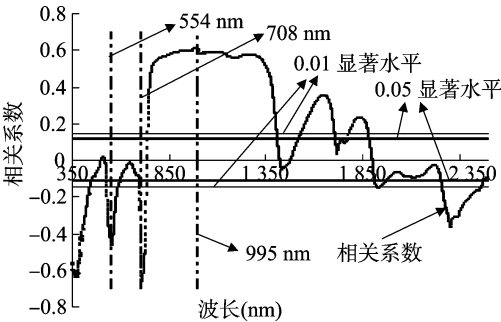


图2 光谱与叶绿素含量的相关性

输入变量时精度已趋于稳定水平,而且此时的  $R^2$  均高于植被指数数量为 1~10 个的 MLR 模型的  $R^2$ ,同时  $RMSE$  低于 1~10 个植被指数 MLR 模型的  $RMSE$ ,说明 5 个植被指数的 RF 建模精度高于 10 个植被指数的 MLR 模型。

2.6 叶绿素含量估测与验证

MLR 和 RF 模型的叶绿素含量验证精度结果见图 3。可以看出,基于原始光谱的 MLR 模型、RF 模型的验证  $r^2$  分别为 0.899、0.411, $RMSE$  分别为 8.297、14.455  $\mu\text{g}/\text{cm}^2$ ,基于植被指数的 MLR 最优模型、RF 最优模型的验证  $r^2$  分别为 0.854、0.843, $RMSE$  分别为 11.242、11.034  $\mu\text{g}/\text{cm}^2$ ,说明基于原始光谱的 MLR 模型估测精度高于其他 3 个模型。基于

表 4 基于原始光谱的 MLR 模型和 RF 模型精度

模型	变量	回归方程	$RMSE(\mu\text{g}/\text{cm}^2)$	$R^2$
MLR	$R_{554}, R_{708}, R_{995}$	$CHL=20.176+138.149\times R_{554}-572.728\times R_{708}+241.580\times R_{995}$	8.240	0.578
RF	$R_{554}, R_{708}, R_{995}$		8.728	0.527

表 5 |r| 和 OOB 计算植被指数与叶绿素含量关联性顺序

植被指数	r	r 排序	植被指数	袋外数据 重要性值	袋外数据 排序
NDVI2	0.733 *	1	NDVI2	6 494.599	1
MMTCI	0.732 *	2	MMTCI	5 612.955	2
VOG2	0.727 *	3	MTCI	5 191.305	3
VOG	0.718 *	4	VOG2	2 805.643	4
MTCI	0.712 *	5	ARI	2 773.980	5
DCNI I	0.710 *	6	DCNI I	2 277.633	6
MNDVI2	0.699 *	7	VOG	2 058.152	7
DI	0.694 *	8	SR 1	1 611.043	8
MSR2	0.693 *	9	MSR2	1 587.592	9
SR1	0.680 *	10	MNDVI2	1 562.748	10

注：“\*”表示在 0.01 水平显著相关。

植被指数的 MLR 模型、RF 模型估测都具有较好的稳定性,2 个模型精度相近,表明可以采用基于植被指数的 RF 最优模型进行苹果叶片叶绿素含量的估测。

由图 3-a 可以看出,基于原始光谱的 MLR 模型的估测效果相对最好。由图 3-b 可以看出,基于原始光谱的 RF 模型的估测效果较差,当实测值在 60 μg/cm<sup>2</sup> 以下时,模型的估测值出现高估现象;当实测值大于 75 μg/cm<sup>2</sup> 时,部分估测值出现低估现象。由图 3-c 可以看出,苹果叶片叶绿素含量部分值出现了低估现象。由图 3-d 可以看出,当叶片叶绿素含量的实测值在 60 μg/cm<sup>2</sup> 以下时(开花期叶绿素含量),RF 模型的估测值出现高估,而且估测值基本不随实测值的增加而改变;当实测值大于 80 μg/cm<sup>2</sup> 时,部分估测值出现低估现象。

3 讨论

本研究分别利用原始光谱和植被指数结合 MLR 模型和

表 6 不同植被指数数量构建的 MLR 模型精度比较

植被指数数量 (个)	MLR 回归方程	RMSE (μg/cm <sup>2</sup> )	R <sup>2</sup>
1	$CHL = -124.738 + 323.653 \times NDVI2$	8.627	0.538
2	$CHL = -56.684 + 167.573 \times NDVI2 + 19.997 \times MMTCI$	8.507	0.550
3	$CHL = -59.637 + 173.966 \times NDVI2 + 22.821 \times MMTCI + 38.888 \times VOG2$	8.505	0.550
4	$CHL = 266.362 + 293.148 \times NDVI2 + 37.238 \times MMTCI - 900.215 \times VOG2 - 322.695 \times VOG$	8.226	0.580
5	$CHL = 242.331 + 343.985 \times NDVI2 + 75.006 \times MMTCI - 937.115 \times VOG2 - 319.778 \times VOG - 21.194 \times MTCI$	8.072	0.595
6	$CHL = 303.89 + 466.533 \times NDVI2 + 82.521 \times MMTCI - 1028.980 \times VOG2 - 419.570 \times VOG - 9.969 \times MTCI - 4.775 \times DCNI1$	8.047	0.598
7	$CHL = 297.886 + 471.769 \times NDVI2 + 85.827 \times MMTCI - 1037.060 \times VOG2 - 432.020 \times VOG - 10.734 \times MTCI - 5.516 \times DCNI1 + 33.860 \times MNDVI2$	8.045	0.598
8	$CHL = 223.994 + 349.047 \times NDVI2 + 96.334 \times MMTCI - 1045.110 \times VOG2 - 334.304 \times VOG - 1.389 \times MTCI - 8.389 \times DCNI1 + 55.556 \times MNDVI2 - 50.834 \times DI$	8.041	0.598
9	$CHL = 229.919 + 361.790 \times NDVI2 + 75.616 \times MMTCI - 1106.870 \times VOG2 - 351.628 \times VOG - 2.829 \times MTCI - 4.235 \times DCNI1 + 85.657 \times MNDVI2 - 26.689 \times DI - 5.346 \times MSR2$	8.030	0.599
10	$CHL = 64.056 + 844.557 \times NDVI2 + 66.343 \times MMTCI - 1044.250 \times VOG2 - 207.144 \times VOG - 4.484 \times MTCI - 3.461 \times DCNI1 - 235.757 \times MNDVI2 - 76.938 \times DI + 3.417 \times MSR2 - 19.139 \times SR1$	8.004	0.602

表 7 不同植被指数数量的 RF 建模精度比较

植被指数数量 (个)	RMSE (μg/cm <sup>2</sup> )	R <sup>2</sup>
1	10.305	0.380
2	9.199	0.483
3	8.752	0.527
4	8.765	0.525
5	7.930	0.609
6	8.067	0.597
7	8.046	0.599
8	8.031	0.600
9	8.008	0.602
10	7.966	0.606

RF 模型构建叶绿素含量估测模型。结果表明,基于原始光谱的 MLR 模型的叶绿素含量估测精度高于其他 3 个模型,可以实现苹果叶绿素含量的精确估测。

关于 RF 模型的应用,基于原始光谱的 RF 模型和基于植被指数的 RF 模型的学习能力都较强,但是基于植被指数的

RF 模型的估测精度较好,基于原始光谱的 RF 模型估测精度较差。主要原因可能是基于原始光谱的 RF 模型用于训练的数据仅有 3 个敏感波段的反射率,模型的训练数据较少,RF 模型适用于较大的数据集,对于小规模数据优势不明显<sup>[38]</sup>。基于植被指数的 RF 模型的估测效果好于基于敏感单波段的 RF 模型,这与王丽爱等得出的结论<sup>[10]</sup>一致,相比于单波段结合 RF 模型进行估测,采用多波段参数与 RF 算法结合的方法提高了估测精度。

本研究利用 |r| 和 OOB 重要性筛选的前 10 个植被指数中,有 9 个共同的植被指数,表明 OOB 重要性估计与 |r| 方法排序存在很好的一致性,同时说明 OOB 重要性估计对变量的排序是可信的,这与 Verikas 等的研究结果不一致,Verikas 等研究表明,OOB 重要性排序并非完全可信<sup>[39]</sup>。

本研究还有一些不足,例如,由于气候原因,2013 年的生育期与 2012 年不完全一致,而且 2012 年第 1 次采集数据的时间是春梢旺长期,2013 年第 1 次采集数据的时间是开花期,而开花期叶绿素含量偏低,导致 2013 年开花期叶绿素含量低于 2012 年建模数据的叶绿素含量,超出建模数据范围,

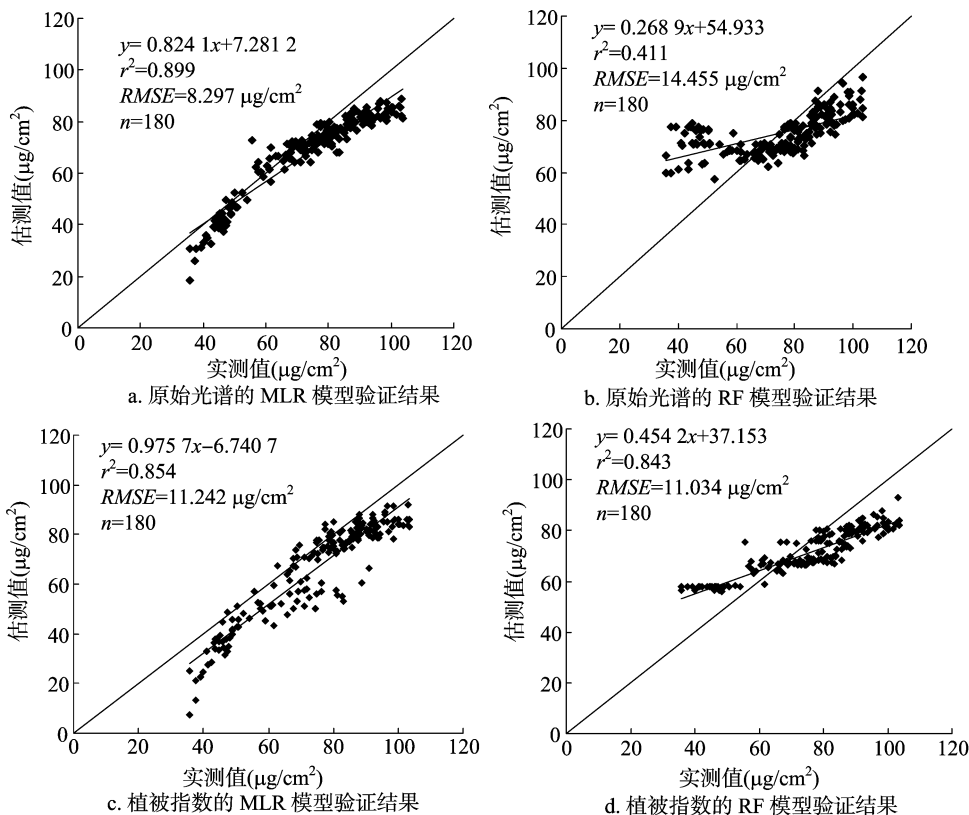


图3 MLR 和 RF 模型的叶绿素含量验证结果

这可能是图 3 中开花期叶绿素含量的估测值不随实测值的增加而改变的原因。另外,2 年同一个生育期的数据采样时间也不完全一致,数据采集时植被内在的理化状况不同,导致估测和验证模型存在高估或低估现象。由于本研究仅使用了 2 年的试验数据,如要大面积应用,仍然需要更多的试验数据积累,以进一步提高模型的估测精度和稳定性。

#### 4 结论

本研究使用 2012、2013 年 2 年采集的数据,建立了基于原始光谱的 MLR 模型和 RF 模型,以及基于植被指数的 MLR 最优模型和 RF 最优模型,得出如下结论:

(1)通过原始光谱与叶片叶绿素含量的相关性分析,筛选出与叶片叶绿素含量相关性最好的波段是 554、708、995 nm,可以作为构建 MLR 模型和 RF 模型的变量。

(2)使用  $|r|$  和 OOB 重要性估计 2 种不同方法进行植被指数的筛选并排序,在各自前 10 个植被指数中,有 9 个共同的植被指数。说明  $|r|$  和 OOB 重要性估计方法在计算植被指数与叶绿素含量的关联性方面一致性很高。

(3)基于原始光谱的 MLR 和 RF 模型的  $R^2$  分别为 0.578 和 0.527,  $RMSE$  分别为 8.240 和 8.728  $\mu\text{g}/\text{cm}^2$ , 验证模型的  $r^2$  分别为 0.899 和 0.411,  $RMSE$  分别为 8.297 和 14.455  $\mu\text{g}/\text{cm}^2$ 。

(4)基于 10 个植被指数的 MLR 最优模型和 5 个植被指数的 RF 最优模型建模  $R^2$  分别为 0.602、0.609,  $RMSE$  分别为 8.004、7.930  $\mu\text{g}/\text{cm}^2$ ;在模型验证方面,基于植被指数的 MLR 最优模型和 RF 最优模型的  $r^2$  分别为 0.854 和 0.843,  $RMSE$

分别为 11.242 和 11.034  $\mu\text{g}/\text{cm}^2$ , 2 个模型精度相近。研究结果表明,基于原始光谱的 MLR 模型的叶绿素含量估测精度高于其他 3 个模型,能够较精确地估测苹果叶片叶绿素含量。另外,基于植被指数的苹果叶片叶绿素含量 RF 估测模型也具有一定的应用潜力。

#### 参考文献:

- [1] 蒋金豹,陈云浩,黄文江. 用高光谱微分指数估测条锈病胁迫下小麦冠层叶绿素密度[J]. 光谱学与光谱分析,2010,30(8): 2243-2247.
- [2] 徐新刚,赵春江,王纪华,等. 新型光谱曲线特征参数与水稻叶绿素含量间的关系研究[J]. 光谱学与光谱分析,2011,31(1): 188-191.
- [3] Mark S, Anatoly G, Donald C R. Nondestructive estimation of leaf chlorophyll content in grapes[J]. American Journal of Enology & Viticulture,2008,59(3):299-305.
- [4] 张 瑶,郑立华,李民赞,等. 基于光谱特征分析的苹果树叶片营养素预测模型构建[J]. 农业工程学报,2013(8):171-178.
- [5] 李敏夏. 苹果叶片光谱反射率与叶绿素和全氮含量的相关研究[D]. 杨凌:西北农林科技大学,2009.
- [6] 李 萍. 基于光谱分析的库尔勒香梨叶片全氮、全磷含量估测模型建立[D]. 乌鲁木齐:新疆农业大学,2013.
- [7] 张蕾蕾. 苹果幼树叶片叶绿素与水分含量的高光谱估测研究[D]. 泰安:山东农业大学,2013.
- [8] 王丽爱,周旭东,朱新开,等. 基于 HJ-CCD 数据和随机森林算法的小麦叶面积指数反演[J]. 农业工程学报,2016(3):149-154.
- [9] 李粉玲,王 力,刘 京,等. 基于高分一号卫星数据的冬小麦叶

- 片 SPAD 值遥感估算[J]. 农业机械学报, 2015, 46(9): 273 – 281.
- [10] 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.
- [11] 韩兆迎, 朱西存, 房贤一, 等. 基于 SVM 与 RF 的苹果树冠 LAI 高光谱估测[J]. 光谱学与光谱分析, 2016, 36(3): 800 – 805.
- [12] 岳继博, 杨贵军, 冯海宽. 基于随机森林算法的冬小麦生物量遥感估算模型对比[J]. 农业工程学报, 2016(18): 175 – 182.
- [13] Lichtenthaler H K. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes[J]. *Methods in Enzymology*, 1987, 148: 350 – 382.
- [14] Peñuelas J, Gamon J A, Fredeen A L, et al. Reflectance indices associated with physiological changes in nitrogen – and water – limited sunflower leaves[J]. *Remote Sensing of Environment*, 1994, 48(2): 135 – 146.
- [15] Peñuelas J, Filella I, Lloret P, et al. Reflectance assessment of mite effects on apple trees[J]. *International Journal of Remote Sensing*, 1995, 16(14): 2727 – 2733.
- [16] Daughtry C S T, Walthall C L, Kim M S, et al. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance[J]. *Remote Sensing of Environment*, 2000, 74(2): 229 – 239.
- [17] Haboudane D, Miller J R, Pattey E, et al. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture[J]. *Remote Sensing of Environment*, 2004, 90(3): 337 – 352.
- [18] Haboudane D, Miller J R, Tremblay N, et al. Integrated narrow – band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture[J]. *Remote Sensing of Environment*, 2002, 81(2): 416 – 426.
- [19] Dash J, Curran P J. The MERIS terrestrial chlorophyll index[J]. *International Journal of Remote Sensing*, 2004, 25(23): 5403 – 5413.
- [20] Sims D A, Gamon J A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages[J]. *Remote Sensing of Environment*, 2002, 81(2/3): 337 – 354.
- [21] Tucker C J. Red and photographic infrared linear combinations for monitoring vegetation[J]. *Remote Sensing of Environment*, 1979, 8(2): 127 – 150.
- [22] Gitelson A A, Merzlyak M N, Chivkunova O B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves[J]. *Photochemistry and Photobiology*, 2001, 74(1): 38 – 45.
- [23] Roujean J L, Breon F M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements[J]. *Remote Sensing of Environment*, 1995, 51(3): 375 – 384.
- [24] Qi J, Chehbouni A, Huete A R, et al. A modified soil adjusted vegetation index[J]. *Remote Sensing of Environment*, 1994, 48(2): 119 – 126.
- [25] Horler D N H, Dockray M, Barber J. The red edge of plant leaf reflectance[J]. *International Journal of Remote Sensing*, 1983, 4(2): 273 – 288.
- [26] Vincini M, Frazzi E, D'Alessio P. Narrow – band vegetation indexes from hyperion and directional chris/proba data for canopy chlorophyll density[C]//Envisat Symposium 2007. Montr eux, Switzerland, 2007.
- [27] Gitelson A A, Merzlyak M N. Remote estimation of chlorophyll content in higher plant leaves[J]. *International Journal of Remote Sensing*, 1997, 18(12): 2691 – 2697.
- [28] Vogelmann J E, Rock B N, Moss D M. Red edge spectral measurements from sugar maple leaves[J]. *International Journal of Remote Sensing*, 1993, 14(8): 1563 – 1575.
- [29] Zarco – Tejada P J, Miller J R, Noland T L, et al. Scaling – up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2001, 39(7): 1491 – 1507.
- [30] Mahlein A K, Steiner U, Dehne H W, et al. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases[J]. *Precision Agriculture*, 2010, 11(4): 413 – 431.
- [31] Zarco – Tejada P J, Pushnik J C, Dobrowski S, et al. Steady – state chlorophyll a fluorescence detection from canopy derivative reflectance and double – peak red – edge effects[J]. *Remote Sensing of Environment*, 2003, 84(2): 283 – 294.
- [32] Kim M S, Daughtry C T, Chapelle E W. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (Apar.)[C]//Proceedings of the Sixth Symposium on Physical Measurements and Signatures in Remote Sensing. Val D'Isure, France, 1994: 299.
- [33] Chen P F, Haboudane D, Tremblay N, et al. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat[J]. *Remote Sensing of Environment*, 2010, 114(9): 1987 – 1997.
- [34] Jin X, Li Z, Feng H, et al. Newly combined spectral indices to improve estimation of total leaf chlorophyll content in cotton[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2014, 7(1): 4589 – 4600.
- [35] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5 – 32.
- [36] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123 – 140.
- [37] 王惠文, 孟 洁. 多元线性回归的预测建模方法[J]. 北京航空航天大学学报, 2007, 33(4): 500 – 504.
- [38] 李旭青, 刘湘南, 刘美玲, 等. 水稻冠层氮素含量光谱反演的随机森林算法及区域应用[J]. 遥感学报, 2014, 18(4): 923 – 945.
- [39] Verikas A, Gelzinis A, Bacauskiene M. Mining data with random forests: a survey and results of new tests[J]. *Pattern Recognition*, 2011, 44(2): 330 – 349.