

基于高光谱技术的香肠亚硝酸盐快速检测方法

Rapid detection method of sausage nitrite based on hyperspectral technology

刘 峥 殷 勇

LIU Zheng YIN Yong

(河南科技大学食品与生物工程学院, 河南 洛阳 471023)

(College of Food and Bioengineering, Henan University of Science and Technology, Luoyang, Henan 471023, China)

摘要:选取 7 个不同储藏时期的香肠分别进行亚硝酸盐含量检测和对应的光谱数据采集,并用 Savitzky-Golary 法进行光谱数据预处理,以减少光谱数据的噪声;在预处理后的光谱数据基础上,用偏最小二乘回归系数法提取出 29 个特征波长;对比分析了特征波长和全波长下香肠中亚硝酸盐含量预测模型的检测精度。结果表明:全波长下的回归模型预测结果均高于特征波长下,且全波长下偏最小二乘回归模型优于主成分回归模型,表征偏最小二乘回归模型精度的决定系数和均方根误差分别为 0.982 9 和 0.059 2。说明全波长下的光谱信息更适用于香肠储藏过程中亚硝酸盐含量高光谱检测模型的构建。

关键词:香肠;储藏时期;亚硝酸盐;高光谱;特征波长;回归模型

Abstract: Seven different storage period sausages were selected for nitrite content detection and corresponding spectral data collection, and uses Savitzky-Golary method to preprocess spectral data to reduce the noise of spectral data. Then based on the pre-processed spectral data, 29 characteristic wavelengths were extracted by partial least squares regression coefficient method. Finally, the detection accuracy of the prediction model of nitrite in sausages at characteristic wavelength and full wavelength were analyzed. The results showed that the prediction results of the regression model based on full wavelength were all higher than that based on characteristic wavelength, and the full-wavelength partial least squares regression model was superior to that of the principal component regression model, and the coefficient of determination of the accuracy of the partial least squares regression

model was determined. The R^2 and root mean square errors were 0.982 9 and 0.059 2, respectively. The dissertation studies show that the spectral information at full wavelength is more suitable for the construction of hyperspectral detection model of nitrite content in sausage storage.

Keywords: sausage; storage time; nitrite; hyperspectral; characteristic wavelength; regression model

在香肠的制作过程中,为了呈现良好的色泽和防止腐蚀,会加入一定量的亚硝酸盐。随着香肠存放时间的延长,原先添加的亚硝酸盐会分解消耗掉一部分,而香肠中的肉品本身也会产生一部分亚硝酸盐,导致香肠在储藏过程中亚硝酸盐含量不断变化,存在不确定性。由于亚硝酸盐的毒性强,摄入过多时会让血液中的血红蛋白转化成高铁血红蛋白,失去输送氧气的能力,使组织产生缺氧,引起紫绀现象甚至死亡^[1]。

现阶段,检测食品中亚硝酸盐的主要方法有高效液相色谱法^[2]、国际格里斯(Griess)试剂比色法^[3]、荧光光度法^[4-5]、毛细管电泳法^[6]、离子色谱法^[7]等。虽然这些方法较为普遍,但是存在操作复杂,对环境敏感,适用范围窄等问题。而高光谱技术具有可视性、快速性等优点,在水果^[8-9]、小麦^[10]、玉米^[11]、茶叶^[12]等农副产品品质检测中得以广泛地运用。文献检索发现,利用高光谱技术针对肉制品中亚硝酸盐检测的研究报道比较少。陈晓东等^[13]利用高光谱技术研究了香肠中亚硝酸盐的预测方法,指出了高光谱技术检测香肠中亚硝酸盐的可行性。但是,该文是基于主成分分析提取的特征所构建的预测模型还不能较好地满足实用要求,模型最高预测精度仅为 0.918。本试验拟从回归模型输入信息的选择方面作为切入点尝试提高高光谱技术检测香肠在储藏过程中亚硝酸盐含量的准确性。

基金项目:河南省科技攻关项目(编号:182102110422)

作者简介:刘峥,女,河南科技大学在读硕士研究生。

通信作者:殷勇(1966—),男,河南科技大学教授,硕士生导师,博士。E-mail: yinyong08@126.com

收稿日期:2019-02-05

1 材料与方法

1.1 试验材料

香肠样品:双汇润口香甜王玉米风味香肠,购于超市。样品保质期为 120 d,储藏方式与购买时超市的存放方式相同,即自然条件(温室环境)下储藏。由于购买日期是生产日期的第 29 天,所以试验选取储藏 30,50,70,90,110,130,150 d 的香肠样品进行亚硝酸盐含量检测。每个样品分别选取 40 个样本,每个样本含量为 (24.0 ± 0.5) g,包含 21 个香肠切片。其中,任选 30 个样本构造训练集,共 210 个样本;剩余的 10 个样本构造测试集,共 70 个样本。在模型构建中,随机生成 3 组训练集及其对应的测试集来分别构建模型和校验模型,以说明研究结果的可靠性。

1.2 高光谱系统

实验采用的高光谱系统是由计算机、光谱仪(IST 50-3810 型,德国 Inno-spec 公司,光谱范围为 371.05 ~ 1 023.82 nm,涵盖可见光和部分近红外光谱)、4 个 500 W 的光纤卤素灯(RK90000420108 型,德国 Esylux 公司)和传送装置等组成。其示意图与性能参数与刘燕德等^[14]使用的装置相同。

1.3 样本高光谱数据采集和校正

在采集样品高光谱数据时,将香肠样品切片平铺在规格为 10 cm×1 cm 的培养皿中,再将盛有样品的培养皿放置在传送带上,带速 2 mm/s,用 SICap-STVR V1.0.x 控制驱动,获得高光谱信息采集结果。每个测试样本的高光谱反射值采用 ENVI 4.8 软件提取,最终可以采集到 1 288 个波段的高光谱反射值。在采集样本的高光谱图像时,须进行黑白校正。具体校正方法为:白板校正使用白色特氟龙(Teflon)标准矫正板进行扫描得到全白的标定图像,黑板校正则是关闭光源及相机镜头得到全黑的标定图像。

其他数据处理方法均在 Matlab 2014a 平台上实现。

1.4 光谱预处理

采集高光谱信息时,会受到高光谱仪器的电路噪声干扰,而且样品表面不平整也会影响到原始光谱数据的采集。所以,为了减少这些外界因素对光谱信息的干扰,在建立模型之前需对原始光谱数据进行预处理。采用 Savitzky-Golary 卷积平滑法(SG 平滑)对原始光谱数据进行预处理^[15-16]。Savitzky-Golary 卷积平滑法是通过多项式来对移动窗口内的数据进行多项式最小二乘拟合,其实质是一种加权平均法,更强调中心点的中心作用^[17]。

1.5 香肠中亚硝酸盐含量测定

采用 GB 5009.33—2016 的检测方法。对所选择的每个储存时间的样品进行 3 次平行样本测试,取平均值作为检测结果。

1.6 特征波长的提取

因高光谱信息共有 1 288 个波段的光谱数据,数据繁多,会提高建模时的复杂度,所以在高光谱分析中通常会进行特征波长的提取^[18-19]。在特征波长提取方法上,常用偏最小二乘回归系数大小作为选择特征波长的依据^[20-21]。因此,采用偏最小二乘回归系数提取特征波长。

1.7 香肠中亚硝酸盐的定量分析方法

1.7.1 多元回归 多元回归是对相关随机变量进行预测,确定这些变量之间数量关系的可能形式,并用数学模型来表示。多元回归模型的精确度由决定系数(determination coefficients, R^2)、均方根误差(root mean squared error, RMSE)2 个指标决定, R^2 越接近于 1,精度越高,模型越稳定, RMSE 越小,模型的预测能力越高。

1.7.2 主成分回归 主成分回归(PCR)是目前处理高维复杂数据时非常有效的方法之一。它可以对复杂的高维数据进行降维,在不丢失主要数据信息的情况下选择维数较少的新变量来代替原来较多的变量,以排除众多信息共存中相互重叠的现象以及夹杂的噪声等干扰,还可以解决高维数据的多重共线性问题,从而使预测结果更加准确合理。本研究采用 R^2 和 RMSE 进行评价, R^2 越接近于 1,精度越高,模型越稳定, RMSE 越小,模型的预测能力越高。

1.7.3 偏最小二乘回归 偏最小二乘回归(PLSR)是一种多元统计数据分析方法,可以同时实现回归建模、简化数据结构和分析 2 组变量间的相关性。PLSR 模型预测精度取决于 R^2 、RMSE 两个指标, R^2 越接近于 1,精度越高,模型越稳定, RMSE 越小,模型的预测能力越高。

2 结果与分析

2.1 香肠中亚硝酸盐的含量

由表 1 可知,香肠在储藏过程中亚硝酸盐含量是变化的、不确定的。因此,对香肠储藏过程中亚硝酸盐的检测、监控是必要的。

2.2 光谱预处理

为了减少外界因素对于光谱信息的干扰,需对原始

表 1 亚硝酸盐理化试验结果

Table 1 Physicochemical test results of nitrite content

储藏天数/d	亚硝酸盐含量/(mg·kg ⁻¹)
30	1.232
50	0.507
70	1.087
90	1.159
110	2.100
130	1.521
150	1.232

光谱数据进行 SG 平滑处理。从图 1、2 中可以看出,经过 SG 平滑法处理过的光谱曲线相较于原始光谱曲线更平滑,受到的噪音影响很小,有利于后期模型的建立。

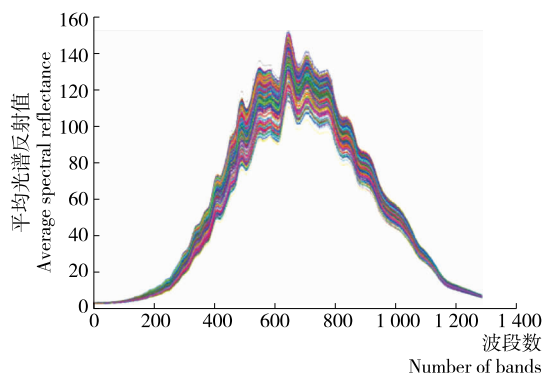


图 1 原始光谱图

Figure 1 The original spectrum

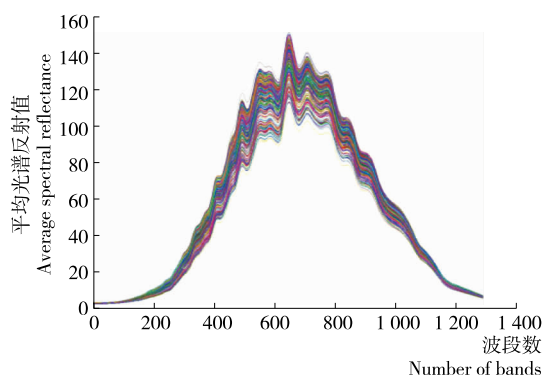


图 2 经 SG 平滑处理后的光谱图

Figure 2 Spectrogram after smoothing by SG

2.3 特征波长的筛选

可见—近红外高光谱中波长范围是 371.05 ~ 1 023.82 nm,但试验操作过程中由于首尾波段受环境及仪器噪声影响较为严重,因此在光谱信息分析的过程中应只考虑 400~1 000 nm 波段下的信息。对全波长数据经偏最小二乘回归分析后,得到的回归曲线如图 3 所示。第 1 主成分涵盖的信息较为全面和常见,波动较小,不能较好地体现样本间的差异;第 2 主成分有一定的波动,可部分体现样本间的差异,所以选取第 2 主成分回归系数的最大值和最小值所对应的波长作为 2 个特征波长;第 3 主成分波动明显,可较好地体现样本间的差异,所以选取第 3 主成分所有波峰和波谷的回归系数所对应的波长作为特征波长。这样,第 2 主成分有 2 个特征波长,第 3 主成分有 27 个特征波长,共计 29 个特征波长。

2.4 特征波长下的定量分析

2.4.1 特征波长下多元回归建模 直接将 29 个特征波长作为模型的输入变量,进行回归分析。图 4 给出了第 1 组数据集的预测结果。由图 4 可以看出,预测值与真实

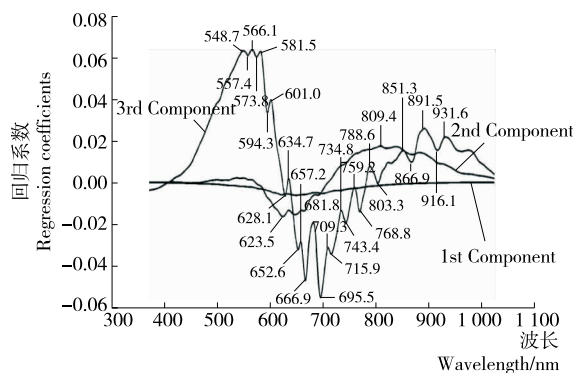


图 3 权重系数图

Figure 3 The weight coefficient diagram

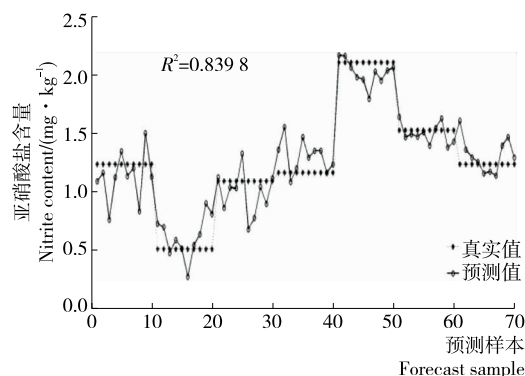


图 4 特征波长下多元回归结果

Figure 4 Multiple regression results at characteristic wavelength

表 2 特征波长下的回归模型结果

Table 2 Results of regression models at characteristic wavelengths

测试集编号	R^2	RMSE
1	0.839 8	0.179 3
2	0.839 2	0.180 2
3	0.858 8	0.168 7

值相差较大,预测结果精确度不高。表 2 给出了 3 组测试集模型预测结果的 R^2 与 RMSE,从表 2 中可以看出,3 组预测集 R^2 最高为 0.858 8,对应的 RMSE 为 0.168 7,预测结果不理想。考虑到各个特征波长之间会存在一定的相关性,影响建模的精度,故尝试建立特征波长下主成分回归和偏最小二乘回归的预测模型。

2.4.2 主成分回归模型的定量分析 对 29 个特征波长变量进行 PCA 分析,得到按贡献率从大到小排序的 29 个主成分。经比较,在提取前 26 个主成分时预测结果准确性相对较高,如图 5 所示(以第 1 组测试集为例,下同)。表 3 为 3 组测试集模型预测结果的 R^2 与 RMSE,从表 3 中可以看出,3 组测试集的预测结果的 R^2 最高为 0.896 1,对应的 RMSE 为 0.148 8,模型精度仍不理想。

表 3 特征波长下的主成分回归与偏最小二乘回归结果

Table 3 Principal component regression and partial least squares regression results at characteristic wavelengths

测试集编号	主成分回归		偏最小二乘回归	
	R^2	RMSE	R^2	RMSE
1	0.888 2	0.149 6	0.911 1	0.139 7
2	0.893 1	0.145 5	0.906 1	0.141 3
3	0.896 1	0.148 8	0.901 4	0.141 7

2.4.3 偏最小二乘回归模型的定量分析 对于 PLSR 模型的构建,经比较,同样在提取前 26 个主成分时预测结果相对较高,如图 6 所示。表 3 给出了 3 组测试集模型预测结果的决定系数与均方根误差,从表 3 中可以看出,3 组测试集的预测结果的决定系数 R^2 最高为 0.911 1,对应的均方根误差 RMSE 为 0.139 7。模型精度有所提高,但还是不太理想。

综合基于特征波长的检测结果来看,在特征波长下建立 PCR 和 PLSR 预测模型的结果虽然比直接回归建模分析的结果较优,但仍不很理想,且模型的变量仅降到

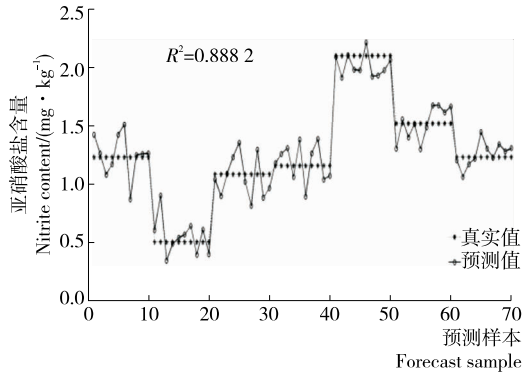


图 5 特征波长下主成分回归结果

Figure 5 Principal component regression results at characteristic wavelength

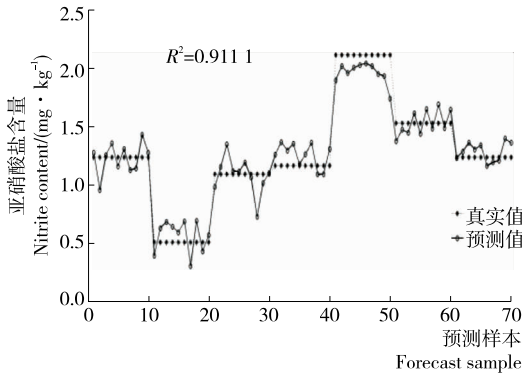


图 6 特征波长下偏最小二乘回归的结果

Figure 6 Results of partial least squares regression with characteristic wavelength

26 个,比 29 个特征波长并无明显的减少,建模的复杂度仍较高。陈晓东等^[13]在用主成分分析方法提取特征波长的基础上构建了预测模型,但预测精度只有 0.918,且数据处理过程繁杂。这可能是选择特征波长表达的信息不够全面,不能充分体现原始数据的信息,从而导致了预测模型的精度不高。而岳学军等^[22]采用全波长数据信息作为模型输入向量较好地实现了对柑橘叶片叶绿素含量的检测。受此启发,本试验尝试了在全波长下建立 PCR 和 PLSR 预测模型的效果。

2.5 全波长下的定量分析

为了克服特征波长不能充分表征原始光谱信息这一问题,本试验对 1 288 个全波长数据进行 PCA 分析;同时选取前 29 个主成分作为回归模型的输入向量,这样可以与特征波长下的定量分析结果形成对比。

2.5.1 主成分回归的定量分析 在全波长下建立 PCR 预测模型,基于前 29 个主成分构建的模型得到的预测结果如图 7 所示。从图 7 中可以看出,预测值与真实值较为接近,预测结果精确度提高。表 4 为 3 组测试集模型预测结果的 R^2 与 RMSE,从表 4 中可以看出,3 组测试集模型预测结果的 R^2 最高为 0.952 2,对应的 RMSE 为 0.097 4。模型精度较高,预测结果较为理想。

2.5.2 偏最小二乘回归的定量分析 在全波长下建立 PLSR 预测模型,其主成分变量仍为前 29 个,得到的预测结果如图 8 所示。图 8 表明,预测值与真实值非常接近,预测结果精度很高。表 4 为 3 组测试集模型预测结果的 R^2 与 RMSE,从表 4 中可以看出,3 组测试集模型预测结果的 R^2 最高为 0.982 9,对应的 RMSE 为 0.059 2。模型精度很高,预测结果非常理想。

预测结果可以表明,全波长下建立的预测模型结果精确度较高,且偏最小二乘回归的结果高于主成分回归的,决定系数可以高达 0.978 8 以上。因此,在全波长下建立偏最小二乘回归的预测模型可以满足检测要求。综上所述认为,在全波长下建立的主成分回归和偏最小二乘回归预测模型,将变量也降低到了 29 个,远低于 1 288 个波长,省去了特征波长的选择计算,不仅简化了数据处理步骤,而且还能得到较为理想的预测结果。另外,也可以认为,全波长下的前 29 个主成分可以较充分地表征原始光谱数据的信息。

表 4 全波长下的主成分回归与偏最小二乘回归结果

Table 4 Principal component regression and partial least squares regression results at full wavelength

测试集编号	主成分回归		偏最小二乘回归	
	R^2	RMSE	R^2	RMSE
1	0.952 2	0.097 4	0.982 9	0.059 2
2	0.938 3	0.112 9	0.978 8	0.064 7
3	0.950 7	0.100 2	0.980 2	0.066 2

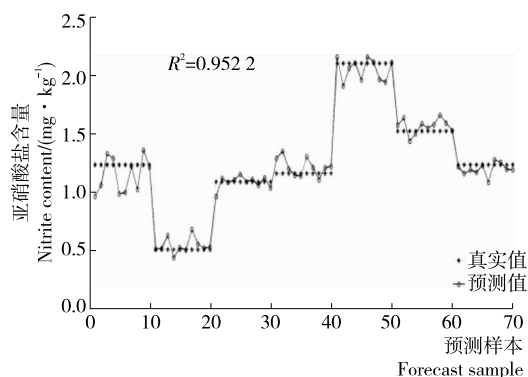


图 7 全波长下主成分回归的结果

Figure 7 Results of principal component regression at full wavelength

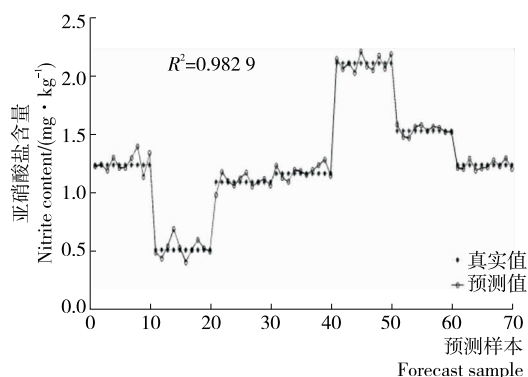


图 8 全波长下偏最小二乘回归的结果

Figure 8 Results of partial-wavelength least squares regression at full wavelength

3 结论

采用高光谱技术检测香肠储藏中亚硝酸盐含量时,在提取特征波长的基础上进行回归模型构建,得到的模型精度不高,效果不理想,预测精度最高只达到了 0.911 1,且数据处理过程复杂。直接在全波长之下建立的预测模型既可以提高预测结果的精度,又可以降低从预处理到建模计算过程中的复杂性。所以全波长信息作为香肠储藏过程中亚硝酸盐含量高光谱检测模型信息的输入是合适的。另外,考虑到提取特征波长在高光谱研究中的优势,在今后的研究中,应该对特征波长的提取方法进行更多的尝试,以提高检测模型的预测精度。

参考文献

- [1] 谢燕丹,刘零怡,楼乔明,等.加工蔬菜中亚硝酸盐的消除技术研究进展[J].食品与发酵工业,2016,42(8):279-286.
- [2] 战旭梅,刘靖,刘萍.高效液相色谱法测定香肠中亚硝酸盐含量[J].食品与机械,2014,30(6):72-74.
- [3] 白雪娟,徐世众.格里斯试剂比色法测定肉制品中亚硝酸盐含量[J].肉类工业,2008(7):46-48.
- [4] HORNYAK I, KOZMA L, LAPAT A, et al. Spectrofluorimetric determination of diethazine and promethazine in pharmaceutical preparations[J]. Biomed.Chromatogr, 1997, 11(2): 99-101.
- [5] 杜红霞,贺稚非,李洪军.食品中亚硝酸盐检测技术研究进展[J].肉类研究,2006(1):12,46-50.
- [6] MERUSI C, CORRADINI C, CAVAZZA A, et al. Determination of nitrates, nitrites and oxalates in food products by capillary electrophoresis with pH-dependent electroosmotic flow reversal[J]. Food Chem., 2010, 120(2): 615-620.
- [7] 蔡芸梅,侯如燕,高柱,等.离子色谱法测定茶叶中的硝酸盐和亚硝酸盐[J].茶叶科学,2012,32(2):95-99.
- [8] 罗霞,洪添胜,罗阔,等.高光谱技术在无损检测火龙果可溶性固形物中的应用[J].激光与光电子学进展,2015,52(8):315-323.
- [9] 詹白勺,倪君辉,李军.高光谱技术结合 CARS 算法的库尔勒香梨可溶性固形物定量测定[J].光谱学与光谱分析,2014,34(10):2752-2757.
- [10] 吴静珠,吴胜男,刘翠玲,等.近红外和高光谱技术用于小麦籽粒蛋白含量预测探索[J].传感器与微系统,2013,32(2):60-62.
- [11] 于慧春,姜楠,殷勇,等.基于高光谱技术及 SPXY 和 SPA 的玉米毒素检测模型建立[J].食品科学,2018,39(16):328-335.
- [12] 于英杰,孙威江.近红外光谱及高光谱技术在茶叶上的应用[J].亚热带农业研究,2014,10(4):269-273.
- [13] 陈晓东,郭培源.基于主成分分析法提取高光谱图像特征检测香肠亚硝酸盐含量[J].肉类研究,2016,30(12):22-27.
- [14] 刘燕德,张光伟.高光谱成像技术在农产品检测中的应用[J].食品与机械,2012,28(5):223-226.
- [15] 梁逸曾,俞汝勤.分析化学手册(10):化学计量学[M].北京:化工出版社,2001:162-165.
- [16] 刘树深,易忠胜.基础化学计量学[M].北京:科学出版社,1999:41-43.
- [17] 褚小立,袁洪福,陆婉珍.近红外分析中光谱预处理及波长选择方法进展与应用[J].化学进展,2004(4):528-542.
- [18] 丁希斌,张初,刘飞,等.高光谱成像技术结合特征提取方法的草莓可溶性固形物检测研究[J].光谱学与光谱分析,2015,35(4):1020-1024.
- [19] 于雷,章涛,朱亚星,等.基于 IRIV 算法优选大豆叶片高光谱特征波长变量估测 SPAD 值[J].农业工程学报,2018,34(16):148-154.
- [20] 薛建新,张淑娟,张晶晶,等.壶瓶枣自然损伤的高光谱成像检测[J].农业机械学报,2015,46(7):220-226.
- [21] 薛建新,张淑娟,张晶晶.基于高光谱成像技术的沙金杏成熟度判别[J].农业工程学报,2015,31(11):300-307.
- [22] 岳学军,全东平,洪添胜,等.柑橘叶片叶绿素含量高光谱无损检测模型[J].农业工程学报,2015,31(1):294-302.