

基于近红外光谱技术快速监测小球藻生长过程中的关键指标

宋雯辉 杜世杰 刘言* 王桥 柳鑫 官智勇

(武汉轻工大学食品科学与工程学院, 武汉 430023)

摘要 采用传统检测方法对小球藻生长过程中的生物量、蛋白质和叶绿素含量等关键指标进行监测时存在操作复杂、检测速度慢和难以大规模应用等问题。本研究通过近红外光谱技术结合化学计量学方法,建立了一种快速高效监测小球藻生长过程中关键指标的方法。利用近红外光谱仪采集不同生长阶段的小球藻藻液在近红外光谱,使用标准方法检测对应样品的生物量、蛋白质和叶绿素含量 3 种关键指标,基于偏最小二乘回归法(Partial least squares regression, PLSR)建立定量分析模型。为了提升模型的预测能力,使用多元散射校正(Multiplicative scatter correction, MSC)减少散射对原始光谱(Raw spectrum, RS)的干扰;采用标准正态变换(Standard normal variate, SNV)对原始光谱数据进行规范化处理,消除样品间差异;使用连续小波变换(Continuous wavelet transform, CWT)获取光谱数据中的关键特征;采用一阶导数(1st derivative, 1st)处理增强原始光谱特征的分度;使用随机化检测(Randomization test, RT)和蒙特卡罗-无信息变量消除法(Monte Carlo-uninformative variable elimination, MC-UVE)筛选波长中的有效变量。通过评价不同模型的预测能力,最终确定小球藻生物量、蛋白质和叶绿素成分含量的定量分析模型。研究表明,基于 1st 结合 RT 处理光谱方法建立的模型对小球藻营养成分含量具有更好的预测能力,对生物量、蛋白质和叶绿素的预测值的均方根误差(Root mean square error of prediction, RMSEP)分别为 0.041、0.012 和 0.517,决定系数(Coefficient of determination, R^2)分别为 0.933、0.973 和 0.962。结果表明,本研究提出的模型具有良好的实际应用价值,能够快速准确地同时预测小球藻的生物量、蛋白质与叶绿素含量。

关键词 近红外光谱; 小球藻; 光谱预处理; 波长筛选; 偏最小二乘回归

全球微藻类市场需求量呈现稳定增长的态势,预计年复合增长率约为 5%,到 2030 年底其市值有望达到 200 亿元^[1]。在当今食品科学领域,利用微藻作为原料进行新资源食品开发以及营养功能强化已成为重要趋势。在新食品资源、功能性食品、动物饲料、化妆品工业、生物柴油以及医疗健康等领域,研究者均对微藻进行了开发研究^[2]。小球藻是绿藻门中一种微小单细胞藻类生物,其细胞内含有叶绿素和类胡萝卜素等色素,外观呈现绿色或黄绿色,富含蛋白质、脂肪、维生素和矿物质等营养成分^[3]。鉴于其营养价值、生物活性功能以及在不同环境(如淡水和人工培养基)中的良好生长适应性,小球藻在食品、医药、生物和新能源等领域广泛应用^[4]。随着小球藻应用范围扩大以及市场需求增加,在养殖生产、应用研发和市场监管等方面,都需要对小球藻生长过程中的品质进行检测、监控和评估,以便即时指导生产。

小球藻的生长状态可由一些生物指标反映,选取其中的关键指标进行监测,可即时了解其生长状态以及质量水平。通常,生物量可用于评估小球藻的总质量或生长数量,反映小球藻有效成分的积累量。蛋白质作为小球藻的初级代谢产物,是其主要营养成分之一,蛋白质的积累量体现了小球藻的品质,是评价产品品质以及指导生产的关键指标^[5]。叶绿素是小球藻进行光合作用的重要色素,可反映小球藻的生长发育情况。通过监测生物量、蛋白质以及叶绿素这 3 个关键指标的含,即可大致评估出小球藻的生长情况,进而即时指导生产。传统的理化检测方法虽能够实现对上述 3 种指标的检测,但存在检测时间长、过程复杂、有机溶剂使用过多以及检测效率低等问题^[6-7]。近红外光谱分析方法是一种快速无损的检测方法,依据不同样品对近红外光吸收、散射与反射的不同,得到不同的光谱数据,再结合化学计量法

2024-12-25 收稿; 2025-02-19 接受

国家重点研发计划青年科学家项目(No. 2022YFF1102500)资助。

* E-mail: liuyanwhpu@163.com

提取关键数据,从而实现样品的定性和定量分析^[8-9]。近红外光谱分析技术分析速度快,适用于快速检测不同领域的大批量样品,具有广阔的应用前景^[10-13]。由于近红外光谱对样品成分变化敏感,同时受环境、光源和仪器等因素影响,在实际应用中,需采用物理、化学、数学与信息科学等手段处理数据,建立稳定精确的分析模型,提高预测准确性^[14-15]。在食品领域,近红外光谱分析技术应用广泛,包括农产品中水分、脂肪和蛋白质等营养的含量检测^[16]以及水产品新鲜度监测^[17-18]、食品掺伪掺假检测、乳制品与昂贵精油的质量鉴别^[19-21]等。

目前,基于近红外光谱分析技术实现对小球藻综合指标检测的相关研究较少,部分研究仅集中于油脂或蛋白质等单个指标的检测。蛋白质由氨基酸通过肽键连接而成,其中的肽键在近红外区域有特征吸收,氨基酸残基中的C—H、N—H和O—H等化学键的振动也会在近红外区产生吸收,这些吸收峰的位置和强度与蛋白质的结构和含量密切相关。随着小球藻生物量增加、细胞数量增多,光在细胞间的散射和反射也会增强,导致漫反射光的强度发生变化。小球藻叶绿素分子中含有共轭双键系统,在近红外区域具有独特的吸收特性^[22]。鉴于此,本研究基于近红外光谱技术,开发了一种快速同时检测小球藻生长过程中多个关键指标的方法。本方法基于偏最小二乘回归(Partial least squares regression, PLSR)技术构建分析模型,用于准确快速评估小球藻生长过程的关键指标含量。为了增强模型预测的准确性,分别使用多元散射校正(Multiplicative scatter correction, MSC)、标准正态变换(Standard normal variate, SNV)、连续小波变换(Continuous wavelet transform, CWT)和一阶导数(1st derivative, 1st)方法对原始光谱(Raw spectrum, RS)数据进行预处理。在此基础上,进一步采用蒙特卡洛-无信息变量消除法(Monte Carlo-uninformative variable elimination, MC-UVE)删除无关变量,采用随机化检测(Randomization test, RT)验证变量的重要性,从而筛选出光谱中的关键变量,进一步精炼模型并提升其预测性能。利用所构建的模型对小球藻的各项生长参数进行预测,并将预测结果与通过标准方法获得的参考数据进行对比验证,最终建立了快速稳定的小球藻生物量、蛋白质和叶绿素定量分析方法。

1 实验部分

1.1 仪器与试剂

Frontier 近红外光谱仪(美国Perkins Elmer公司); Multiskan FC型酶标仪(上海赛默飞世尔仪器有限公司); 恒温培养箱(山东欧莱仪器有限公司); GR85DA压力蒸汽灭菌器(厦门致微仪器有限公司); 电子天平(上海梅特勒-托利多仪器有限公司); BBS-V800超净工作台(济南贝西生物技术有限公司); YM-040PLUS超声波清洗机(深圳市方奥微电子有限公司); DHG-9123A电热鼓风干燥箱(上海一恒科技有限公司); Water-DUF超纯水器(上海和泰仪器有限公司); BCA蛋白浓度测定试剂盒(上海碧云天生物技术有限公司); TGL-16M离心机(湖南湘仪有限公司)。

小球藻藻种(中国科学院淡水藻种库); BG11培养基(上海光语生物科技有限公司); 无水乙醇和丙酮(分析纯,国药集团化学试剂有限公司)。

1.2 培养方法

将小球藻藻种按6%的体积比接种至培养瓶中,放置于小球藻培养光反应器架上,并将硅胶管接到培养瓶的通气管上。实验室的室内温度保持在25℃,相对湿度约为80%。使用生物日光灯进行光照,光照强度设置为2500 Lux,光照周期为12 h/d。培养过程中,培养瓶通过硅胶管持续通入空气,以确保CO₂在瓶中的溶解量。接种后持续培养,每天定时从各组培养瓶中收集小球藻样品进行光谱采集和标准值数据测定。实验设置2批次独立样本:第一批次样本3组培养瓶,收集53份样品;第二批次样本1组培养瓶,收集33份样品。

1.3 标准值测定

每天在同一时间段取样测定生物量、蛋白质和叶绿素含量的标准值,所有标准值均平行检测3次后取平均值。采用干重法检测小球藻生物量:移取50 mL小球藻匀浆液于离心管中,以6000 r/min离心10 min,弃去上清液,将藻泥置于60℃烘箱中干燥6~8 h至恒重,称重得到干物质含量。采用BCA法检测蛋白质含量,按照试剂盒说明书使用。参照我国农业行业标准NY/T 3082—2017^[23],采用分光光度法

测定微藻中叶绿素含量。

1.4 光谱采集

采集培养瓶中充分混匀的小球藻样品,置于 1 mm 光程的石英比色皿中,使用 Frontier 近红外光谱仪采集小球藻样品的透射光谱,光谱波长范围为 5284~10000 cm^{-1} ,光谱采集间隔为 4 cm^{-1} 。每个小球藻样品扫描 3 次取平均值,近红外光谱仪工作环境温度恒定为 25 $^{\circ}\text{C}$,相对湿度保持在 20%~35%之间。

1.5 建模计算

在第一批次 3 组培养瓶中随机选取 2 组,将其中采集的样品划为训练集,另一组培养瓶中采集的样品划为预测集。第二批次培养瓶所得样本作为验证集。训练集的样品主要用于建立定量分析模型以及对模型进行优化,预测集的样品主要用于评价定量分析模型的预测能力,验证集的样品用于验证模型的预测结果。

在构建定量分析模型的前期阶段,针对训练集与预测集中的样品光谱,依次进行 SNV、CWT、MSC 和 1st 数据处理。随后,依托 PLSR 技术,使用预处理后的光谱数据,建立针对小球藻生物量、蛋白质以及叶绿素含量的定量分析模型。其中,PLSR 技术在建模中涉及的因子数采用 MCCV 确定。为初步评估已构建模型的预测效能,采取留一交叉验证法(Leave one out cross validation, LOOCV)对训练集样品的生物量、蛋白质与叶绿素含量进行预测^[24],然后将模型的预测结果与理化方法测定结果进行比较。计算预测结果的交叉验证均方根误差(Root mean square error of cross validation, RMSECV)、决定系数(Coefficient of determination, R^2)和相对性能偏差(Relative performance deviation, RPD),用于初步评估模型的预测性能,并以此为基础确定最佳光谱处理方法。随后,采用 MC-UVE 删除无关变量,RT 法验证变量的重要性,对光谱数据进行波长筛选,进一步精炼模型。根据筛选得到的波长数据,重新构建一个定量分析模型,并重新评价新模型的性能。利用优化的新模型,依据小球藻样品的预测集光谱数据,推算小球藻样品的生物量、蛋白质和叶绿素的含量水平,并将其与已知的标准值进行对比分析,计算 RMSEP、 R^2 以及 RPD,对模型进一步评估。整个模型的建立和计算分析工作均借助 Matlab 2023a 软件完成。

2 结果与讨论

2.1 光谱预处理

分别利用 RS 及经过多种预处理方法组合优化后的光谱数据,构建了相应的 PLSR 定量分析模型。由 PLSR 定量分析模型的初步评价结果(表 1)可知,光谱原数据经过预处理后,建立的定量分析模型的预测能力有所提高。对于不同的指标,提高效果有所不同。对于生物量指标,8 种光谱预处理在模型性能提升方面有限,但是经过 1st 处理后对应的预测模型效果优于其它光谱预处理方式。对于蛋白质含量指标,利用 RS 构建的 PLSR 模型预测结果的 RMSECV 为 0.017, R^2 为 0.924, RPD 为 3.667,而采用经过 1st 处理 RS 后得到的光谱数据建立的 PLSR 模型,其预测结果的 RMSECV 降至 0.009, R^2 提升至 0.980, RPD 提升至 7.113,明显优于 RS 的模型预测结果,也较其它光谱预处理方式的预测结果更优。由此可见,1st 处理后的光谱模型预测可以得到更好的结果。针对叶绿素含量预测,采用 RS 构建的 PLSR 模型,其预测结果的 RMSECV 为 0.586, R^2 为 0.955, RPD 为 4.760。当 RS 经过 1st 处理后,所建立的 PLSR 模型预测性能得到了显著提升,效果也明显优于其它处理方法。综合考虑,选择 1st 处理方式优化小球藻样品的 RS 数据,进行光谱筛选。

2.2 波长选择

样品的 RS 如图 1 所示。样品的光谱数据中包含了大量信息,其中一部分信息与目标分析无关,甚至可能干扰分析结果^[25]。为增强 PLSR 模型的预测精确度并降低无关数据的影响,选用 MC-UVE 与 RT 两种计算方法,对 1st 处理的光谱数据进行波长选择,得到与目标信息相关的光谱波长点^[26]。针对不同的检测指标,各类光谱的筛选结果如图 2 所示,针对不同分析对象,筛选得到的光谱波长点呈现一定的差异性。对于生物量,两种算法筛选得到的波长点主要集中于 7500~10000 cm^{-1} 范围内,这些吸收峰主要与 C—H 的伸缩振动、 NH_2 的二倍频振动等基团相关;对于蛋白质和叶绿素,筛选得到的光谱主要集中于 5300~7200 cm^{-1} 范围内,主要与 NH 一级倍频和 N—H 的伸缩振动等基团相关^[27]。对比两种算法,RT 算

表 1 基于不同光谱预处理方法建立的偏最小二乘回归(PLSR)模型的评价结果

Table 1 Evaluation results of partial least squares regression (PLSR) models established with different spectral preprocessing

预测指标 Predictive index	预处理方法 Pretreatment methods	训练集 Training set			预测集 Prediction set		
		RMSECV	R_t^2	RPD	RMSEP	R_p^2	RPD
生物量 Biomass	RS	0.032	0.934	3.956	0.047	0.869	2.203
	1st	0.032	0.933	3.932	0.042	0.932	2.460
	CWT	0.032	0.933	3.932	0.044	0.926	2.312
	MSC	0.036	0.918	3.542	0.046	0.895	2.218
	SNV	0.035	0.920	3.586	0.056	0.860	1.834
	1st+MSC	0.035	0.922	3.627	0.064	0.705	1.611
	1st+SNV	0.035	0.921	3.612	0.064	0.705	1.613
	CWT+MSC	0.031	0.940	4.126	0.062	0.798	1.660
	CWT+SNV	0.033	0.931	3.847	0.063	0.778	1.624
蛋白质 Protein	RS	0.017	0.924	3.667	0.021	0.935	3.574
	1st	0.009	0.980	7.113	0.018	0.945	4.220
	CWT	0.013	0.956	4.813	0.030	0.852	2.519
	MSC	0.017	0.926	3.732	0.026	0.930	2.943
	SNV	0.017	0.924	3.682	0.027	0.931	2.888
	1st+MSC	0.009	0.980	7.145	0.036	0.763	2.103
	1st+SNV	0.009	0.980	7.130	0.036	0.763	2.104
	CWT+MSC	0.011	0.972	6.060	0.031	0.829	2.438
	CWT+SNV	0.011	0.970	5.882	0.033	0.811	2.341
叶绿素 Chlorophyll	RS	0.586	0.955	4.760	0.818	0.922	3.248
	1st	0.420	0.977	6.642	0.779	0.943	3.408
	CWT	0.414	0.977	6.730	0.811	0.928	3.275
	MSC	0.572	0.957	4.876	0.695	0.931	3.819
	SNV	0.593	0.954	4.707	0.720	0.931	3.691
	1st+MSC	0.565	0.958	4.941	1.690	0.653	1.571
	1st+SNV	0.565	0.958	4.938	1.689	0.653	1.572
	CWT+MSC	0.592	0.954	4.712	1.393	0.749	1.906
	CWT+SNV	0.596	0.953	4.683	1.388	0.751	1.913

注(Note): RS, 原始光谱(Raw spectrum); 1st, 一阶导数(1st derivative); CWT, 连续小波变换(Continuous wavelet transform); MSC, 多元散射校正(Multiplicative scatter correction); SNV, 标准正态变换(Standard normal variate); RMSECV, 交叉验证均方根误差(Root mean square error of cross validation); RMSEP, 预测均方根误差(Root mean square error of prediction); RPD, 相对性能偏差(Relative performance deviation); R_t^2 , 训练集决定系数(Coefficient of determination of the training set); R_p^2 , 预测集决定系数(Coefficient of determination of the prediction set)。

法筛选得到的光谱波长点数量少于 MC-UVE 算法筛选得到的光谱波长点,但二者在波长点的分布上具有一致性。

使用筛选后的波长点建立新的 PLSR 模型,利用 LOOCV 对小球藻训练集样品进行生物量、蛋白质及叶绿素含量的预测。计算出预测结果的 RMSECV、 R^2 和 RPD,用于评估模型性能,结果见表 2。比较分析表 1 与表 2 中的数据可知,利用筛选得到的波长点建立 PLSR 模型,预测能力进一步提升。特别是在对蛋白质含量与叶绿素含量的预测方面,此模型展现出更显著的预测效果。由表 1 可知,使用经过 1st 方法处理后的光谱建立的蛋白质 PLSR 定量分析模型预测结果的 RMSECV、 R_t^2 和 RPD 分别为 0.009, 0.980 和 7.113,而使用 MC-UVE 和 RT 算法筛选出 450 个和 102 个波长点后,使用筛选出的波长点建立新的 PLSR 模型(表 2),其预测结果的 RMSECV 分别下降至 0.006 和 0.007, R_t^2 分别提升至 0.992 和 0.989, RPD 分别提升至 11.061 和 9.717。类似地,小球藻叶绿素的原始光谱经 1st 处理后,新建立的 PLSR 模型的预测结果的 RMSECV 为 0.420, R_t^2 为 0.977, RPD 为 6.642; 原始光谱经过 1st 处理与 MC-UVE 波长筛选

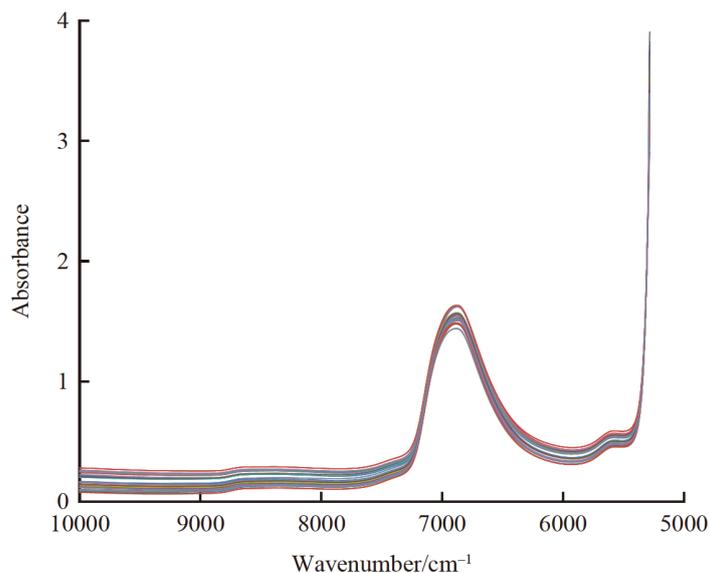


图1 小球藻的近红外光谱图

Fig.1 Near-infrared spectra of chlorella algal

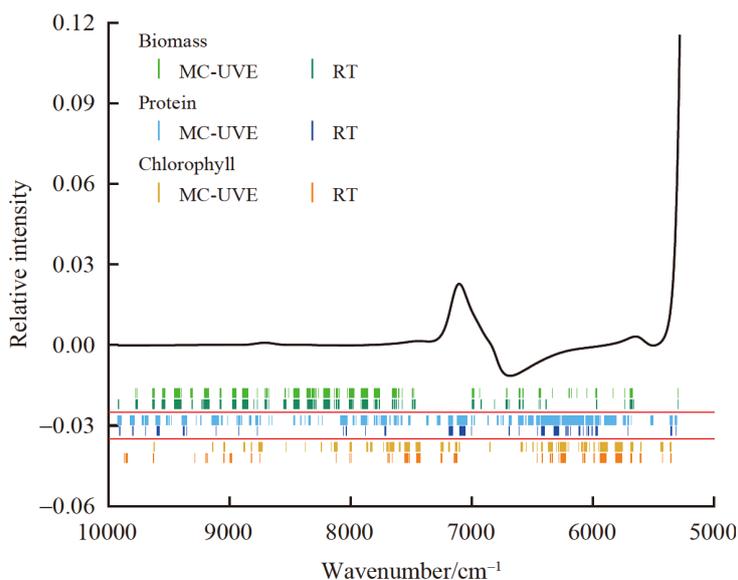


图2 不同光谱波长筛选结果

Fig.2 Screening results of different spectral wavelengths

MC-UVE,蒙特卡洛无信息变量剔除法(Monte Carlo-uninformative variable elimination); RT,随机测试法(Randomization test)

表 2 基于不同光谱预处理结合波长筛选建立的 PLSR 模型的预测结果

Table 2 Prediction results of PLSR models established by combining different spectral preprocessing with wavelength selection

预测指标 Predictive index	预处理方法 Processing method	波长点数 Wavelength point	训练集 Training set			预测集 Prediction set		
			RMSECV	R_1^2	RPD	RMSEP	R_p^2	RPD
生物量 Biomass	1st+MC-UVE	224	0.032	0.934	3.951	0.041	0.934	2.480
	1st+RT	222	0.032	0.934	3.935	0.041	0.933	2.476
蛋白质 Protein	1st+MC-UVE	450	0.006	0.992	11.061	0.015	0.963	5.270
	1st+RT	102	0.007	0.989	9.717	0.012	0.973	6.135
叶绿素 Chlorophyll	1st+MC-UVE	215	0.342	0.985	8.148	0.676	0.952	3.929
	1st+RT	124	0.328	0.986	8.512	0.517	0.962	5.138

后,对应光谱建立的 PLSR 模型的预测结果的 RMSECV 下降至 0.342, R^2 提升至 0.985, RPD 提升至 8.148; RS 经过 1st 处理与 RT 筛选后,对应光谱建立的 PLSR 模型的预测结果的 RMSECV 下降至 0.328, R^2 提升至 0.986, RPD 提升至 8.512。而对于生物量指标,波长选择仅起到精简模型的效果,对于预测能力的提升效果有限。综合来看,采用两种方法进行波长筛选处理后,对应光谱建立的 PLSR 模型的预测能力均得到提升,1st 与 RT 的组合略微优于 1st 与 MC-UVE 组合。

2.3 模型验证

为了评估多种 PLSR 模型在实际分析中的预测效果,采用预测集样本的光谱数据,并结合上述建立的 PLSR 模型预测样本中的生物量、蛋白质含量及叶绿素含量。将这些计算结果与通过标准方法获取的参考值进行对比分析,各模型的预测结果见表 1 和表 2。相较于直接使用 RS 建立的模型,在处理后的光谱数据基础上构建的 PLSR 模型的预测性能更优,此结果与训练集的结果呈现出相似性。进一步研究表明,采用波长筛选技术挑选出有效的波长子集后,模型的预测精确度再次得到了显著提升。以蛋白质指标为例(表 1),基于 RS 数据建立的预测模型的预测结果的 RMSEP、 R_p^2 和 RPD(预测)分别为 0.021、0.935 和 3.574; 而使用经 1st 处理后的光谱建立的 PLSR 模型的 RMSEP、 R_p^2 和 RPD(预测)分别为 0.018、0.945 和 4.220。由表 2 可知,采用 RT 方法从经过 1st 处理后的光谱数据中筛选出 102 个有效变量,以这些变量重新建立 PLSR 模型,其预测结果的 RMSEP、 R_p^2 和 RPD(预测)进一步提升至 0.012、0.973 和 6.135。类似地,采用 MC-UVE 方法从经 1st 处理后的光谱中筛选出 450 个有效变量,以其重新建立 PLSR 模型,其预测结果的 RMSEP、 R_p^2 和 RPD(预测)分别提升至 0.015、0.963 和 5.270。针对生物量和叶绿素含量,采用光谱预处理和波长筛选技术后,PLSR 模型的预测能力均得到了相应增强。这些经过优化的模型均表现出良好的预测性能。

图 3 以散点图形式呈现了采用最优方法组合(即 1st+RT 处理光谱)后所构建的 PLSR 模型针对训练

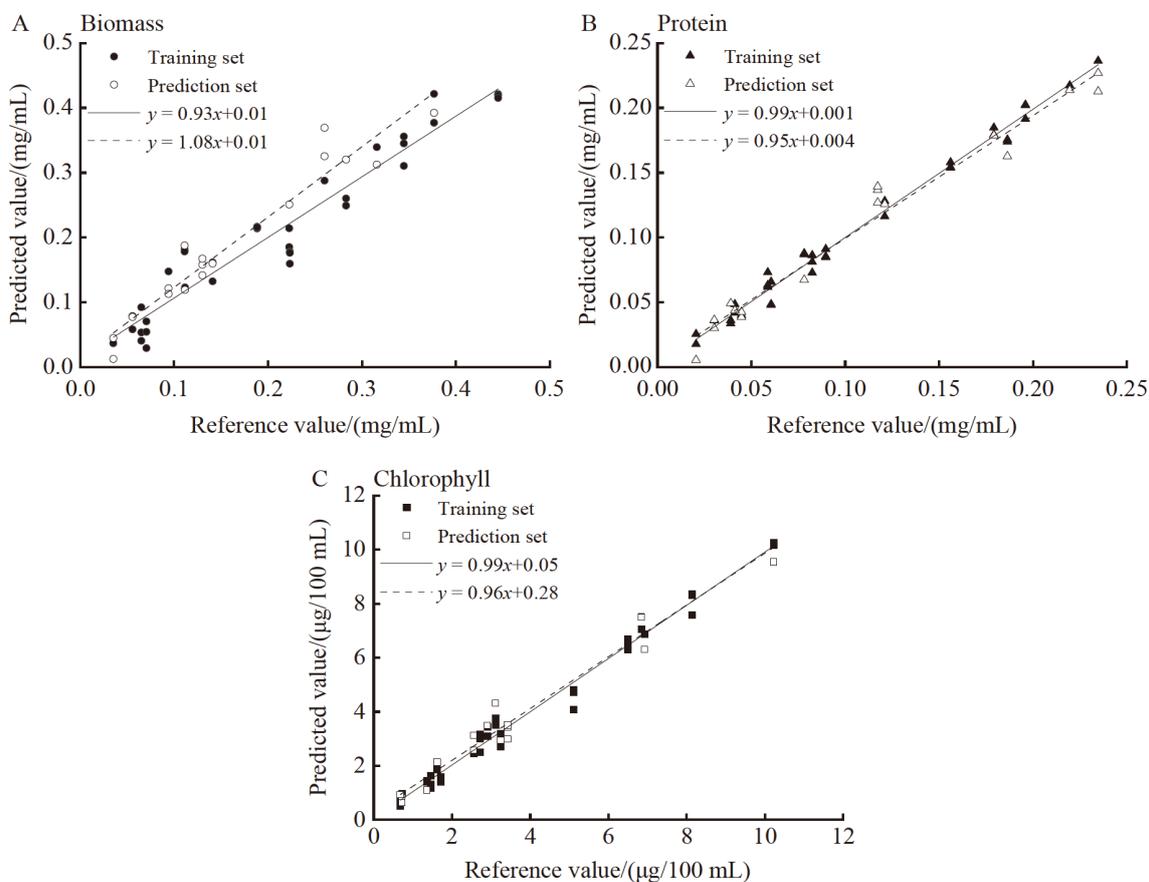


图3 模型预测值与标准值的相关性: (A)生物量拟合图; (B)蛋白质拟合图; (C)叶绿素拟合图

Fig.3 Correlation between model predicted values and reference values: (A) Fitting plot for biomass; (B) Fitting plot for protein; (C) Fitting plot for chlorophyll

集和预测集中小球藻样本的生物量、蛋白质含量以及叶绿素含量的预测结果。由图 3 可见,训练集(黑色)和预测集(白色)的样品分布均匀,通过 PLSR 模型计算得出的生物量、蛋白质含量及叶绿素含量与实验室标准理化方法检测结果之间的差异较小,并且这种差异与浓度水平的变化无明显关联。对训练集与预测集的样品分别进行了直线拟合分析,结果表明,3 个子图中对应的两条直线斜率均近似为 1。由图 3A 可见,由于个别样品的生物量指标预测误差较大,导致两条拟合曲线之间存在一定的偏差。然而,对于蛋白质和叶绿素含量,两条拟合曲线几乎完全重合(图 3B 和 3C),说明训练集与预测集样品之间的预测偏差较小。上述结果表明,经过优化的 PLSR 模型对实际样品具有良好的预测性能,完全满足小球藻样品中生物量、蛋白质及叶绿素含量的定量分析要求。

为了清晰直观地展示最终模型的预测性能,进一步采用第二批次样品作为验证集检验模型的预测效果。将模型预测的验证值与测得的参考值进行对比,进一步评估模型的性能。如图 4 所示,采用散点图的形式清晰地呈现了验证值与参考值的变化趋势。在图 4 中,每个样品的验证值(红色)与参考值(黑色)一一对应,而连续的线条直观地展示了二者随样本顺序的变化趋势。可以直观地看到验证值与参考值的接近程度较高,变化趋势一致,说明本模型能够较准确地预测小球藻叶绿素、蛋白质和生物量的实际含量,并且模型不易受样品批次的影响,泛化能力强,具有良好的实际应用价值。

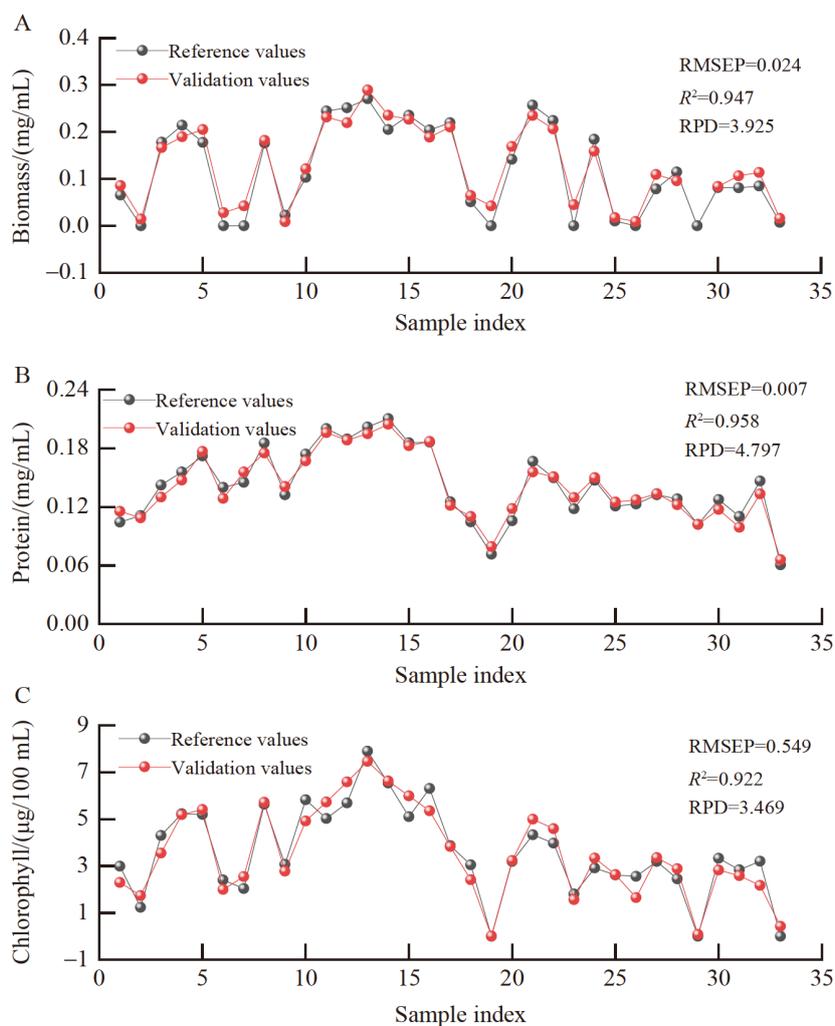


图 4 模型验证值与标准值的相关性: (A)生物量散点图; (B)蛋白质散点图; (C)叶绿素散点图

Fig.4 Correlation between model validation values and reference values: (A) Scatter plot of biomass; (B) Scatter plot of protein; (C) Scatter plot of chlorophyll

3 结论

考察了近红外光谱分析技术对小球藻生长过程中生物量、蛋白质含量和叶绿素含量 3 种关键质量指标进行快速检测的可能性。采用 PLSR 技术建立了小球藻生物量、蛋白质含量和叶绿素含量的定量分析模型。首先,采用 MSC、SNV、CWT 和 1st 等多种方式对样本 RS 进行预处理,初步提升模型的预测能力;然后,采用 MC-UVI 与 RT 算法筛选出有效波长,增强模型的预测效能;最后,采用此模型预测小球藻的生长指标,并与标准检测结果进行对比验证。结果表明,光谱经过 1st 预处理后,采用 RT 算法进行波长筛选,所构建的 PLSR 模型具有良好的预测效果。利用优化的模型,对小球藻样本中的生物量、蛋白质含量和叶绿素含量进行预测,预测结果与标准方法测得结果具有较好的一致性。本方法具有良好的实际应用价值,可快速便捷地获得小球藻综合营养含量信息,及时反馈生长过程信息,进而指导实际生产。

References

- [1] WU Zhi-Gang, ZHAO Lu. *Agric. Prod. Process.*, 2023, (2): 73-75, 80.
吴志刚, 赵璐. *农产品加工*, 2023, (2): 73-75, 80.
- [2] ZHAO Zhen-Yu, LIU Ping-Huai, WANG Sheng-Lin, NIU Qing. *Food Ind.*, 2017, 38(11): 275-278.
赵震宇, 刘平怀, 王盛林, 牛晴. *食品工业*, 2017, 38(11): 275-278.
- [3] CAMACHO F, MACEDO A, MALCATA F. *Mar. Drugs*, 2019, 17(6): 312.
- [4] KIM S W, XIAO M Z, SHIN H J. *Biotechnol. Bioprocess Eng.*, 2016, 21(6): 743-750.
- [5] HU Quan, WANG Zhi-Hong. *Guangdong Chem. Ind.*, 2015, 42(10): 152-153.
胡权, 王志红. *广东化工*, 2015, 42(10): 152-153.
- [6] KUMAR R, HEGDE A S, SHARMA K, PARMAR P, SRIVATSAN V. *Food Res. Int.*, 2022, 157: 111338.
- [7] SHI Chun-Zhe. *Mod. Food*, 2023, 29(24): 37-39.
石春哲. *现代食品*, 2023, 29(24): 37-39.
- [8] PASQUINI C. *Anal. Chim. Acta*, 2018, 1026: 8-36.
- [9] NG W, MALONE B P, MINASNY B. *Geoderma*, 2017, 289: 150-160.
- [10] WANG S Y, ZHANG J, CAI W S, SHAO X G. *Chin. Chem. Lett.*, 2019, 30(5): 1024-1026.
- [11] MUNCAN J, MATOVIC V, NIKOLIC S. *Talanta*, 2020, 206: 120253.
- [12] WANG L, SUN D W, PU H B, CHENG J H. *Crit. Rev. Food Sci. Nutr.*, 2017, 57(7): 1524-1538.
- [13] ZHENG S Y, WEI Z S, LI S, ZHANG S J, XIE C F, YAO D S, LIU D L. *Food Chem.*, 2020, 332: 127419.
- [14] HENN R, KIRCHLER C G, SCHIRMEISTER Z L, ROTH A, MÄNTELE W, HUCK C W. *J. Biophotonics*, 2018, 11(7): e201700365.
- [15] ZHANG Jin, HU Yun, ZHOU Luo-Xiong, LI Bo-Yan. *J. Instrum. Anal.*, 2020, 39(10): 1196-1203.
张进, 胡芸, 周罗雄, 李博岩. *分析测试学报*, 2020, 39(10): 1196-1203.
- [16] WANG Rong, ZHENG En-Rang, CHEN Bei. *J. Chin. Cereals Oils Assoc.*, 2023, 38(6): 141-148.
王蓉, 郑恩让, 陈蓓. *中国粮油学报*, 2023, 38(6): 141-148.
- [17] LI Rui, XIA Zhen-Zhen, WANG Chao, WANG Qiao, DUAN Shuo, LIU Yan. *J. Food Saf. Qual.*, 2022, 13(8): 2419-2425.
李锐, 夏珍珍, 王超, 王桥, 段烁, 刘言. *食品安全质量检测学报*, 2022, 13(8): 2419-2425.
- [18] ZHAN Ke, CHEN Ji-Wang, XU Yan, NI Yang-Fan, LIU Yan, ZOU Sheng-Bi. *Food Sci.*, 2024, 45(2): 299-307.
占可, 陈季旺, 徐言, 倪杨帆, 刘言, 邹圣碧. *食品科学*, 2024, 45(2): 299-307.
- [19] LIU Bo-Yang, GAO An-Ping, YANG Jian, GAO Yong-Liang, BAI Peng, TERI-GELE, MA Li-Jun, ZHAO San-Jun, LI Xue-Jing, ZHANG Hui-Ping, KANG Jun-Wei, LI Hui, WANG Hui, YANG Si, LI Chen-Xi, LIU Rong. *Spectrosc. Spectral Anal.*, 2023, 43(10): 3009-3014.
刘伯扬, 高安平, 杨骛, 高永亮, 白鹏, 特日格乐, 马利军, 赵三军, 李雪晶, 张慧萍, 康俊巍, 李慧, 王慧, 杨斯, 李晨曦, 刘蓉. *光谱学与光谱分析*, 2023, 43(10): 3009-3014.
- [20] TANG Yun-Feng, CHAI Qin-Qin, LIN Shuang-Jie, HUANG Jie, LI Yu-Rong, WANG Wu. *Spectrosc. Spectral Anal.*, 2020, 40(1): 202-208.
唐云峰, 柴琴琴, 林双杰, 黄捷, 李玉榕, 王武. *光谱学与光谱分析*, 2020, 40(1): 202-208.
- [21] GUO Wen-Chuan, ZHU De-Kuan, ZHANG Qian, DU Rong-Yu. *Trans. Chin. Soc. Agric. Mach.*, 2020, 51(9): 350-357.
郭文川, 朱德宽, 张乾, 杜荣宇. *农业机械学报*, 2020, 51(9): 350-357.
- [22] LIU Jing-Hua, CHEN Jun, QIN Song, QI Ze-Ming, HUANG Qing. *Spectrosc. Spectral Anal.*, 2019, 39(1): 79-86.
刘京华, 陈军, 秦松, 戚泽明, 黄青. *光谱学与光谱分析*, 2019, 39(1): 79-86.

- [23] NY/T 3082—2017. Determination of Chlorophyll Content in Fruits, Vegetables and Derived Products—Spectrophotometry Method. Agricultural Industry Standard of the People's Republic of China.
水果、蔬菜及其制品中叶绿素含量的测定 分光光度法. 中华人民共和国农业行业标准. NY/T 3082—2017.
- [24] WANG Jia-Jun, LI Zheng-Feng, WANG Luo-Ping, BIAN Xi-Hui, CAI Wen-Sheng, SHAO Xue-Guang. *Comput. Appl. Chem.*, 2015, 32(12): 1530-1536.
王家俊, 李正风, 王萝萍, 卞希慧, 蔡文生, 邵学广. *计算机与应用化学*, 2015, 32(12): 1530-1536.
- [25] ZHU Xue-Mei, TUO Xian-Guo, ZHANG Gui-Yu, ZHAI Shuang, LUO Lin, LUO Qi. *Mod. Food Sci. Technol.*, 2023, 39(1): 196-204.
朱雪梅, 庾先国, 张贵宇, 翟双, 罗林, 罗琪. *现代食品科技*, 2023, 39(1): 196-204.
- [26] GUO Tuo, XU Feng-Jie, MA Jin-Fang, XIAO Huan-Xian. *Spectrosc. Spectral Anal.*, 2024, 44(3): 737-743.
郭拓, 徐凤捷, 马晋芳, 肖环贤. *光谱学与光谱分析*, 2024, 44(3): 737-743.
- [27] JERRY W J, LOIS W. *Practical Guide to Interpretive Near-Infrared Spectroscopy*, Translated by CHU Xiao-Li, XU Yu-Peng, TIAN Gao-You. Beijing: Chemical Industry Press, 2009: 19, 69-76.
JERRY W J, LOIS W. *近红外光谱解析实用指南*. 褚小立, 许玉鹏, 田高友 译. 北京: 化学工业出版社, 2009: 19, 69-76.

Rapid Monitoring of Key Indicators in Growth Process of *Chlorella* Using Near-Infrared Spectroscopy Technology

SONG Wen-Hui, DU Shi-Jie, LIU Yan^{*}, WANG Qiao, LIU Xin, GONG Zhi-Yong
(College of Food Science and Engineering, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract The traditional detection methods for monitoring the biomass, protein, chlorophyll content and other key indicators in the growth of *Chlorella* have some problems, including complicated operation, slow detection speed and difficult large-scale application. In this study, a fast and efficient monitoring method for the key indicators in the growth of *Chlorella* was established using near infrared spectroscopy and chemometrics. Near-infrared spectroscopy was used to collect near-infrared spectra of *Chlorella* algal fluid at different growth stages, and standard methods were used to detect the biomass, protein and chlorophyll contents of corresponding samples. A quantitative analysis model was established based on partial least squares regression (PLSR). To improve the prediction ability of the model, multiplicative scatter correction (MSC) was used to reduce the interference of scattering on the raw spectrum (RS), standard normal variate (SNV) was used to normalize the original spectral data to eliminate differences between samples, continuous wavelet transform (CWT) was used to obtain the key features of spectral data, the first derivative (1st) was used to enhance the differentiation of the original spectral features, and monte carlo-uninformative variable elimination (MC-UVE) and randomization test (RT) were used to screen the valid variables in the wavelength. By evaluating the prediction ability of different models, the quantitative analysis models of *Chlorella* biomass, protein and chlorophyll content were finally determined. The results showed that the model based on 1st combined with RT spectra had better predictive ability for *Chlorella* nutrient content detection, and the root mean square errors of prediction (RMSEP) and coefficients of determination (R^2) were 0.041 and 0.933 for biomass, 0.012 and 0.973 for protein, and 0.517 and 0.962 for chlorophyll, respectively. This model showed practical application value, and could realize the rapid and accurate detection of *Chlorella* biomass, protein and chlorophyll content at the same time.

Keywords Near infrared spectroscopy; *Chlorella*; Spectral pretreatment; Wavelength screening; Partial least squares regression

(Received 2024-12-25; accepted 2025-02-19)

Supported by the National Key Research and Development Program of China for Young Scientists (No. 2022YFF1102500).