

基于近红外超连续激光光谱的水稻种子 活力无损分级检测研究

金文玲^{1,2}, 曹乃亮¹, 朱明东³, 陈伟¹, 张佩光¹, 赵庆磊¹, 梁静秋¹, 余应弘⁴, 吕金光^{1*}, 阚瑞峰^{1*}

(1. 中国科学院长春光学精密机械与物理研究所应用光学国家重点实验室, 吉林长春 130033;

2. 中国科学院大学, 北京 100049;

3. 湖南省水稻研究所, 湖南长沙 410125;

4. 湖南省农业科学院, 湖南长沙 410125)

摘要:针对目前农业种植选种应用对于带稃壳水稻种子活力分级检测的迫切需求,以及现有通用的糙米检测技术存在的问题,本文提出一种基于近红外超连续激光光谱的水稻种子活力透射光谱检测方法。首先,设计了种子活力近红外吸收光谱检测系统,测量了3种不同年份的带稃壳的水稻种子的近红外吸收光谱,结果显示,水稻种子的活力梯度与近红外吸收光谱的特征吸收峰值相关。然后,采用归一化、二阶导数校正法和正交信号校正相结合优化了种子光谱的预处理算法。最后,建立主成分分析(PCA)模型,对光谱进行降维,确定最佳主成分数目,应用偏最小二乘判别分析(PLS-DA)建立了水稻种子活力分析鉴别模型。分析结果表明,本文设计的透射式吸收光谱检测系统结合PLS-DA判别模型可对不同活力的水稻种子进行分类,校正集和验证集的准确率分别为94.44%和95.92%,筛选后水稻种子的发芽率可达97.17%。研究表明,本文提出的基于近红外光谱信息实现水稻种子活力无损分级的方法可行,且具有较高的预测精度。

关键词:近红外光谱;种子活力;偏最小二乘判别分析;主成分分析

中图分类号:O433.1

文献标志码:A

doi:10.37188/CO.2020-0027

收稿日期:2020-02-24; 修订日期:2020-03-25

基金项目:湖南省农业科技创新基金项目(No. 2018NK1020);国家自然科学基金(No. 61627819, No. 61805239, No. 61727818);吉林省科技发展计划(No. 20190303063SF, No. 20180201024GX, No. 20150520101JH);中国科学院青年创新促进会基金(No. 2018254)

Supported by Hunan Agricultural Science and Technology Innovation Fund Project (No. 2018NK1020); National Natural Science Foundation of China (No. 61627819, No.61805239, No. 61727818); Jilin Province Science and Technology Development Plan (No. 20190303063SF, No. 20180201024GX, No.20150520101JH); Foundation of Youth Innovation Promotion Association of Chinese Academy of Sciences (No. 2018254)

Nondestructive grading test of rice seed activity using near infrared super-continuum laser spectrum

JIN Wen-ling^{1,2}, CAO Nai-liang¹, ZHU Ming-dong³, CHEN Wei¹, ZHANG Pei-guang¹, ZHAO Qing-lei¹, LIANG Jing-qiu¹, YU Ying-hong⁴, LV Jin-guang^{1*}, KAN Rui-feng^{1*}

(1. State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. Hunan Rice Research Institute, Changsha 410125, China;

4. Hunan Academy of Agricultural Science, Changsha 410125, China)

* Corresponding author, E-mail: jinguanglv@163.com; rfkan@ciomp.ac.cn

Abstract: In view of the urgent need for seed selection technology in agriculture and for grading detection of the vigor of three different years of unpeeled rice seeds, we proposed a new method of detecting the vigor of rice seeds based on near-infrared super-continuous laser spectrum to overcome the significant issues in pre-existing universal brown rice detection technology. Firstly, we design a near-infrared absorption spectroscopy system with which we detect seed viability and measure the NIR spectra of three different years of unpeeled rice seeds. The results show that the activity gradient of the rice seeds is correlated with the characteristic absorption peak of their NIR absorption spectrum. Then, the spectrum of seed is optimized with a pre-treatment algorithm of normalization, second derivative correction and orthogonal signal correction. Finally, a Principal Component Analysis (PCA) model is established to reduce the dimension of the spectrum and determine the optimal number of principal components. A Partial Least Squares Discriminant Analysis (PLS-DA) model is established. The analysis results show that the transmission absorption spectrum detection system designed in this paper combined with the PLS-DA discrimination model can classify rice seeds of different vigor with an accuracy of 94.44% and 95.92%. After screening, the germination rate of rice seeds can reach 97.17%. The results show that it is feasible to achieve non-destructive classification of rice seed activity using near-infrared spectroscopy with high accuracy.

Key words: near-infrared spectroscopy; seed vigor; partial least-squares discriminant analysis; principal component analysis

1 引 言

近红外光谱 (Near Infrared Spectroscopy, NIRS) 分析作为一种先进的快速无损检测手段, 通过测量物质发射、吸收或散射光的波长和强度, 可以对物质进行定性和定量分析。自 20 世纪 80 年代后期以来, 近红外光谱分析技术已经在

矿物分析^[1]、农业食品^[2]及石油化工^[3]等领域取得了许多令人瞩目的应用成就。

水稻是我国第一大粮食作物, 占粮食总产量的 42%^[4], 是我国农业领域在国际市场唯一具有较强优势的产业。然而, 水稻产业一直面临着种子质量不高、库存积压严重等问题, 而水稻种子的优劣又直接影响水稻的产量与质量。在我国, 每年因为作物种子选育不当造成的农产品减产比

例在 10%~20% 之间^[5]。因此,水稻种子的活力检测至关重要。另一方面,通过在积压的种子中筛选高活力种子,可以大大降低经济损失。此外,由于种子活力的高遗传性,选用高活力的水稻种子进行杂交可以从遗传层面对种子的活力进行改良,进而选育高活力的新品种^[6-7]。因此,对水稻种子活力的高精度检测需求十分迫切。

传统检测方法主要包括感官与物理检测、生理检测及化学检测等,这些方法通过测量种子与苗的外形尺寸或电导率等参数进行活力判断,往往需要投入大量的人力物力,且检测周期长,主观性强,容易对种子造成损伤,难以满足现代农业生产对自动化、高通量种子活力检测的需求。近年来,国际上提出了种子活力的光谱学检测方法,利用种子内部物质对光谱的吸收表征其含量变化。该方法无需对水稻种子进行预处理,经过筛选的种子样品仍可用于生产或继续储藏。迄今为止,国内外研究人员已经取得了许多研究进展。例如, Ambrose 等^[2]采用拉曼光谱检测和傅立叶红外光谱检测,通过主成分分析法和偏最小二乘判别分析法对玉米种子进行活力判定,预测准确率达 100%;朱银等^[8]利用偏最小二乘算法结合傅立叶变换近红外光谱分析技术,对小麦种子的发芽率进行判定,得到的小麦种子发芽率的预测值和标准值有很高的相关性。在水稻检测方面,李欢欢等^[9]利用声光光谱技术结合最小二乘支持向量机回归(LS-SVR)对糙米活力进行测定,为便携式水稻活力光声光谱仪的研制提供了理论依据;李美凌等^[10]通过高光谱成像仪采集了 240 粒不带壳水稻种子的光谱图像,采用支持向量机建立水稻种子活力鉴别模型,区分不同活力的种子;宋乐等^[11]采用近红外光谱分析技术,通过漫反射积分球采集 4 种水稻种子光谱,并对光谱建立主成分分析模型,成功区分出种子是否具有发芽潜力。国内外已进行的选种分级方面的研究具有以下特点:首先,国外大部分活力检测分析的研究对象都是小麦玉米等谷物的种子,鲜有水稻种子活力分析的报道,而国内大部分活力分析模型的研究对象都是针对不带稃壳的糙米种子;其次,针对水稻种子活力的鉴别一般将种子分成发芽与未发芽两

类,没有对具有活力的种子进行细化分级。近年来,农业种植技术的快速发展对于大规模种子活力的精细判断提出了新的要求,迫切需要对大量带有稃壳的水稻种子进行分级检测,以避免因种子内部器官受到病虫等的损伤而影响发芽率。同时需要对种子活力进行级别精细化检测,加强对幼苗强壮程度的预判,以适用于不同的种植需求,从而进一步提高水稻产量。

本论文基于近红外光谱检测技术,分析了带有稃壳的水稻种子的吸收光谱特性,通过化学计量学算法建立了活力分析模型,完成了水稻种子的活力分级检测。

2 水稻种子活力检测原理与检测系统设计

2.1 水稻种子活力近红外光谱检测原理

物质中的不同组分或同一组分的不同基团在近红外(NIR)光谱区域具有丰富的光谱特征,因此,近红外光谱能反映物质内部成分的结构和组成信息。水稻种子在储藏的过程中,其内部的物质成分会发生变化,导致生物大分子的含量也随之变化,进而引起活力的变化^[10],这种变化可以通过近红外光谱检测中物质对光的能量吸收特性进行检测和表征。水稻种子的蛋白质、淀粉、脂肪、纤维素等有机化合物中均具有丰富的含氢基团,如 C-H(脂肪烃)、C-H(芳香烃)、C=O(羧基)、O-H(羟基)及 N-H(氨基)等,这些含氢基团在近红外光谱区具有较强的吸收。水稻种子中的有机分子在近红外区的光谱吸收特征因其所含基团的种类和振动模式不同,导致吸收波长与强度都有明显差别。因此,利用水稻种子的近红外吸收光谱,可以对种子中的物质组成与含量进行表征,从而对种子的活力进行鉴定。

当水稻种子中的有机分子受到近红外区域的电磁辐射时,分子中基团原子间振动的非谐性使分子振动从基态向高能级跃迁,从而选择性吸收对应频率的光。而近红外光谱主要反映含氢基团基频振动的倍频和组合频的吸收情况,有机分子在近红外区的这种特有的吸收模式,可以表达

被测种子的化学成分信息。基团效应是近红外光谱中最主要的效应,文献[12]中示出了分子中各种基团在近红外光谱区的倍频与组合频光谱分布。同时,由于光谱的能量与波长成反比,相较于中长波红外波段,近红外光更容易透过稃壳,获得种子内部的光谱信息^[13-14]。因此,采用近红外光谱检测方法可以有效获取种子的活力信息。

2.2 活力光谱检测系统设计

水稻种子由稃壳和糙米两部分组成,在物理角度,稃壳和糙米的颜色、密度和形态均有不同,在化学角度,稃壳和糙米的蛋白质、脂肪、淀粉、纤维素等化学物质的组成也不同,而对水稻种子活力起主要作用的物质是种子内部糙米中的化学成分,因此在种子活力检测中要尽量增强糙米的信息而减少稃壳对测量的影响^[15]。近红外光谱检测一般采用反射光谱法或透射光谱法^[16-17]。对带稃壳的水稻种子进行活力测量的反射法与透射法分别如图1(彩图见期刊电子版)所示,其中红色光线表示稃壳反射的光源的光,蓝色光线表示在种子内部经多次反射后携带糙米信息的光。对于反射光谱测量法而言,入射光只有部分光入射到种子内部,其携带种子内部信息后透射出种子表面,其余的光则被稃壳反射,两部分光一起由探测器接收。而采用透射光谱测量时,光源发出的光垂直入射到种子中,除少部分入射光被稃壳反射外,其余光线均入射到种子内部。入射到种子内部的光,一部分携带种子内部信息以一定发散角出射,另一部分被内壁光滑的稃壳反射后透射出种子,最终都被探测器接收。光线图表明透射光谱法不仅可以得到整个种子内部糙米的组成物质信息,同时可以避免表面种壳的反射光和外部光线的干扰,而且可以消除由于种子内部蛋白质、脂肪、淀粉、纤维素等成分位于不同部位而造成的测定误差,十分适用于完整水稻种子活力的检测。对于水稻稃壳来说,其密度较低且厚度较薄,所以光程较短,吸光度比较小。同时,由于内壁光滑的稃壳包裹着糙米,测量时,相当于在糙米周围引入了反射镜,光进入种子内部后,经过种壳的多次反射,光程增加几倍,使得透射光谱中糙米的吸光度得到显著提高^[18],实现吸收光谱信息的增强,通过精确测定水稻种子糙米中蛋白质、淀粉、脂

肪、纤维素等成分信息,从而获取其活力梯度,所以更适合采用透射光谱法进行测量。

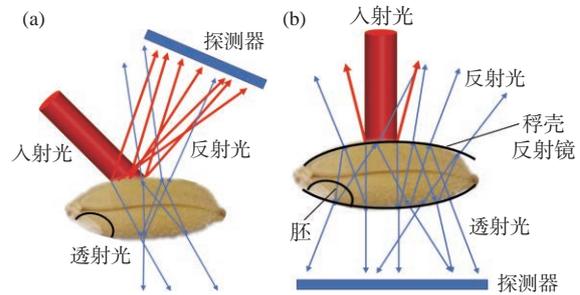


图1 (a) 反射式光谱检测方法 (b) 透射式光谱检测方法测量原理示意图

Fig. 1 Schematic diagram of (a) reflection type spectral detection method, and (b) transmission type spectral detection method

基于以上分析,本文设计了水稻种子活力的透射光谱检测系统。该系统由超连续激光光源,准直镜,孔径光阑,聚焦镜,石英光纤及光栅光谱仪几部分组成,其中聚焦镜包括两块材料分别为熔融石英(Fused Silica)和氟化钙(CaF₂)的双透镜组。其工作原理如图2(彩图见期刊电子版)所示。

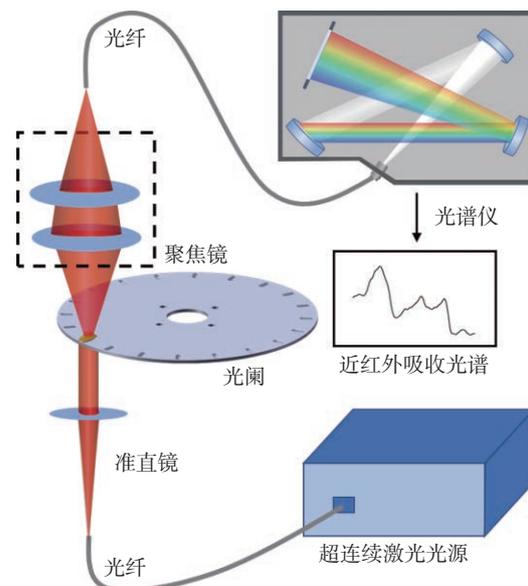


图2 水稻种子活力的透射光谱检测系统

Fig. 2 Transmission spectrum detection system for rice seed vigor

采用具有宽波段高能量密度的超连续激光光源(NKT SuperK Compact)作为入射光,相较于传统近红外光谱检测中常用的卤钨灯光源,其具有

更均匀的输出光谱,在保证光源谱带足够宽的同时,光功率输出比卤钨灯高出 10^3 数量级,能够实现带有稃壳水稻种子宽波段高信噪比的光谱测量^[19]。为了防止光源的入射光直接进入光谱仪,在光源和种子之间设置尺寸小于水稻种子的孔径光阑。同时,为了避免光阑对种子内部透射光的阻挡,光源采用自下而上照射,依次经过光阑和种子。入射到种子中的光在种子内部经过多次吸收反射后,便携带了种子内部的组成和结构信息,并以一定发散角出射,经过聚焦镜被耦合到光纤中,最终由光谱仪 (NIR Quest 512-2.5) 采集。根据光谱仪中光栅的色散,在阵列探测器上获得不同波长处的吸收光谱信息,由吸收光谱曲线便可以对水稻种子的活力进行表征。

3 水稻种子光谱测量与分析

用图 2 的光谱检测系统分别测量 2016、2017、2018 三种年份及三种年份随机混合的 4 组不同活力的日本晴水稻种子在 900-2500 nm 波段的透射光谱,然后将所有种子放入垂直发芽板,在光照培养箱中进行标准发芽实验^[20],根据第七天的发芽情况与表型数据作为活力判定依据,将高活力与低活力种子占有被测种子的百分比记为种子原始发芽率,得到其活力表征及原始发芽率如表 1 所示。

表 1 筛选前水稻种子的活力情况

Tab. 1 The seed vigor parameters of rice seeds before selecting

年份	活力高	活力低	不发芽	发芽率
2018	192	52	44	84.72%
2017	154	71	63	78.13%
2016	106	93	89	69.09%
随机混合	133	84	71	75.351%
总计	585	300	267	—

由于光谱仪测量波段两端的仪器噪声较大,故选取 1100~2100 nm 波段光谱数据进行特性分析,根据发芽结果对不同活力水平种子的平均光谱进行对比,如图 3(a) 所示。

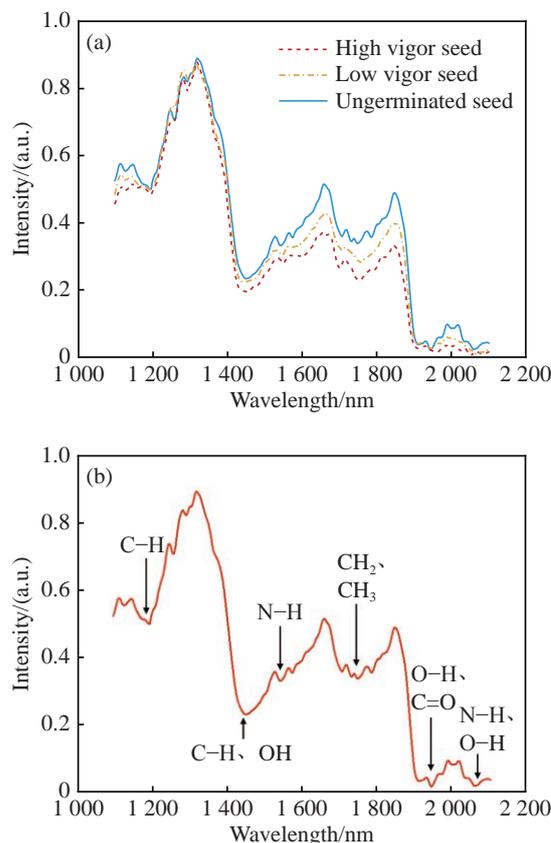


图 3 水稻种子的近红外吸收光谱图。(a)不同活力的水稻种子近红外光谱平均曲线;(b)高活力水稻种子近红外吸收光谱

Fig. 3 Near-infrared absorption spectrum of rice seeds. (a) The average curves of transmission spectra of rice seeds with different vigors. (b) Near-infrared absorption spectrum of high vigor seeds

从水稻种子的近红外吸收光谱图 3(a) 中可以看出,不同活力水平的水稻种子吸收光谱整体比较相似,主要在 1150~1220 nm、1410~1450 nm、1510~1540 nm、1660~1800 nm、1910~1950 nm、1980~2100 nm 等波段处有明显的吸收波谷,这是由于水稻中蛋白质、淀粉和纤维素等主要有机物中的氨基 N-H、羟基 O-H 和羰基 C=O 等基团的倍频及组合频吸收^[21]。图 3(b) 是高活力水稻种子的近红外吸收光谱,在 1150~1220 nm 处的波谷是甲基 CH_3 与亚甲基 CH_2 中 C-H 二级倍频和组合频振动吸收,在 1410~1450 nm 处的波谷是甲基 CH_3 与亚甲基 CH_2 中 C-H 及蛋白质中的羟基 O-H 的组合频振动吸收,在 1510~1540 nm 处的波谷是氨基酸和蛋白质等物质中所含的氨基

N-H 的一级倍频振动吸收,在 1660~1800 nm 处的波谷是脂肪族 CH_3 及 CH_2 的一级倍频吸收,在 1910~1950 nm 处的波谷是脂类、纤维素等多糖物质所含的 C=O 二级倍频吸收振动及淀粉与纤维素中 O-H 伸缩和 O-H 变形振动的组合频吸收,在 2100 nm 处的波谷则对应于 N-H 的组合频、二级倍频吸收及淀粉中 O-H 的组合频吸收^[22-23]。这些吸收谱带的强度分别代表了蛋白质和碳水化合物的含量。由水稻种子近红外光谱的吸收谱带特征可以看出,与水稻种子活力相关的蛋白质、淀粉和纤维素等有机化合物含量的变化可以由吸收谱带的光谱特性表征出来。从图 3(a) 可以看出,虽然不同活力的水稻种子的近红外吸收光谱特征峰具有一定差异,但其谱峰较宽,并且各种成分的吸收峰谱带重叠严重,难以直接从图谱中判别活力大小。针对这一问题,本文将通过化学计量学方法从光谱中提取相关信息进行定性分析。

4 光谱处理与数据分析

4.1 光谱信号校正

在红外光谱测量过程中,光的散射、样品的粒度及仪器噪声等因素会对近红外光谱仪器所采集的光谱产生影响,从而使采集的光谱除包含样品自身成分信息外还包含其他无关信息和噪声。为了消除系统或环境的影响,在建立判别模型之前需要对原始光谱进行一定的预处理^[24-25]。由于种子形貌、颜色等的差异会对光谱基线产生影响,从图 3 的水稻种子的近红外吸收光谱图也可以看出光谱存在一定的基线漂移,这会对多元分析校正产生非常不利的影响。测量的水稻种子吸收光谱因内部各组分吸收带间的相互干扰会有一些的谱带重叠,使用差分法可以消除这种干扰。为了初步对测得的所有种子的光谱数据的光谱线型、谱线吸收峰及光谱展宽等性质进行对比和观察,需对光谱曲线进行一阶导数 (FD)、二阶导数 (SD) 等处理,以消除基线漂移的影响。此外,还需利用多元散射校正 (MSC)、标准正态变量校正 (SNV) 及正交信号校正 (OSC) 等方法进行处理,来减小谱线重叠的影响^[26-27]。经过二阶导数预处

理的光谱如图 4 所示,可以看出基线漂移基本已被消除掉。

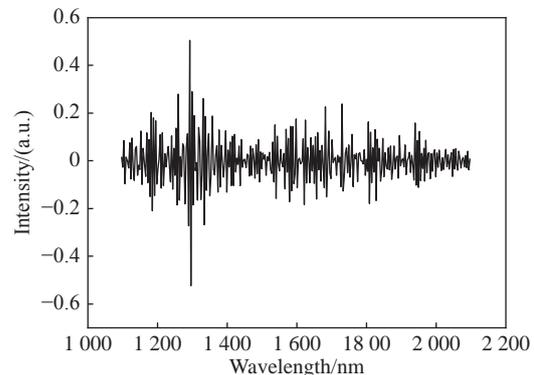


图 4 二阶导数处理后的光谱信息

Fig. 4 The absorption spectrum of rice seed after second derivative processing

由图 4 可知,经过二阶导数预处理之后,光谱的基线漂移问题得到了有效改善,但谱线峰值并不明显。因此,在预处理之后,还需要对数据进行信号校正以简化模型、提高光谱信噪比。正交信号校正 (OSC) 可以从光谱矩阵中去除与目标预测矩阵完全不相关的成分;多元散射校正 (MSC) 可以减少粒子分布不均和大小不一对散射产生的影响;标准正态变量校正 (SNV) 则可以消除表面散射、固体颗粒大小及光程变化对光谱的影响。图 5(a) 与图 5(b) 分别为对经过预处理的图像使用正交信号校正 (OSC)、标准正态变量校正 (SNV) 后的光谱图。

可以看出,对于该种类型的光谱,正交信号校正后的光谱较标准正态变量校正后的光谱谱线特征更明显。虽然对光谱进行校正能以较低的校正误差或简单的校正模型来描述待测目标,但由于不同类型的光谱特性不同,没有通用的校正方法,需要对几种预处理方法进行组合,筛选出最适合的模型^[28]。采用准确率来判断模型效果,判别结果的准确率表示模型能够正确筛选出的高活力、低活力及不发芽水稻种子总数占被测种子数量的百分率。几种预处理方法的建模筛选结果如表 2 所示。由表 2 可知,在 1100~2100 nm 光谱范围内,在主成分数为 3 时,选用归一化+SD+OSC 的预处理方法对样品进行活力判别时,模型的准确率最高,可以用该方法对本类型光谱建立判别分析模型。

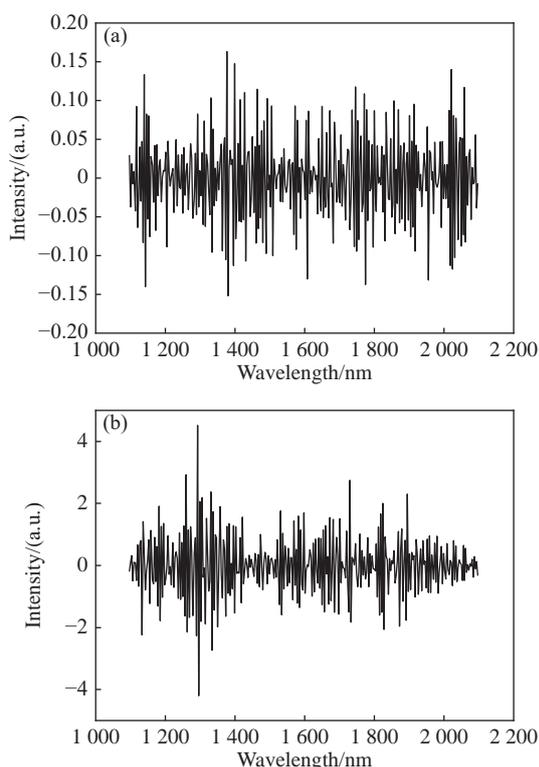


图 5 两种预处理方法结果对比。(a)正交信号校正处理后的光谱;(b)标准正态变量校正后的光谱

Fig. 5 Comparison of absorption spectrum of rice seed processed by two pretreatment methods. (a) Orthogonal signal correction; (b) standard normal variate correction

表 2 不同预处理方法对样品的活力鉴别情况

Tab. 2 The vitality identification results of samples by different pretreatment methods

预处理方法	光谱范围/nm	主成分数	准确率/%
未处理	1100 ~ 2100	5	68.19
MSC	1100 ~ 2100	3	75.63
SNV	1100 ~ 2100	3	71.78
OSC	1100 ~ 2100	3	77.05
归一化+MSC	1100 ~ 2100	3	70.33
SD+MSC	1100 ~ 2100	3	77.91
SD+SNV	1100 ~ 2100	3	79.85
SD+OSC	1100 ~ 2100	3	78.57
归一化+SD+MSC	1100 ~ 2100	3	83.97
归一化+SD+SNV	1100 ~ 2100	3	88.06
归一化+SD+OSC	1100 ~ 2100	3	94.13
归一化+SD+OSC	1100 ~ 2100	2	82.85
归一化+FD+OSC	1100 ~ 2100	3	89.39

4.2 主成分分析

选取最优的预处理方法对光谱数据进行处理后,还要进行建模分析,如果变量太多会增加计算的复杂性,更容易引入与类别无关的变量从而使模型失真。水稻种子中很多物质成分含量间具有一定相关性,当一个变量可以投影到其它变量上时,可以解释为光谱信息有一定的重叠。主成分分析(PCA)可以从光谱数据中提取出所有变量,将具有线性相关的变量删掉,其可以在减少需要分析指标数量的同时避免损失主要光谱信息,从而尽可能多地保留反应种子物质信息含量光谱的原有信息。

主成分分析将原始光谱矩阵 $X(n \times m)$ 分解为得分矩阵 T 和载荷矩阵 L 的乘积,矩阵 T 与 L 中的列分别为得分向量和载荷向量,即:

$$X = t_1 l_1^T + t_2 l_2^T + t_3 l_3^T + \dots + t_n l_n^T, \quad (1)$$

其中, t 为得分向量, l 为载荷向量,也称为主成分,各得分向量之间及各载荷向量之间是相互正交的。当光谱矩阵 X 中的向量线性相关时可以将 X 中的 n 维特征投影到 f 维上 ($f < n$), X 的变化将主要体现在前几个主成分上,在后几个载荷上的投影很小^[29]。由于绝大部分方差可由前几个主成分加以解释,因此,可以将矩阵 X 重建为:

$$X = t_1 l_1^T + t_2 l_2^T + t_3 l_3^T + \dots + t_f l_f^T + E, \quad (2)$$

式中, E 表示误差矩阵,由于引起误差矩阵 E 的主要因素是测量噪声,因此将 E 忽略掉不会引起数据中大量信息缺失,同时还起到消除噪声的作用。对矩阵 X 进行主成分分析等效于对得分矩阵 T 的协方差矩阵 $T^T T$ 进行特征向量分析,如果将 $T^T T$ 的特征值做如下排列:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m, \quad (3)$$

则这些特征值对应的特征向量 l_1, l_2, \dots, l_m , 即为矩阵 X 的主成分。前 f 个主成分的累积贡献率定义为矩阵 T 的协方差矩阵的前 f 个特征值的和 $\sum_{i=1}^f \lambda_i$ 除以它的所有特征值 $\sum_{i=1}^{\min(n,m)} \lambda_i$ 之和,累积贡献率表示提取的载荷矩阵能够代表的原始变量信息的百分比。

选用 4.1 中的最优光谱预处理方法对水稻种子光谱进行预处理后,用 PCA 提取数据主成分,不同主成分数对原始水稻种子光谱的解释情况如

表3所示。表3表明,随着主成分数的增加,模型累计贡献率逐渐增大,但单个主成分的贡献率越来越小,表明其对原始数据信息表示量的贡献越来越小,可以看到当主成分数为3时,对模型已经有了很好的解释能力,累计贡献率达到93.5%,主成分数继续增加时,一些与参数无关的信息被引入,模型出现过拟合,致使预测结果降低。

表3 主成分数与模型贡献率

Tab.3 Number of principal components and model contribution rate

主成分数	1	2	3	4	5	6
模型准确率/%	55.3	82.9	95.9	94.2	94.1	92.7
累积贡献率/%	62.4	85.7	93.5	96.1	98.2	99.6

4.3 基于水稻种子红外光谱信息的活力判别分析

4.3.1 PLS-DA 判别分析

对种子光谱数据进行预处理并选取了最优主成分数之后,采用偏最小二乘判别分析(PLS-DA)法对水稻种子进行活力分类。偏最小二乘判别分析根据种子活力梯度的类别进行数值变量标定,结合种子近红外光谱特征采用偏最小二乘法(PLS)建立校正模型,对种子活力进行判别。

PLS-DA方法是基于PLS回归方法进行种子分类和识别的,将种子的光谱数据看作自变量 X ,其行为为种子样本序号,其列为光谱变量序号,将种子的活力类别作为因变量 Y ,其行对应种子样本序号, Y 的数值对应不同类别。 Y 中元素值为0对应于不发芽的种子, Y 中元素值为1对应于低活力种子, Y 中元素值为2对应于高活力种子。

PLS-DA法利用PLS算法将种子光谱矩阵 X 和活力的类别矩阵 Y 分别进行分解,在对 X 矩

阵进行分解得到得分矩阵 T 的同时,对类别矩阵 Y 矩阵进行正交分解:

$$X = TP^T + E, \quad (4)$$

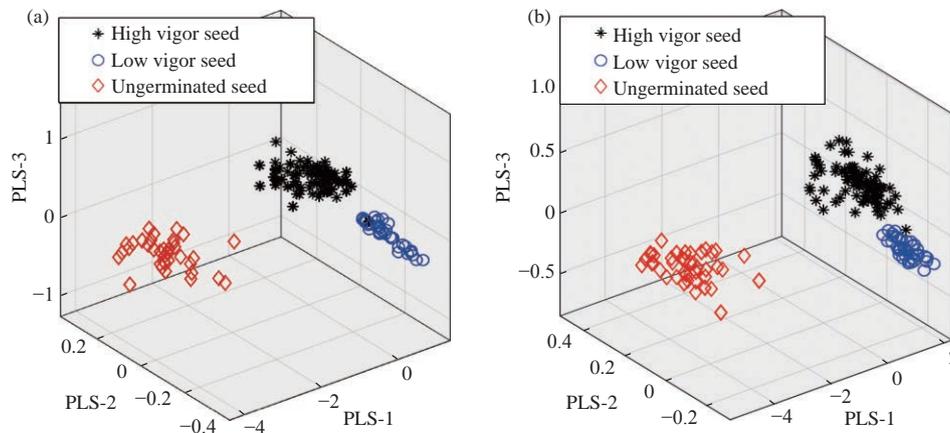
$$Y = UQ^T + F, \quad (5)$$

其中, T 和 U 分别为 X 与 Y 的得分矩阵, P 和 Q 分别为 X 、 Y 的载荷矩阵, E 和 F 为误差项的矩阵,从而找到多元空间的变量和方向,判别校正集类别。在预测时,首先根据载荷矩阵 P 求出未知样品光谱矩阵 X_{test} 的得分,然后,计算出回归的类别信息矩阵 Y_{test} ,根据 Y_{test} 与0、1、2的接近程度确定种子样品属于哪一类,从而实现对种子活力类别的判别。

PLS-DA可同时提取光谱矩阵 X 和类别矩阵 Y 的特征向量,从而加强了光谱分解时种子类别信息的作用,故可以提取出与样本活力梯度最相关的光谱信息,从而使提取的不同类别光谱之间的差异最大化。作为一种监督学习的判别分类模型,PLS-DA数据集的每个样本都是有类别输出的,通过投影分别将预测变量和观测变量映射到一个低维度的空间,在剔除与类标签无关的信息后,使每一类别数据的投影点尽可能聚集,不同类数据间的方差尽可能大,得到线性回归模型。

4.3.2 PLS-DA 判别分析结果

将全部1152份样本按3:1分为建模集和验证集。所有的样本光谱经过归一化+SD+OSC预处理后,在1100-2100 nm光谱范围内建立偏最小二乘判别分析模型,在特征波段上分析水稻种子活力情况,2016、2017、2018及混合年份的水稻种子活力分级结果如图6所示。



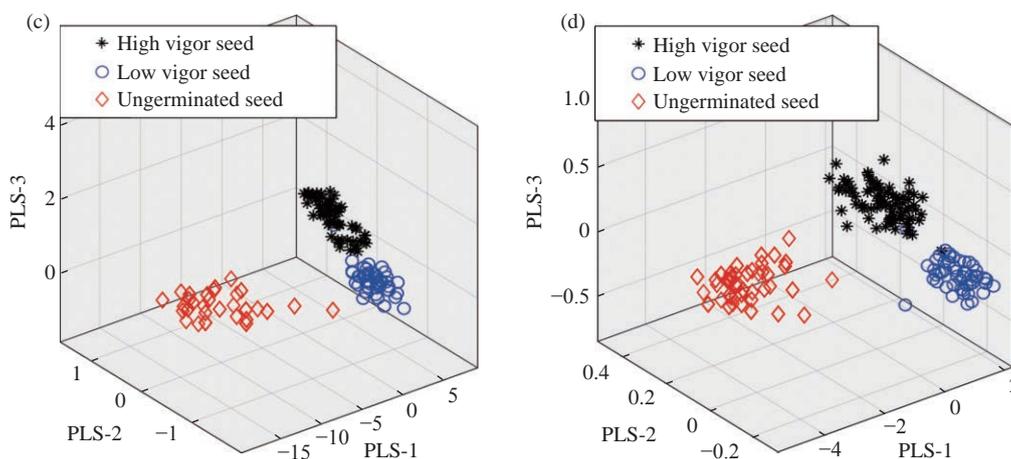


图 6 PLS-DA 模型对 2018 年(a)、2017 年(b)、2016 年(c)及随机混合(d)的水稻种子活力判别结果

Fig. 6 Evaluation results of vigor of the rice seed in 2018 (a)、2017 (b)、2016 (c) and random mixing (d) determined by PLS-DA model

对不同年份水稻种子的光谱数据与活力类别建立偏最小二乘判别分析模型,通过识别检测样本的准确率判断模型的预测精度,分类后种子的发芽率如表 4 所示。从表 4 可以看出,该模型分级的准确率可以达到 90% 以上,筛选后种子的发芽率可以到达 95% 以上。

表 4 PLS-DA 模型判别准确率及筛选后种子发芽率

Tab. 4 Total accuracy of PLS-DA model and seed germination of rice seeds after screening

年份	校正集准确率/%	验证集准确率/%	筛选后发芽率/%
2018	94.44	95.92	97.17
2017	93.98	94.69	96.52
2016	91.53	92.37	95.06
随机混合	92.74	93.16	96.07

5 结 论

为了实现带稃壳水稻种子的活力分级,本文

参考文献:

- [1] 杨振发,肖航,张雷,等.基于近红外光谱的水泥生料氧化物含量快速测定方法研究[J].分析化学,2020,48(2):275-281.
YANG ZH F, XIAO H, ZHANG L, *et al.*. Rapid determination of oxides content in cement raw meal based on near infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2020, 48(2): 275-281. (in Chinese)
- [2] AMBROSE A, LOHUMI S, LEE W H, *et al.*. Comparative nondestructive measurement of corn seed viability using fourier transform near-infrared (FT-NIR) and raman spectroscopy[J]. *Sensors and Actuators B: Chemical*, 2015, 224:

设计并搭建了一套透射式近红外光谱检测系统,获取了 3 种不同年分的日本晴水稻种子在 1100~2100 nm 波段内的近红外光谱信息。光谱分析结果表明:水稻种子的活力梯度与其近红外吸收光谱的吸收峰值相关。结合光谱特性建立了主成分分析模型,主成分数目为 3 时模型累积贡献率为 93.5%,此时水稻的近红外吸收光谱与其振动吸收特性有较好的一致性。最后,通过采用偏最小二乘方法建立了不同年份水稻种子的活力判别模型。结果表明:相较于 SNV、MSC 等算法,采用归一化、SD 和 OSC 算法对原始光谱进行校正后,再利用 PLS-DA 模型对不同年份水稻种子活力梯度的分级具有较高的预测能力,校正集和验证集的分级准确率均可达到 90% 以上,筛选后水稻种子的发芽率可达 95% 以上。利用本论文设计的检测系统与优化的光谱处理方法实现了不同活力水稻种子的快速无损分类,为提高苗的质量、积压种子的筛选、优质水稻选育生产等应用需求提供了一种有效的技术手段。

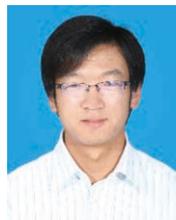
- 500-506.
- [3] 李茂刚, 闫春华, 薛佳, 等. 近红外光谱结合小波变换—随机森林法快速定量分析甲醇汽油中甲醇含量[J]. 分析化学, 2019, 47(12): 1995-2003.
LI M G, YAN CH H, XUE J, *et al.*. Rapid quantitative analysis of methanol content in methanol gasoline by near infrared spectroscopy coupled with wavelet transform-random forest[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(12): 1995-2003. (in Chinese)
- [4] 何龙生. 水稻种子活力测定方法的初步研究[D]. 杭州: 浙江农林大学, 2018: 1-2.
HE L SH. The preliminary study on methods for determination of rice seed vigor[D]. Hangzhou: Zhejiang A&F University, 2018. (in Chinese)
- [5] 王岳含. 我国种子质量可追溯系统研究[D]. 北京: 中国农业科学院, 2016.
WANG Y H. Seed quality tracing system of china research[D]. Beijing: Chinese Academy of Agricultural Sciences, 2016. (in Chinese)
- [6] 牟致远. 小麦种子活力及其遗传分析[J]. 西南农业大学学报, 1987, 9(4): 421-425.
MU ZH Y. Seedling vigor and genetic analysis in wheat varieties[J]. *Journal of Southwest Agricultural University*, 1987, 9(4): 421-425. (in Chinese)
- [7] 王青峰, 宫庆友, 沈凌云, 等. 超甜玉米种子活力研究[J]. 种子, 2007, 26(6): 4-7.
WANG Q F, GONG Q Y, SHEN L Y, *et al.*. Study of combining ability of seed vigor in super sweet corn[J]. *Seed*, 2007, 26(6): 4-7. (in Chinese)
- [8] 朱银, 颜伟, 杨欣, 等. 基于近红外光谱的小麦种子发芽率测试[J]. 江苏农业科学, 2015, 43(12): 111-113.
ZHU Y, YAN W, YANG X, *et al.*. Test of germination rate of wheat seeds based on near infrared spectroscopy[J]. *Jiangsu Agricultural Sciences*, 2015, 43(12): 111-113. (in Chinese)
- [9] 李欢欢, 卢伟, 杜昌文, 等. 基于光声光谱结合LS-SVR的稻种活力快速无损检测方法研究[J]. 中国激光, 2015, 42(11): 1115003.
LI H H, LU W, DU CH W, *et al.*. Study on rapid and non-destructive detection of rice seed vigor based on photoacoustic spectroscopy combined with LS-SVR[J]. *Chinese Journal of Lasers*, 2015, 42(11): 1115003. (in Chinese)
- [10] 李美凌, 邓飞, 刘颖, 等. 基于高光谱图像的水稻种子活力检测技术研究[J]. 浙江农业学报, 2015, 27(1): 1-6.
LI M L, DENG F, LIU Y, *et al.*. Study on detection technology of rice seed vigor based on hyperspectral image[J]. *Acta Agriculturae Zhejiangensis*, 2015, 27(1): 1-6. (in Chinese)
- [11] 宋乐, 王琦, 王纯阳, 等. 基于近红外光谱的单粒水稻种子活力快速无损检测[J]. 粮食储藏, 2015, 44(1): 20-23.
SONG L, WANG Q, WANG CH Y, *et al.*. Qualitative analysis of single rice seed vigor using near infrared reflectance spectroscopy[J]. *Grain Storage*, 2015, 44(1): 20-23. (in Chinese)
- [12] ZHOU X B, ZHAO J W, POVEY M J W, *et al.*. Variables selection methods in near-infrared spectroscopy[J]. *Analytica Chimica Acta*, 2010, 667(1-2): 14-32.
- [13] 史永刚, 栗斌, 田高友, 等. 化学计量学方法及 MATLAB 实现[M]. 北京: 中国石化出版社, 2010.
SHI Y G, LI B, TIAN G Y, *et al.*. *Chemometrics Method and MATLAB Implementation*[M]. Beijing: China Petrochemical Press, 2010. (in Chinese)
- [14] 郭帅, 苏杭, 黄星灿, 等. 光学无创血糖浓度检测方法的研究进展[J]. 中国光学, 2019, 12(6): 1235-1248.
GUO SH, SU H, HUANG X C, *et al.*. Research progress in optical methods for noninvasive blood glucose detection[J]. *Chinese Optics*, 2019, 12(6): 1235-1248. (in Chinese)
- [15] 王纯阳. 基于近红外光谱的单籽粒水稻种子品质检测的方法研究[D]. 合肥: 中国科学技术大学, 2017.
WANG CH Y. The nondestructive quality analysis of single rice seed using near infrared spectroscopy[D]. Hefei: University of Science and Technology of China. (in Chinese)
- [16] 高升, 王巧华, 李庆旭, 等. 基于近红外光谱的红提维生素C含量、糖度及总酸含量无损检测方法[J]. 分析化学, 2019, 47(6): 941-949.
GAO SH, WANG Q H, LI Q X, *et al.*. Non-destructive detection of vitamin C, sugar content and total acidity of red globe grape based on near-infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(6): 941-949. (in Chinese)

- [17] 付丹丹, 王巧华, 高升, 等. 不同品种鸡蛋贮期S—卵白蛋白含量分析及其可见/近红外光谱无损检测模型研究[J]. 分析化学, 2020, 48(2): 289-297.
FU D D, WANG Q H, GAO SH, *et al.*. Analysis of S-Ovalbumin content of different varieties of eggs during storage and its nondestructive testing model by visible-near infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2020, 48(2): 289-297. (in Chinese)
- [18] 刘宁武, 许林广, 周胜, 等. 量子级联激光光谱在土壤生态系统中的应用[J]. 光学学报, 2019, 39(11): 1130001.
LIU N W, XU G L, ZHOU SH, *et al.*. Application of quantum-cascade laser spectroscopy to soil ecosystems[J]. *Acta Optica Sinica*, 2019, 39(11): 1130001. (in Chinese)
- [19] 谢臣瑜, 翟文超, 李健军, 等. 超连续激光单色仪系统级光谱响应度定标比对验证[J]. 红外与激光工程, 2020, 49(2): 0205005.
XIE CH Y, ZHAI W CH, LI J J, *et al.*. System-level spectral responsivity calibration comparison and validation of supercontinuum laser and monochromator[J]. *Infrared and Laser Engineering*, 2020, 49(2): 0205005. (in Chinese)
- [20] 曹栋栋, 阮晓丽, 詹艳, 等. 杂交水稻种子不同活力测定方法与其田间成苗率的相关性[J]. 浙江农业学报, 2014, 26(5): 1145-1150.
CAO D D, RUAN X L, ZHAN Y, *et al.*. Relativity analysis between seedling percentage in field and different seed vigor testing methods of hybrid rice seeds[J]. *Acta Agriculturae Zhejiangensis*, 2014, 26(5): 1145-1150. (in Chinese)
- [21] 刘燕德, 叶灵玉, 孙旭东, 等. 基于光谱指数的蜜橘成熟度评价模型研究[J]. 中国光学, 2018, 11(1): 83-91.
LIU Y D, YE L Y, SUN X D, *et al.*. Maturity evaluation model of tangerine based on spectral index[J]. *Chinese Optics*, 2018, 11(1): 83-91. (in Chinese)
- [22] 董盈红. 油菜籽的傅里叶变换红外光谱鉴别[J]. 保山学院学报, 2018, 37(2): 38-41.
DONG Y H. Identification of rapeseed by fourier transform infrared spectroscopy[J]. *Journal of Baoshan Teachers College*, 2018, 37(2): 38-41. (in Chinese)
- [23] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011.
CHU X L. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications*[M]. Beijing: Chemical Industry Press, 2011. (in Chinese)
- [24] 王昕, 吕世龙, 李岩, 等. 基于基线漂移模型的气体光谱自动基线校正[J]. 光谱学与光谱分析, 2018, 38(12): 3946-3951.
WANG X, LV SH L, LI Y, *et al.*. Automatic baseline correction of gas spectra based on baseline drift model[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(12): 3946-3951. (in Chinese)
- [25] 王凡, 李永玉, 彭彦昆, 等. 基于可见/近红外透射光谱的番茄红素含量无损检测方法研究[J]. 分析化学, 2018, 46(9): 1424-1431.
WANG F, LI Y Y, PENG Y K, *et al.*. Nondestructive determination of lycopene content based on visible/near infrared transmission spectrum[J]. *Chinese Journal of Analytical Chemistry*, 2018, 46(9): 1424-1431. (in Chinese)
- [26] 高升, 王巧华, 付丹丹, 等. 红提糖度和硬度的高光谱成像无损检测[J]. 光学学报, 2019, 39(10): 355-364.
GAO SH, WANG Q H, FU D D, *et al.*. Nondestructive detection of sugar content and firmness of red globe grape by hyperspectral imaging[J]. *Acta Optica Sinica*, 2019, 39(10): 355-364. (in Chinese)
- [27] 范雪婷, 朱明东, 杨晨光, 等. 利用近红外吸收光谱对水稻种子活力的判别方法[J]. 杂交水稻, 2019, 34(4): 62-67.
FAN X T, ZHU M D, YANG CH G, *et al.*. Assessment of rice seed vigor using near infrared spectroscopy[J]. *Hybrid Rice*, 2019, 34(4): 62-67. (in Chinese)
- [28] SAMPAIO P S, SOARES A, CASTANHO A, *et al.*. Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms[J]. *Food Chemistry*, 2018, 242: 196-204.
- [29] LIU D L, WU Y X, GAO Z M, *et al.*. Comparative non-destructive classification of partial waxy wheats using near-infrared and raman spectroscopy[J]. *Crop & Pasture Science*, 2019, 70(5): 437-441.

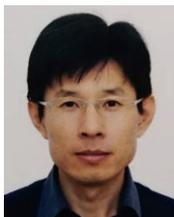
作者简介:



金文玲(1994—), 女, 辽宁大连人, 硕士研究生, 2017年于沈阳理工大学获得学士学位, 主要从事光学与光谱学检测系统设计及方法研究。E-mail: jinwenling17@mails.ucas.ac.cn



吕金光(1984—)男, 吉林蛟河人, 博士, 副研究员, 硕士生导师, 2008年于吉林大学物理学院获得学士学位, 2013年于中国科学院长春光学精密机械与物理研究所获得博士学位, 主要从事微小光学系统设计与光学信息处理方面的研究。E-mail: jinguanglv@163.com



阚瑞峰(1977—), 男, 辽宁锦州人, 研究员, 博士生导师, 主要从事激光光谱检测方法及其在环境污染、生产安全、航空航天流场诊断、深海溶解气体检测等方面应用的研究。E-mail: rfkan@ciomp.ac.cn

《发光学报》

—EI 核心期刊(物理学类; 无线电电子学、电信技术类)

《发光学报》是中国物理学会发光分会与中国科学院长春光学精密机械与物理研究所共同主办的中国物理学会发光分会的学术会刊。该刊是以发光学、凝聚态物质中的激发过程为专业方向的综合性学术刊物。

《发光学报》于1980年创刊, 曾于1992年, 1996年, 2000年和2004年连续四次被《中文核心期刊要目总览》评为“物理学类核心期刊”, 并于2000年同时被评为“无线电电子学、电信技术类核心期刊”。2000年获中国科学院优秀期刊二等奖。现已被《中国学术期刊(光盘版)》、《中国期刊网》和“万方数据资源系统”等列为源期刊。英国《科学文摘》(SA)自1999年; 美国《化学文摘》(CA)和俄罗斯《文摘杂志》(AJ)自2000年; 美国《剑桥科学文摘社网站》自2002年; 日本《科技文献速报》(CBST, JICST)自2003年已定期收录检索该刊论文; 2008年被荷兰“Elsevier Bibliographic Databases”确定为源期刊; 2010年被美国“EI”确定为源期刊。2001年在国家科技部组织的“中国期刊方阵”的评定中, 《发光学报》被评为“双效期刊”。2002年获中国科学院2001~2002年度科学出版基金“择重”资助。2004年被选入《中国知识资源总库·中国科技精品库》。本刊内容丰富、信息量大, 主要反映本学科专业领域的科研和技术成就, 及时报道国内外的学术动态, 开展学术讨论和交流, 为提高我国该学科的学术水平服务。

《发光学报》自2011年改为月刊, A4开本, 144页, 国内外公开发行。国内定价: 40元, 全年480元, 全国各地邮局均可订阅。《发光学报》欢迎广大作者、读者广为利用, 踊跃投稿。

地 址: 长春市东南湖大路3888号
《发光学报》编辑部
邮 编: 130033
电 话: (0431)86176862, 84613407
E-mail: fgxbt@126.com

国内统一刊号: CN 22-1116/04
国际标准刊号: ISSN 1000-7032
国内邮发代号: 12-312
国外发行代号: 4863BM
http://www.fgxb.org