

基于高光谱技术反演大豆生理信息的特征波长提取方法研究

刘爽, 于海业, 朴兆佳, 陈美辰, 于通, 孔丽娟, 张蕾, 党敬民, 隋媛媛*

吉林大学生物与农业工程学院, 吉林 长春 130022

摘要 生理信息的准确获取及预测可为种植的精细化管理提供依据。传统的大豆生理信息反演方法检测效率低、操作过程繁琐且多为有损检测。利用高光谱技术建立大豆生理信息的快速无损反演方法。以大豆开花结荚期叶片为研究对象, 在2个日期(D1和D2)获取高光谱、叶绿素含量、净光合速率和光合有效辐射数据。首先分别采用多元散射校正(MSC)、标准正态变量变换(SNV)、一阶导数(FD)、二阶导数(SD)、Savitzky-Golay平滑(SG)、MSC-SG-FD、MSC-SG-SD、SNV-SG-FD和SNV-SG-SD共9种方法对原始光谱数据进行预处理, 随后结合偏最小二乘法(PLS)建立全波段模型, 比较分析, 选出最优预处理方法。再分别利用竞争性自适应权重取样法(CARS)、连续投影法(SPA)和相关系数法(CC)对特征波长进行筛选提取。最后将优选出的预处理方法与特征波长变量进行PLS建模并对比分析, 以校正集和预测集相关系数 R_c 和 R_p 为模型评价指标, 最终优选出与大豆生理信息相关性最高的反演模型。结果表明: 采用MSC-SG-FD预处理后建立的叶绿素含量全波段PLS模型的 R_c 和 R_p 最高, 分别为0.909和0.882(D1), 0.909和0.880(D2), 采用SNV-SG-FD预处理后建立的光能利用率全波段PLS模型的 R_c 和 R_p 最高, 分别为0.913和0.894, 0.902和0.869, 与原始及其他预处理后建立的模型相比表现出最高的模型性能特征。进一步对比3种特征波长提取方法的建模, 发现SPA法筛选出的变量能将叶绿素含量反演模型的建模变量数由512个压缩至20个(D1)和23个(D2), 变量压缩率高达96.09%和95.51%, 同时能将光能利用率反演模型的建模变量数压缩至27个和37个, 变量压缩率高达94.73%和92.77%。最终得出反演叶绿素含量的最优建模方法为MSC-SG-FD-SPA-PLS, R_c 值为0.944(D1)和0.941(D2), R_p 值为0.911和0.903, 反演光能利用率的最优建模方法为SNV-SG-FD-SPA-PLS, R_c 值为0.929(D1)和0.925(D2), R_p 值为0.912和0.907, 所建模型精度较高, 可为大面积检测大豆生理信息提供技术支持。

关键词 高光谱; 大豆; 生理信息; 特征波长; 反演模型

中图分类号: S565.1 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)11-3542-07

引言

生理信息的准确获取可使管理者根据其生长状况合理控制作物生长的环境参数, 使水、肥、农药等得到精确的管理, 因此, 生理信息准确获取技术的研究可提高农业生产的科学性, 有效促进我国农业整体水平的快速发展。近年来光谱技术的应用推动了作物生理信息研究方面的快速发展, 尤其是高光谱技术, 获取的光谱波段是连续的, 光谱信息较精准, 因此被广泛应用于作物的生理信息研究^[1]。但是, 高光谱数据的波长间往往会包含一些其他的多余信息, 增加大量的计算工作, 因此对全波段的波长变量进行优选和精简十分必

要。目前的研究表明, 无关信息的存在会极大地降低模型的稳健性, 因此, 对高光谱数据进行特征波长的筛选提取是建立最优大豆生理信息反演模型的重要基础^[2-4]。

近年来提出了一些新型的特征波长变量优选方法, 主要包括竞争性自适应权重取样法(competitive adaptive re-weighted sampling, CARS)、连续投影法(successive projections algorithm, SPA)、相关系数法(Correlation coefficient, CC)和其他算法等^[5]。国内外学者针对不同特征波长提取方法在建立判别模型上的应用开展了大量的研究, Dai等^[6]以虾为研究对象, 以冷冻和非冷冻为试验条件, 将SPA与SNV结合建立新鲜度检测的判别模型, 模型预测集的正确率均高于95%。刘泽蒙等^[7]应用离散萤火虫算法(DFA)进行

收稿日期: 2019-09-21, 修订日期: 2020-01-19

基金项目: 吉林省科技发展计划项目(20170204020NY)和国家自然科学基金青年科学基金项目(31801259)资助

作者简介: 刘爽, 1993年生, 吉林大学生物与农业工程学院博士研究生 e-mail: 13756900435@163.com

* 通讯联系人 e-mail: suiyuan@jlu.edu.cn

光谱特征变量筛选建立发酵液丁二酸含量的 PLSR 模型, R_c 和 R_p 值高达 0.986 和 0.969, 模型精度优于全光谱建模。牛智友等^[8]应用 CARS 算法进行生物质秸秆中 N 和 O 元素含量的高光谱敏感变量优选, 用 PLS 方法建立了元素定量分析模型, 建模变量数分别为 24 个和 10 个, 相对分析误差 RPD 分别为 3.11 和 2.32, 表明可进行实际应用。目前对特征波长的提取方法研究, 多是应用在营养成分及元素含量等方面, 而针对农作物生理信息的研究鲜有报道。

本文通过高光谱技术采集大豆开花结荚期不同叶片的光谱、叶绿素含量、净光合速率和光合有效辐射数据, 采用 9 种预处理方法及 3 种特征波长筛选算法, 筛选出最优预处理方法和特征波长优选方法, 再结合 PLS 建模探讨不同建模方法组合对反演大豆生理信息(叶绿素含量和光能利用率)准确性的影响, 为研究精准、无损且快速的大豆生理信息检测技术提供理论依据。

1 实验部分

实验于吉林大学生物与农业工程学院日光温室内进行, 供试大豆品种为虎山 60, 采用盆栽试验, 于 2019 年 6 月 19 日播种于塑料花盆中, 每盆播种 1 粒种子, 播种 20 盆, 所有植株均进行一致管理, 最后挑选 10 株长势较好、形态均匀且健康无病虫害的大豆植株作为实验样本。

1.1 数据采集与处理

于 2019 年 7 月 27 日和 8 月 4 日(分别以 D1 和 D2 表示)进行 2 次数据采集, 2 次采集的数据种类相同, 采样时间为 9:00—14:00, 测量时天气晴朗, 共采集 50 片大豆功能叶片

的高光谱、叶绿素含量、净光合速率和光合有效辐射 4 种数据。高光谱采用美国 Analytical Spectral Devices 分析光谱仪器公司产 HH2 地物光谱仪测定, 测量范围 325~1 075 nm, 采样间隔 1.4 nm, 分辨率 3 nm@700 nm, 每片叶获取 10 条光谱数据。叶绿素含量采用日本产 SPAD-502 测定, 每片叶获取 3 条数据。因 SPAD-502 读数与叶绿素含量密切相关, 因此文中将其值代表叶绿素含量^[9]。净光合速率和光合有效辐射采用美国产 LI-6400 型光合作用仪测定, 每片叶获取 3 条数据。以上 4 种数据均取其平均值作为分析所用数据。所用的数据处理与分析软件为 ViewSpec Pro, SPSS 24.0, Matlab R2015b 和 Origin 19.0。

1.2 光能利用率的计算

光能利用率通过下式进行计算^[10]

$$L_{ue} = P_n / P_{ar} \times 100\%$$

式中: L_{ue} 为光能利用率; P_n 为净光合速率; P_{ar} 为光合有效辐射。

2 结果与讨论

2.1 基于不同预处理方法的高光谱建模分析

采用 5 种单一预处理方法分别为 MSC, SNV, SG, FD 和 SD 及 4 种组合方法分别为 MSC-SG-FD, MSC-SG-SD, SNV-SG-FD, SNV-SG-SD 共 9 种方法对原始光谱进行处理, 采用梯度法选取 40 个样本作为校正集, 其余 10 个样本作为预测集建立全波段 PLS 模型, 模型的评价指标为校正集和预测集相关系数 R_c 和 R_p , 其中, RAW 预处理表示原始光谱, 作为其他预处理方法的对照, 结果如表 1。可以看出, 对于

表 1 高光谱预处理方法优选结果

Table 1 Optimization results from hyperspectral pre-processing methods

建模对象	日期	预处理方法	R_c	R_p	建模对象	日期	预处理方法	R_c	R_p
叶绿素含量	D1	RAW	0.842	0.828	光能利用率	D1	RAW	0.862	0.825
		MSC	0.869	0.846			MSC	0.887	0.850
		SNV	0.867	0.839			SNV	0.891	0.854
		SG	0.855	0.838			SG	0.894	0.821
		FD	0.858	0.824			FD	0.885	0.856
		SD	0.859	0.818			SD	0.862	0.827
		MSC-SG-FD	0.909	0.882			MSC-SG-FD	0.900	0.878
	MSC-SG-SD	0.893	0.833	MSC-SG-SD	0.891	0.887			
	SNV-SG-FD	0.896	0.870	SNV-SG-FD	0.913	0.894			
	SNV-SG-SD	0.903	0.838	SNV-SG-SD	0.907	0.862			
	D2	RAW	0.875	0.730	D2	RAW	0.869	0.838	
		MSC	0.903	0.876		MSC	0.881	0.856	
		SNV	0.907	0.867		SNV	0.882	0.857	
		SG	0.887	0.869		SG	0.889	0.787	
FD		0.883	0.814	FD		0.886	0.833		
SD		0.873	0.788	SD		0.870	0.823		
MSC-SG-FD		0.909	0.880	MSC-SG-FD		0.896	0.808		
MSC-SG-SD	0.899	0.873	MSC-SG-SD	0.900	0.865				
SNV-SG-FD	0.909	0.816	SNV-SG-FD	0.902	0.869				
SNV-SG-SD	0.897	0.869	SNV-SG-SD	0.900	0.866				

同一种生理信息, D1 和 D2 建立的最优模型所用预处理方法相同, MSC-SG-FD 预处理后建立的叶绿素含量模型校正集和预测集相关系数最高, D1 所建模型的 R_c 和 R_p 分别为 0.909 和 0.882, D2 所建模型的 R_c 和 R_p 分别为 0.909 和 0.880。SNV-SG-FD 预处理后建立的光能利用率模型校正集和预测集相关系数最高, D1 所建模型的 R_c 和 R_p 分别为 0.913 和 0.894, D2 所建模型的 R_c 和 R_p 分别为 0.902 和 0.869。选取 MSC-SG-FD 作为建立叶绿素含量反演模型所用预处理方法, SNV-SG-FD 作为建立光能利用率反演模型所用预处理方法。

2.2 特征波长优选方法分析

本研究采用的光谱仪有 512 个波长点, 光谱变量较多, 含有较多的冗余和共线性变量, 需对其进行优选以获取最有效的光谱信息。将 MSC-SG-FD 和 SNV-SG-FD 预处理后的光谱进行特征波长筛选以分别建立大豆叶绿素含量和光能利用率的 PLS 反演模型。

2.2.1 CARS 法筛选特征波长

基于 CARS 法进行大豆生理信息反演模型的特征波长提取结果如图 1。CARS 算法中, 每次通过自适应加权采样 (adaptive reweighted sampling, ARS) 保留 PLS 模型中回归系数绝对值权重较大的点作为新的子集, 去掉权重较小的点, 然后基于新的子集建立 PLS 模型, 经多次计算, 选择 PLS 模型交叉验证均方根误差 (RMSECV) 最小的子集中的波长作为特征波长。由于 CARS 法中的蒙特卡罗采样随着采样次数的不同呈现不同的运算结果^[11-12], 所以本文通过设定不同的采样次数后分别进行运算以选取相对较优的波长变量, 图 1 为将采样次数设为 50 次的最优运算结果。如图 1(a) 所示, 采样次数较少时, 由于指数衰减函数 (exponentially decreasing function, EDF) 的作用, CARS 法保留的波长变量数快速降低, 当采样次数逐渐增加时, 保留变量数的降低速度逐渐减缓。由图 1(a) 可见, RMSECV 值呈现不同程度的波动, “*” 线标出最小 RMSECV 值所对应的采样次数, 即从开始到此次采样次数的运算过程中剔除了与大豆生理信息无关的信息, 则此次所选的波长子集即为建立大豆生理信息

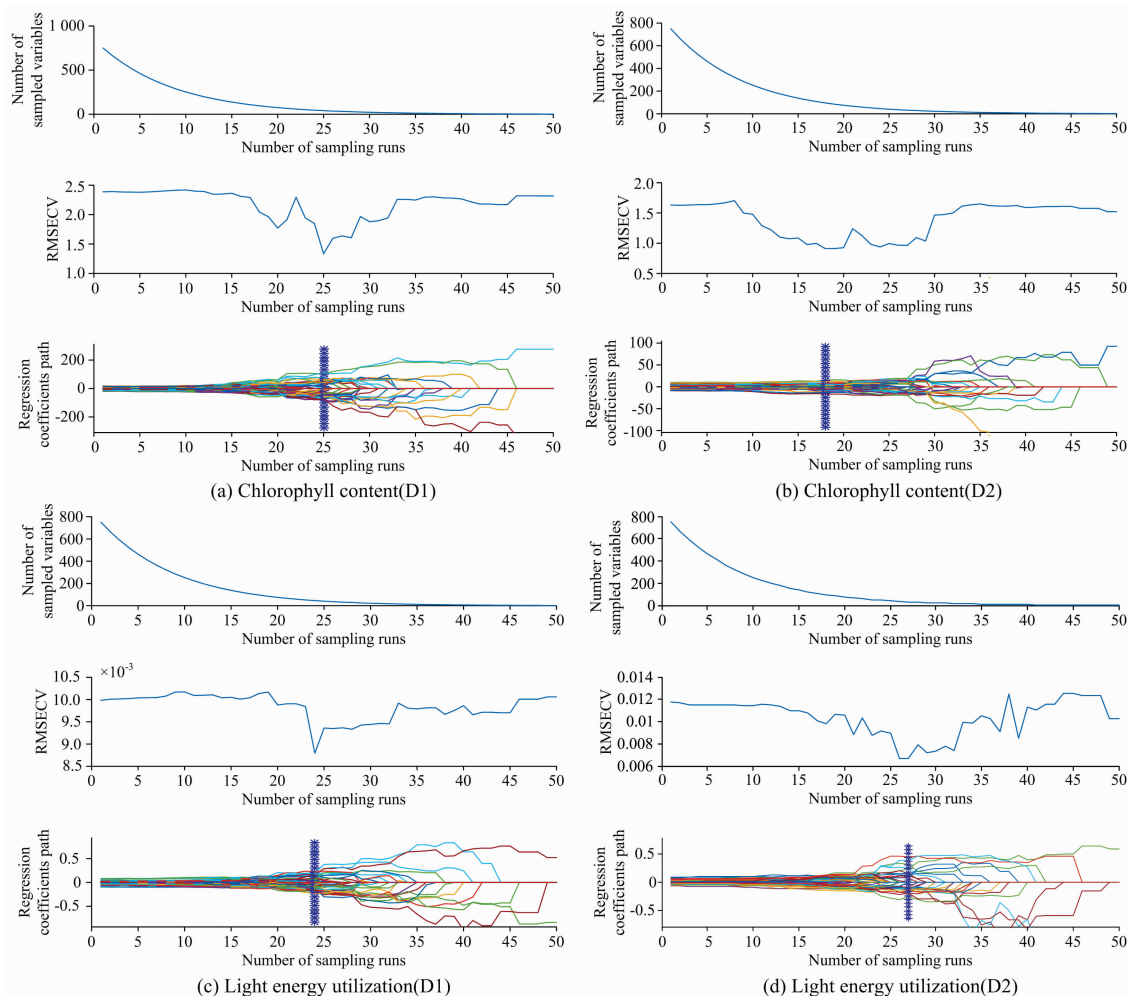


图 1 CARS 法筛选特征波长

(a): 叶绿素含量(D1); (b): 叶绿素含量(D2); (c): 光能利用率(D1); (d): 光能利用率(D2)

Fig. 1 Characteristic wavelengths selected by CARS algorithm

(a): Chlorophyll content (D1); (b): Chlorophyll content (D2);
(c): Light energy utilization (D1); (d): Light energy utilization (D2)

高光谱反演模型的最优波长，所选的波长数量见图 1(a)。由图 1(a)和(b)可知，第 25 次(D1)和第 18 次(D2)采样时，RMSECV 值最小，对应的最优大豆叶绿素含量建模变量数分别为 41 个和 96 个，剔除的变量数分别为 471 个和 416 个，所选变量占全波段的 8.01%和 18.75%。由图 1(c)和(d)可知，第 24 次(D1)和第 27 次(D2)采样时，RMSECV 值最小，对应的最优大豆光能利用率建模变量数分别为 46 个和 32 个，剔除的变量数分别为 466 个和 480 个，所选变量占全波段的 8.98%和 6.25%。

2.2.2 SPA 法筛选特征波长

基于 SPA 法进行大豆生理信息反演模型的特征波长提取结果如图 2。SPA 法在运算过程中通过分析投影向量的大

小进行特征波长变量的筛选，通过计算校正模型的 RMSE 值，其值最小时，对应的波长子集即为优选波长^[13]。由图 2 可见，随变量数的增加，RMSE 值逐渐减小，“□”表示最小 RMSE 值，对应的即为最优特征波长变量。如图 2(a)和(b)所示，经 SPA 法后分别筛选出 20 个(D1)和 23 个(D2)特征波长用于建立最优的大豆叶绿素含量反演模型，剔除的变量数分别为 492 个和 489 个，所选变量占全波段的 3.91%和 4.49%。如图 2(c)和(d)所示，分别筛选出 27 个(D1)和 37 个(D2)特征波长用于建立最优的大豆光能利用率反演模型，剔除的变量数分别为 485 个和 475 个，所选变量占全波段的 5.27%和 7.23%。

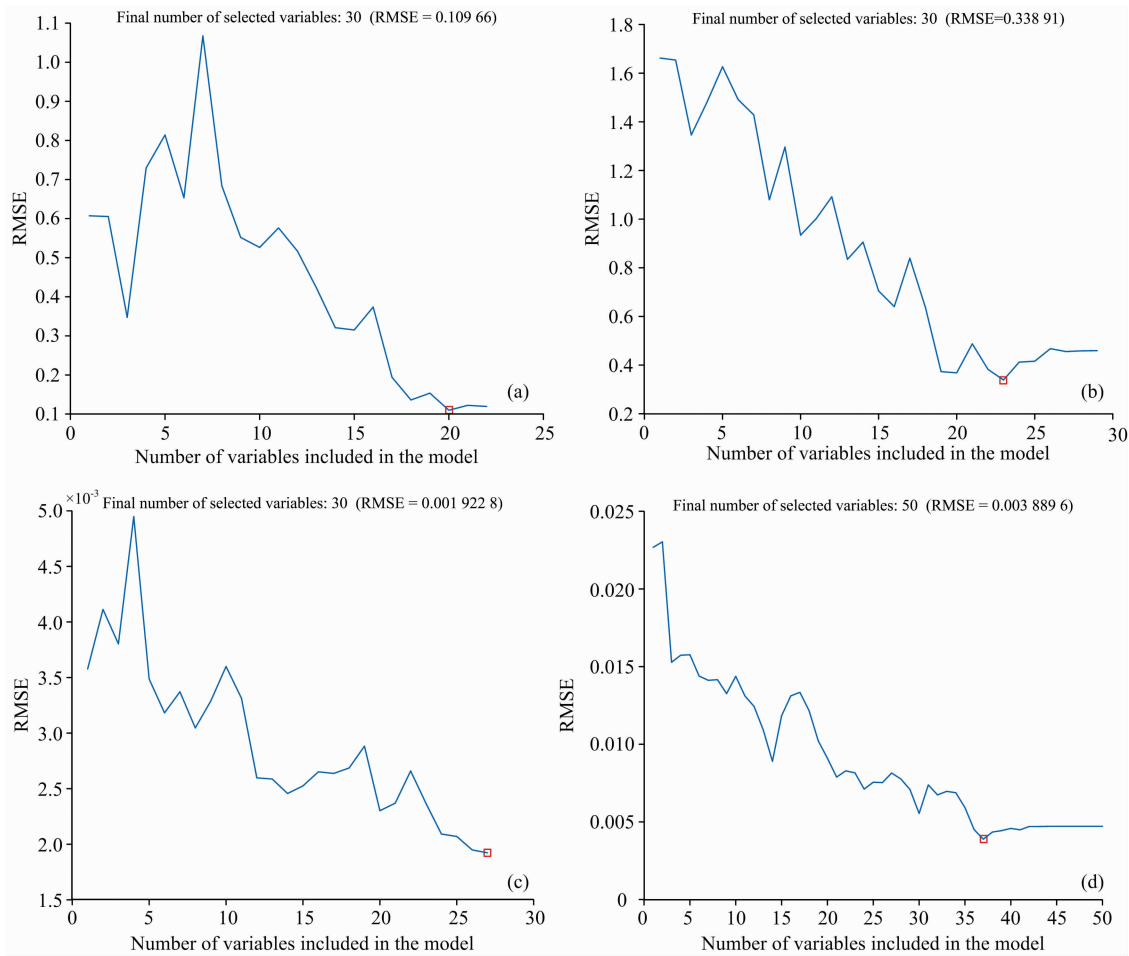


图 2 SPA 法筛选特征波长

(a): 叶绿素含量(D1); (b): 叶绿素含量(D2); (c): 光能利用率(D1); (d): 光能利用率(D2)

Fig. 2 Characteristic wavelengths selected by SPA algorithm

(a): Chlorophyll content (D1); (b): Chlorophyll content (D2);
(c): Light energy utilization (D1); (d): Light energy utilization (D2)

2.2.3 CC 法筛选特征波长

基于 CC 法进行大豆生理信息反演模型的特征波长提取结果如图 3。CC 法通过计算光谱中每一个波长下的光谱值与叶绿素含量和光能利用率的相关系数，其值的绝对值越大，越可能选此波长作为特征波长，结合波长与对应的相关系数

做出波长-相关系数图进行特征波长筛选。由图 3(a)和(b)可知，大豆高光谱与叶绿素含量在 0.05 显著性水平下相关的阈值分别为 -0.254 (D1)和 -0.236 (D2)，阈值线以下的波长为筛选出的特征波长，筛选出的特征波长变量分别为 221 个和 97 个，剔除的变量数分别为 291 个和 415 个，所选变量

占全波段的 43.16% 和 18.95%。由图 3(c) 和 (d) 可知大豆高光谱与光能利用率在 0.05 显著性水平下相关的阈值分别为 -0.279 (D1) 和 -0.236 (D2)，筛选出的特征波长变量分别为 234 个和 224 个，剔除的变量数分别为 278 个和 288 个，所选变量占全波段的 45.7% 和 43.75%。

2.3 大豆生理信息反演模型的建立与比较

将 MSC-SG-FD 和 SNV-SG-FD 预处理后的光谱与不同特征波长优选方法组合，分别对应进行叶绿素含量和光能利用率的 PLS 建模。基于上述 3 种波长优选方法的 PLS 建模结果如表 2 所示。与全波段变量所建模型相比，变量优选后

所建模型的反演结果均有明显提升，通过比较发现，SPA 法所建模型的性能优于原始及另外 2 种特征波长提取方法所建模型。其中，叶绿素含量反演模型中，SPA 法优选出的变量数最少，变量压缩率高达 96.09% (D1) 和 95.51% (D2)，模型的校正 R_c 值从 0.909 分别提高到 0.944 和 0.941，预测 R_p 值从 0.882 和 0.880 分别提高到 0.911 和 0.903，建模因子数从 10 个减少到 7 个。光能利用率反演模型中，SPA 法优选出的变量压缩率高达 94.73% (D1) 和 92.77% (D2)，模型的校正 R_c 值从 0.913 和 0.902 分别提高到 0.929 和 0.925，预测 R_p 值从 0.894 和 0.869 分别提高到 0.912 和 0.907。

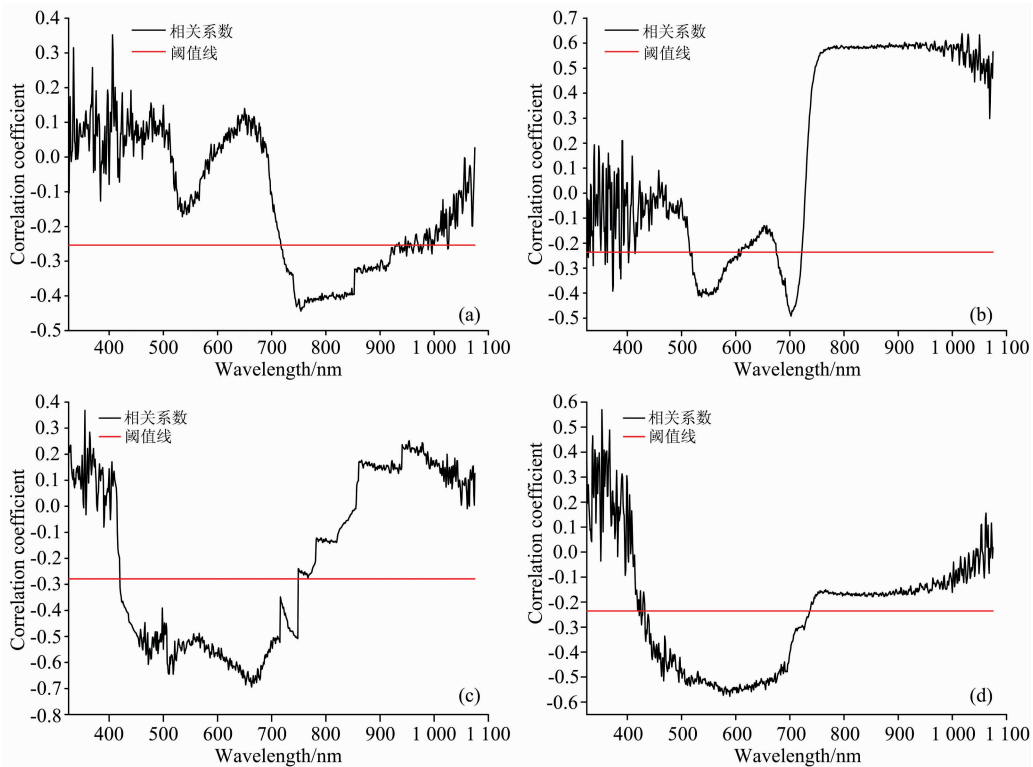


图 3 CC 法筛选特征波长

(a): 叶绿素含量(D1); (b): 叶绿素含量(D2); (c): 光能利用率(D1); (d): 光能利用率(D2)

Fig. 3 Characteristic wavelengths selected by CC algorithm

(a): Chlorophyll content (D1); (b): Chlorophyll content (D2);
(c): Light energy utilization (D1); (d): Light energy utilization (D2)

表 2 大豆生理信息反演模型结果

Table 2 Soybean physiological information inversion model results

建模对象	日期	建模方法	变量保留数	P_{cs}	R_c	R_p	建模对象	日期	建模方法	变量保留数	P_{cs}	R_c	R_p
叶绿素含量	D1	PLS	512	10	0.909	0.882	光能利用率	D1	PLS	512	10	0.913	0.894
		CC-PLS	221	9	0.911	0.906			CC-PLS	234	9	0.919	0.902
		CARS-PLS	41	9	0.927	0.892			CARS-PLS	46	9	0.921	0.909
		SPA-PLS	20	7	0.944	0.911			SPA-PLS	27	7	0.929	0.912
	D2	PLS	512	10	0.909	0.880	D2	PLS	512	10	0.902	0.869	
		CC-PLS	97	9	0.902	0.898		CC-PLS	224	9	0.907	0.885	
		CARS-PLS	96	9	0.914	0.899		CARS-PLS	32	9	0.912	0.898	
		SPA-PLS	23	7	0.941	0.903		SPA-PLS	37	7	0.925	0.907	

0.907, 建模因子数从 10 个减少到 7 个。可见 SPA-PLS 模型的计算和预测性能较 PLS, CARS-PLS 和 CC-PLS 模型均有显著提高, 表明通过 SPA 法筛选后, 尽可能地剔除了噪声数据, 特征波长的有效信息完全释放了出来, 避免了有效特征信息被其他相关性不大的波长所遮蔽的现象, SPA 法对大豆生理信息的反演表现出了最优的性能, 因此反演大豆生理信息模型的最优波长筛选方法为 SPA 法。

3 结 论

以大豆开花结荚期叶片为研究对象, 在 2 个日期(D1 和 D2)进行了大豆叶片高光谱、叶绿素含量、净光合速率和光合有效辐射的数据测量, 采用 5 种不同的单一预处理方法、4 种组合预处理方法及 3 种特征波长变量筛选算法对全波段高光谱数据进行处理, 对筛选出的最优预处理方法和特征波长变量进行组合建立大豆生理信息(叶绿素含量和光能利用率)的反演模型。结果表明: (1) 在 MSC, SNV, SG, FD, SD, MSC-SG-FD, MSC-SG-SD, SNV-SG-FD 和 SNV-SG-SD 9 种全波段光谱 PLS 建模预处理方法中, MSC-SG-FD 组合为大豆叶绿素含量反演模型的最优预处理方法, SNV-SG-FD 组合为大豆光能利用率反演模型的最优预处理方法; (2) 3 种特征波长提取方法 CARS, SPA 和 CC 法均有效减少了建模变量数, 其中 SPA 法优选出的有效变量数最少, 从 512 个变量

中分别优选出了 20 个(D1 叶绿素含量)、23 个(D2 叶绿素含量)、27 个(D1 光能利用率)和 37 个(D2 光能利用率)变量, 对应的变量压缩率分别高达 96.09%, 95.51%, 94.73% 和 92.77%; (3) 反演大豆叶绿素含量的最优建模方法为 MSC-SG-FD-SPA-PLS, 所用变量数分别为 20 个(D1)和 23 个(D2), 建模因子数为 7 个, 对应的 R_c 值为 0.944 和 0.941, R_p 值为 0.911 和 0.903; (4) 反演大豆光能利用率的最优建模方法为 SNV-SG-FD-SPA-PLS, 所用变量数分别为 27 个(D1)和 37 个(D2), 建模因子数为 7 个, 对应的 R_c 值为 0.929 和 0.925, R_p 值为 0.912 和 0.907。本研究可为大田及大范围大面积种植大豆时检测其生理信息提供参考, 具有重要的指导和实践意义。

高光谱技术在作物生理信息反演的实际应用中仍存在一些有待解决, 如不同品种的大豆或同一品种但生长地点不同时, 其吸收能量及光合作用等能力必然会存在不同程度的差异, 导致对光谱曲线及生理信息数据产生影响, 进而影响提取的特征波长变量, 则在进行生理信息反演时需要大量的数据及建模运算, 工作量较大。另外, 本研究及大多数学者的研究多以单一大豆生长期为试验时期, 其结果应用于大豆的整个生长期时效果并不理想。如何将单个时期与整个生长期的生理信息反演模型通用并达到一定的反演精度仍需更深入的探索与实践研究。

References

- [1] Huang L, Ding W, Liu W, et al. *Journal of Plant Pathology*, 2019.
- [2] Zhu H Y, Chu B Q, Zhang C, et al. *Scientific Reports*, 2017, 1(1).
- [3] DONG Chun-wang, LIANG Gao-zhen, AN Ting, et al(董春旺, 梁高震, 安 霆, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2018, 34(24): 306.
- [4] YU Lei, HONG Yong-sheng, ZHOU Yong, et al(于 雷, 洪永胜, 周 勇, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2016, 32(13): 95.
- [5] HE Yong, LIU Fei, LI Xiao-li, et al(何 勇, 刘 飞, 李晓丽, 等). *Spectroscopy and Imaging Technology in Agriculture(光谱及成像技术在农业中的应用)*. Beijing: Science Press(北京: 科学出版社), 2016. 122.
- [6] Dai Q, Cheng J H, Sun D W, et al. *Journal of Food Engineering*, 2015, 149: 97.
- [7] LIU Ze-meng, ZHANG Rui, ZHANG Guang-ming, et al(刘泽蒙, 张 瑞, 张广明, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2016, 36(12): 3931.
- [8] NIU Zhi-you, LI Xiao-jin, GAO Hai-long(牛智有, 李晓金, 高海龙). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2014, 30(22): 181.
- [9] Wakiyama Y. *Japan Agricultural Research Quarterly*, 2016, 50(4): 329.
- [10] ZHOU Li-na, YU Hai-ye, YU Lian-jun, et al(周丽娜, 于海业, 于连军, 等). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2014, 45(7): 255.
- [11] Chen Q, Zhang D, Pan W, et al. *Trends in Food Science and Technology*, 2015, 43(1): 63.
- [12] Kutsanedzie F Y H, Chen Q S, Hassan M M, et al. *Food Chemistry*, 2018, 240: 231.
- [13] Shao Y Y, Xuan G T, Hu Z C, et al. *PLoS ONE*, 2019, 14(9): E0222633.

Study on Extracting Characteristic Wavelength of Soybean Physiological Information Based on Hyperspectral Technique

LIU Shuang, YU Hai-ye, PIAO Zhao-jia, CHEN Mei-chen, YU Tong, KONG Li-juan, ZHANG Lei, DANG Jing-min, SUI Yuan-yuan*

School of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China

Abstract The accurate acquisition and prediction of physiological information can provide a basis for the fine management of planting. The traditional soybean physiological information inversion methods have low detection efficiency, cumbersome operation process and mostly damage detection. To this end, this paper uses hyperspectral technology to establish a rapid non-destructive inversion method for soybean physiological information. The leaves of soybean flowering and pod-forming period were taken as research objects, and the hyperspectral, chlorophyll content, net photosynthetic rate and photosynthetically active radiation data were obtained on 2 dates (D1 and D2). First, multiple scatter correction (MSC), standard normal variable transformation (SNV), first derivative (FD), second derivative (SD), Savitzky-Golay smoothing (SG), MSC-SG-FD, MSC-SG-SD, SNV-SG-FD and SNV-SG-SD methods are used to preprocess the original spectral data, then establish a full-band model by partial least squares (PLS), compare and analyze, and select the optimal preprocessing method. The feature wavelengths are filtered and extracted by competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA) and correlation coefficient (CC), respectively. Finally, the preferred preprocessing method and the characteristic wavelength variable are modeled and compared by PLS. The correlation coefficients R_c and R_p of the correction set and the prediction set are used as model evaluation indexes, and the inversion model with the highest correlation with soybean physiological information is finally selected. The results showed that the R_c and R_p of the full-band PLS model with chlorophyll content established by MSC-SG-FD pretreatment were the highest, 0.909 and 0.882 (D1), 0.909 and 0.880 (D2), respectively, the R_c and R_p of the full-band PLS model with light energy utilization established by SNV-SG-FD pretreatment are the highest, 0.913 and 0.894, 0.902 and 0.869, respectively, which shows the highest model performance characteristics compared with the original and other pre-processed models. Further comparing the modeling of the three characteristic wavelength extraction methods, it is found that the variables selected by the SPA algorithm can compress the modeling variables of the chlorophyll content inversion model from 512 to 20 (D1) and 23 (D2), and the variable compression rate is as high as 96.09% and 95.51%. At the same time, the modeling variables of the light energy utilization inversion model can be compressed to 27 and 37, and the variable compression rate is as high as 94.73% and 92.77%. Finally, the optimal modeling method for inversion of chlorophyll content is MSC-SG-FD-SPA-PLS with R_c values of 0.944 (D1) and 0.941 (D2), R_p values of 0.911 and 0.903, and the optimal modeling method for inversion of light energy utilization is MSC-SG-FD-SPA-PLS with R_c values of 0.929 (D1) and 0.925 (D2), R_p values of 0.912 and 0.907. The model has high precision and can provide technical support for large-area detection of physiological information.

Keywords Hyperspectral; Soybean; Physiological information; Characteristic wavelength; Inversion model

(Received Sep. 21, 2019; accepted Jan. 19, 2020)

* Corresponding author