

鸡胸肌冻干粉赖氨酸近红外定量 预测模型的建立与优化

陶琳丽¹, 陶冶², 甘文斌³, 杜光英¹, 李富银¹, 王天武¹, 尹汝高凡¹, 杨万进¹, 张曦^{1*}, 牛国一^{1*}

(1. 云南农业大学动物科学技术学院, 云南省动物营养与饲料重点实验室, 云南昆明 650201)

(2. 云南省饲料工业协会, 云南昆明 650201) (3. 云南省兽药饲料检测所, 云南昆明 650201)

摘要: 该研究采用偏小二乘法建立鸡胸肌冻干粉赖氨酸近红外定量预测模型, 并通过光谱预处理、建模特征光谱筛选、异常样本剔除、建模和验证样本选择 4 个方面对模型进行优化, 旨在提高模型的预测精度和稳健性。以 263 个鸡胸肌冻干粉为研究对象, 研究 7 种不同光谱预处理方法、4 种特征光谱筛选方法、MCCV 异常样本剔除方法, SPXY 和鸡种 2 种建模验证样本选取方法对鸡胸肌冻干粉赖氨酸近红外定量预测模型的影响。结果表明: 在 1000~2502 nm 建模谱区, 剔除 68 个异常样本后, 使用 SPXY 方法选取 156 个校正样本、39 个外部验证样本, 使用原始光谱所建模型最优; 没有进行异常样本剔除、建模特征谱区选择等处理时, 1000~2502 nm 谱段建模, 光谱需进行 SNV+gapsegment (1#,15,7) 预处理; 异常样本剔除和建模样本选择对建模谱段 1386~1379 nm、1329~1323 nm、1289~1283 nm、1276~1258 nm、1240~1235 nm、1194~1184 nm、1173~1168 nm、1142~1137 nm、1103~1099 nm、1080~1076 nm、1058~1054 nm、1012~1009 nm、1004~1001 nm (540 个光谱), 使用原始光谱所建模型精度影响最小。研究显示, 鸡肉赖氨酸近红外定量预测模型的建立受异常样本和建模样本、验证样本的影响较大, 建模特征光谱对模型的适用性影响较大, 光谱预处理方法仅在全谱段不做其它处理时对建模精度影响较大。

关键词: 近红外光谱; 胸肌冻干粉; 赖氨酸; 定量预测模型

文章编号: 1673-9078(2021)10-270-282

DOI: 10.13982/j.mfst.1673-9078.2021.10.0114

Establishment and Optimization of the Near Infrared Spectroscopy Quantitative Prediction Model for the Lysine in Freeze-dried Chicken Breast Muscle Powder

TAO Lin-li¹, TAO Ye², GAN Wen-bin³, DU Guang-ying¹, LI Fu-yin¹, WANG Tian-wu¹, YIN Ru-gao-fan¹,
YANG Wan-jin¹, ZHANG Xi^{1*}, NIU Guo-yi^{1*}

(1. Faculty of Animal Science and Technology, Key Laboratory of Animal Nutrition and Feed Science of Yunnan Province, Yunnan Agricultural University, Kunming 650201, China) (2. Yunnan Feed Industry Association, Kunming 650201, China) (3. Yunnan Veterinary Medicine and Forage Inspection Institution, Kunming 650201, China)

Abstract: In this study, a near infrared (NIR) quantitative prediction model for the lysine in the freeze-dried chicken breast muscle powder was developed using the partial least squares method. The model was optimized through four aspects, spectrum preprocessing, model characteristic spectrum screening, outlier samples elimination, and sample selection for modeling and verification, in order to

引文格式:

陶琳丽,陶冶,甘文斌,等.鸡胸肌冻干粉赖氨酸近红外定量预测模型的建立与优化[J].现代食品科技,2021,37(10):270-282

TAO Lin-li, TAO Ye, GAN Wen-bin, et al. Establishment and optimization of the near infrared spectroscopy quantitative prediction model for the lysine in freeze-dried chicken breast muscle powder [J]. Modern Food Science and Technology, 2021, 37(10): 270-282

收稿日期: 2021-01-01

基金项目: 国家自然科学基金项目 (31760487); 云南省重大科技专项 (2016ZA008); 云南省现代农业禽蛋产业技术体系项目 (2017KJTX0017)

作者简介: 陶琳丽 (1974-), 女, 博士, 教授, 研究方向: 动物营养与饲料科学, E-mail: tllk@qq.com

通讯作者: 张曦 (1960-), 男, 教授, 研究方向: 动物营养与饲料科学, E-mail: 943727490@qq.com; 共同通讯作者: 牛国一 (1990-), 女, 博士, 讲师, 研究方向: 动物营养与饲料科学, E-mail: 690661078@qq.com

improve the prediction accuracy and robustness of the model. Taking a total of 263 freeze-dried chicken breast muscle powders as the research object, the effects of seven different spectral preprocessing methods, four characteristic spectrum screening methods, MCCV outlier samples elimination methods, and two samples selection methods for modeling and verification, SPXY and chicken breeds, on the NIR quantitative prediction model for the lysine of in freeze-dried chicken breast muscle powder. The results showed that (1) the best model was established on the basis of 156 calibration samples and 39 external verification samples after the elimination of 68 outlier samples from the original spectra in the 1000~2502 nm; (2) The spectral region in the 1000~2502 nm used for modeling should be pretreated by SNV+gapsegment (1#,15,7) method when there was no outlier spectrum or model characteristic spectral region for selection; (3) Least impact on the accuracy of the model established using the original spectra, was exerted by outlier sample elimination and samples selection for modeling in 1386~1379 nm, 1329~1323 nm, 1289~1283 nm, 1276~1258 nm, 1240~1235 nm, 1194~1184 nm, 1173~1168 nm, 1142~1137 nm, 1103~1099 nm, 1080~1076 nm, 1058~1054 nm, 1012~1009 nm, 1004~1001 nm. Research showed that the establishment of a near-infrared quantitative prediction model for the lysine of in freeze-dried chicken breast muscle powder is greatly affected by outlier samples, samples for modeling and external sample selection for verification. The applicability of the model is greatly affected by characteristic spectra regions for modeling. The preprocessing of the whole spectrum without other pretreatments has a greater impact on the accuracy of model.

Key words: near infrared spectroscopy; chicken breast muscle freeze-dried powder; lysine; quantitative prediction model

鸡肉作为人类健康经济的蛋白质来源,一直深受消费者的喜爱,占有着重要的肉品市场份额。赖氨酸是鸡肉蛋白质的重要组成部分,是人的第一限制性氨基酸、鸡的第二限制性氨基酸。因此,检测鸡肉中赖氨酸含量对鸡肉加工利用,鸡育种、养殖技术提升都具有重要意义。

赖氨酸含量检测方法有分光光度法、气相色谱法、高相液相色谱法、液相色谱与质谱联用法、毛细管电泳法等^[1]。目前,鸡肉赖氨酸含量的检测主要使用高相液相色谱法,如柱后衍生氨基酸分析法^[2]、柱前衍生液相色谱法^[3]。这两种方法在进行赖氨酸分析时均需进行复杂的样品前处理,存在耗时长、试剂种类多、操作复杂、仪器精度要求高、检测成本高等问题,制约大批量鸡肉样本氨基酸的检测。近红外光谱定量分析技术是近年来发展迅速的一种绿色安全简便的检测技术,已被广泛的应用于各种有机物的检测。在氨基酸检测中,近红外定量预测模型已被用于水果^[4]、饲料^[5]、中药^[6]等。对于肉类检测,已有关于羊肉中氨基酸近红外光谱建模方面的报道,但近红外在鸡肉化学成分检测上的应用研究仅见于粗蛋白质、粗脂肪、脂肪酸、水分等^[7]指标,鲜见氨基酸方面的报导。不同种类来源的相同化学成分,近红外定量预测模型不能通用,因此鸡肉中赖氨酸定量预测模型的建立对鸡肉产业的发展具有十分重要的意义。

一个实用的近红外定量校正模型应具有预测结果相关性高,标准偏差低,但相关性又不能是由过拟合造成等特点。在建模过程中,一般通过四个方面来提高模型的精度与稳定性:一是通过光谱预处理方法减少以至消除各种非目标因素对光谱的影响,提高光

谱的有效信息量^[8];二是通过特征光谱的选取,提高模型的预测精度^[9];三是删除降低模型预测精度异常样本^[10];四是合理进行建模校正集和外部验证集的选择^[11]。因此,本文以五种不同来源的云南地方鸡胸肌冻干粉中的赖氨酸为研究对象,研究不同建模因素对鸡胸肌赖氨酸近红外定量校正模型建模的影响,以期获得最佳的鸡胸肌赖氨酸近红外定量校正模型,为今后鸡肉中赖氨酸的定量检测提供一种快捷准确的检测方法。

1 材料与方法

1.1 试验材料

采集 300~350 日龄、放养的云南省香格里拉县尼西鸡 48 羽(NX,公 22,母 26)、剑川县青花鸡 58 羽(QH,公 28,母 30)、南涧县无量山乌骨鸡 52 羽(LW,公 26,母 26)、新平县黎明鸡 58 羽(XP,公 30,母 28)、武定县武定鸡 47 羽(WD,公 18,母 29),5 种鸡的胸肌,制成冻干粉,作为研究样本,各品种样本记为 NXX、QHx、LWX、XPx、WDX。

1.2 赖氨酸含量的化学检测方法

采用高效液相色谱内标法进行检测(岛津高效液相色谱仪 LC-20A),赖氨酸检测方法参照天津市博纳艾杰尔科技有限公司 Venusil AA 氨基酸分析方法^[12]进行。

1.3 近红外光谱采集

将样品装入 0.6 mm 厚聚乙烯自封袋^[13],样品厚

度 5 mm。使用岛津傅立叶变换红外光谱仪 IRPrestige-21 (配套近红外附件 FlexIR™ Near-Infrared Fiber Optics module) 扫描样品, 光谱采集软件岛津 IRresolution 1.50。样品扫描前先将 0.6 mm 聚乙烯圆片置于背景金箔面上进行背景去除, 去除 PE 自封袋的影响。光谱扫描的波长范围为 1000~2502 nm, 分辨率 8 cm⁻¹, 每个样品扫描 3 次, 每次扫描 50 遍, 以 3 条近红外光谱的平均值作为样品原始光谱。

1.4 光谱预处理方法筛选

由于样品不同成分之间的相互干扰, 导致近红外光谱谱带重叠, 存在谱峰掩盖及信号噪声等问题, 一般会在获得样品近红外漫反射光谱 (Near Infrared Reflection Spectroscopy, NIRS) 后进行光谱预处理, 即消除光谱噪声和其它谱图不规则因素的影响^[8]。本文采用标准正态变量变换 (SNV)、多元散射校正 (MSC), 一阶导数[gapsegment(1#,15,7)]、二阶导数[gapsegment(2#,15,7)]、标准正态变换结合一阶导数法[SNV+gapsegment(1#,15,7)]、标准正态变换结合二阶导数法[SNV+gapsegment(2#,15,7)]、标准正态变换结合去趋势算法(SNV+Detrending)等 7 种光谱预处理方法对鸡肉 NIRS 进行光谱预处理。具体过程如下:

(1) 将 263 个样本划分为校正集 200 样本 (每种地方鸡取 40 个样本 (除 WDX 公鸡 18 个, 母鸡 22 个外, 其余四种鸡公母各半) 和外部验证集 63 样本, 样本划分时满足验证样品的赖氨酸检测值均包含在校正样品中的要求, 该样本划分方法称为鸡种样本划分法。

(2) 采用 7 种光谱预处理方法对鸡肉 NIRS 进行光谱预处理。

(3) 使用原始光谱、7 种预处理光谱, 在全谱段 (1000~2502 nm) 上采用偏最小二乘法 PLS、内部留一交互验证建立校正模型。根据模型的优劣筛选赖氨酸的光谱预处理方法, 记为 NIRS-preprocess1。

1.5 建模特征光谱筛选

随着对 NIR 结合 PLS 建模方法的深入研究和应用发现, 通过特定方法筛选特征波长或波长区间有可能得到更好的定量校正模型, 波长选择不但可以简化模型, 还可以通过剔除不相关或非线性变量, 得到预测能力强、稳健性好的校正模型^[14]。本文对特征光谱筛选的过程如下:

(1) 根据 5 种鸡胸肌冻干粉光谱形态和主成分分析 (PCA) 结果将样品 NIRS 划分成不同谱区。在这些谱区上, 采用原始光谱和 NIRS-preprocess1 光谱

进行建模。评价各谱区校正模型的优劣, 筛选出建模最佳谱区, 记为 Feature-spectrum1。

(2) 将 1000~2502 nm 全谱区划分为 155 个子区间、77 个子区间、39 个子区间、10 个子区间, Feature-spectrum1 谱区划分为 97 个子区间、40 个子区间、10 个子区间, 使用前向间隔偏最小二乘法 (Forward Interval PLS, FiPLS)^[15]筛选不同子区间上的建模特征谱区, 使用 PLS 建立校正模型。

(3) 根据 (2) 的研究结果, 在 1000~2502 nm 谱区和 Feature-spectrum1 谱区上筛选出较优的子区间划分方法各 2 个, 再在这 4 个子区间上使用后向间隔偏最小二乘法 (Backward interval PLS, BiPLS)^[16]的进行特征光谱区筛选, 在筛选出的 4 个特征光谱区间上使用 PLS 建立校正模型。

(4) 根据 (2) 的研究结果, 在 1000~2502 nm 谱区和 Feature-spectrum1 谱区上筛选出最优的建模特征光谱各 1 个, 在 FiPLS 法筛选的最优特征光谱基础上, 再使用 BiPLS 法进行特征光谱区的筛选, 记为 FBipls 法。在 FBipls 法筛选的特征光谱区间, 使用 PLS1 建立校正模型。

(5) 比较上述不同特征光谱区的建模效果, 筛选出最佳的建模光谱区域。

1.6 异常样本的剔除

异常样本是指在建模过程中加入某些样本后导致所建模型的预测精度大幅降低的样本, 主要包括 2 类, 一类是由操作误差和设备误差导致, 这类样本应视为异常样本进行剔除; 另一类是由样本本身属性的不同导致化学值和光谱值与其他样品相比具有特殊性, 加入这些样本虽会降低模型对一般样本的预测精度, 但可显著提高模型的适配范围^[11]。本文采用蒙特卡洛交叉验证法 (MCCV) 作为异常样本剔除方法, 其 matlab 算法由 Michal Daszykowski 提供网络下载 <http://www.chemometria.us.edu.pl>。

具体步骤如下:

步骤 1: 针对所有样本使用 PLS 确定最佳主成分数;

步骤 2: 使用蒙特卡洛随机取样法选取 80% 样本作为校正集, 剩余 20% 作为验证集, 根据步骤 1 所确定的最佳主成分数, 针对校正集使用 PLS 方法建立回归模型, 并对验证集进行预测, 获得验证集每个样本的预测值;

步骤 3: 将步骤 2 循环 2000 次, 得到每个样本的预测误差分布;

步骤 4: 计算每个样本预测残差的均值 (MEAN)

和方差 (STD), 绘制样本的残差均值—方差分布图, 根据绘制的均值—方差分布图, 确定在一定方差下, 不同较大残差下的样本数, 将这些样本作为异常样本。

步骤 5: 对胸肌冻干粉原始光谱按上述四个步骤建立均值—方差分布图, 根据残差均值、残差方差和剔除数量划分成 2 组具有不同数量的异常样本集 (异常样本集 1、异常样本集 2), 研究剔除不同异常样本数对模型的影响。

1.7 校正集和验证集样本的选择

校正集和验证集样本的选择方法有常规选择和计算机识别^[17]。本文上述研究过程中采用的就是常规选择法, 即根据鸡种进行选择。现增加 SPXY 算法选择校正集和验证集, SPXY 算法的优点是能有效覆盖多维向量空间, 从而改善模型预测能力。在计算样品间距离时 SPXY 算法同时将 x 变量和 y 变量同时考虑在内^[18], 其标准化的 xy 距离公式为:

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, N]} d_y(p, q)}$$

式中:

N —样品数;

x —PLS 确定的主成分;

y —化学检测值;

$d_x(p, q)$ — p 和 q 主成分的欧氏距离;

$d_y(p, q)$ — p 和 q 化学检测值的欧氏距离。

在剔除异常样本集的基础上, 根据计算的欧氏距离进行样本集和验证集的重新划分。

1.8 综合优化建模研究

以原始光谱 PLS 的主成分数为依据, 采用 MCCV 方法筛选出 2 个相同残差方差、不同残差均值的异常样本集, 在全样本集、删除异常样本集 1 和删除异常

样本集 2 的基础上, 采用 SPXY 方法划分校正集和验证集, 形成 3 个新的校正集和验证集, 针对 1.2.2 选定的建模谱段和 1.2.1 选定的光谱预处理方法, 采用 PLS 方法建立胸肌赖氨酸定量预测模型。评价建模效果, 最终筛选出鸡胸肌赖氨酸最佳校正模型。

1.9 数据处理与模型评价

采用校正决定系数 R^2_{CAL} 、校正标准偏差 SE_{CAL} 、留一法内部交互验证决定系数 R^2_{CV} 、留一法内部交互验证标准偏差 SE_{CV} 进行评价, 以最高的 R^2_{CAL} 和 SE_{CV} 确定最优校正模型。

$$R^2_{CAL} = 1 - \frac{\sum_{i=1}^n (y_{i,actual} - y_{i,predicted})^2}{\sum_{i=1}^n (y_{i,actual} - \bar{y}_{i,actual})^2}$$

$$SE_{CAL} = \sqrt{\frac{\sum_{i=1}^n (y_{i,actual} - y_{i,predicted})^2}{n-1}}$$

式中:

$y_{i,actual}$ —第 i 个样品化学检测值;

$y_{i,predicted}$ —用所建模型对第 i 个样品的预测值;

$\bar{y}_{i,actual}$ —建模集中所有样品化学检测值的平均值;

n —建模集的样品数;

R^2_{CV} 的计算公式同 R^2_{CAL} , SE_{CV} 的计算公式同 SE_{CAL} 。

使用外部验证集对校正模型进行外部验证。评价指标是预测决定系数 R^2_p 、预测标准偏差 SE_p , 验证集标准偏差与预测标准偏差的比值 RPD_p。 R^2_p 和 SE_p 的计算公式同校正集。

根据校正模型的评价指标、内部交互验证的评价指标和外部验证评价指标综合评定模型的优劣。

2 结果与讨论

2.1 胸肌冻干粉赖氨酸化学方法检测结果

表 1 胸肌冻干粉赖氨酸含量化学检测结果

Table 1 Statistics of lysine content measured by chemical method for the breast muscles freeze-dried powder

| 样本 | 样本数/个 | 最大值/% | 最小值/% | 平均值/% | 标准偏差 | 变异系数 |
|-------|-------|-------|-------|-------|------|-------|
| 校正集 | 200 | 10.07 | 6.30 | 8.41 | 0.92 | 10.98 |
| 外部验证集 | 63 | 9.90 | 6.76 | 8.34 | 0.83 | 9.89 |
| NXX | 48 | 9.18 | 7.36 | 8.10 | 0.38 | 4.66 |
| QHX | 58 | 9.77 | 7.38 | 8.45 | 0.68 | 8.07 |
| LWX | 52 | 10.07 | 8.49 | 9.63 | 0.29 | 3.03 |
| XPX | 58 | 9.52 | 7.07 | 8.34 | 0.73 | 8.71 |
| WDX | 47 | 8.13 | 6.30 | 7.32 | 0.36 | 4.97 |

根据 1.4 中的鸡种划分法对 263 个样本进行划分, 划分后的校正集和外部验证集赖氨酸含量化学方法检测结果如表 1 所示。校正集样本赖氨酸含量在

10.07%~6.30% 之间, 外部验证集样本的赖氨酸含量为 9.90%~6.76%, 包含在校正集中, 满足近红外模型的建模和验证要求。由表 1 可以看出, 无量山乌骨鸡

胸肌中赖氨酸含量较高,平均值为 9.63%,且变异系数最低 CV 仅为 3.03%;武定鸡胸肌中赖氨酸含量最低,平均值仅为 7.32%;青花鸡和新平黎明鸡胸肌中赖氨酸含量变异较大,其 CV 均大于 8%。

2.2 光谱预处理方法筛选

由表 2 可知,在全谱段 1000~2502 nm 波长范围内,不同光谱预处理方法对建模结果存在一定影响。SNV+gapsegment (1#,15,7)光谱的建模结果最佳,而 SNV+Detrending 光谱所获结果较差,SNV 和 MSC

两种光谱预处理效果相同。因此,本研究的光谱预处理方法为 SNV+gapsegment (1#,15,7)方法。陆艳婷等^[19]使用“一阶导数+多元散射校正”和“一阶导数+矢量归一化”两种光谱预处理方法建立的粳稻赖氨酸 NIRS 预测模型效果较优,其 R^2_{CAL} 均为 0.85。牛智友等^[20]使用“变量标准化,7 点平滑,二阶导数”光谱预处理方法建立的鱼粉赖氨酸 NIRS 预测模型的 R^2_{CAL} 为 0.95。上述研究用光谱预处理方法与本文结论不一致,其原因主要是样本来源不同,样本赖氨酸及其 NIRS 差异所致。

表 2 不同光谱预处理方法建模结果

Table 2 The modeling results of different spectra pretreatment methods

| 光谱预处理方法 | 主成分数 | R^2_{CAL} | SE _{CAL} | R^2_{CV} | SE _{CV} | R^2_p | SE _p | RPD _p |
|--------------------------|------|-------------|-------------------|------------|------------------|---------|-----------------|------------------|
| 原始光谱 | 6 | 0.79 | 0.4262 | 0.70 | 0.5078 | 0.50 | 0.5850 | 1.41 |
| SNV | 4 | 0.75 | 0.4588 | 0.66 | 0.5423 | 0.45 | 0.6101 | 1.35 |
| MSC | 4 | 0.75 | 0.4588 | 0.66 | 0.5423 | 0.46 | 0.6083 | 1.36 |
| Gapsegment (1#,15,7) | 8 | 0.82 | 0.3897 | 0.68 | 0.5205 | 0.56 | 0.5442 | 1.52 |
| Gapsegment (2#,15,7) | 6 | 0.75 | 0.4608 | 0.62 | 0.5736 | 0.58 | 0.5346 | 1.54 |
| SNV+gapsegment (1#,15,7) | 8 | 0.83 | 0.3820 | 0.68 | 0.5198 | 0.57 | 0.5425 | 1.52 |
| SNV+gapsegment (2#,15,7) | 6 | 0.76 | 0.4551 | 0.61 | 0.5787 | 0.45 | 0.6101 | 1.35 |
| SNV+Detrending | 4 | 0.73 | 0.4821 | 0.61 | 0.5747 | 0.53 | 0.5669 | 1.46 |

2.3 建模特征光谱筛选

2.3.1 原始光谱图和 SNV+gapsegment (1#,15,7)光谱图

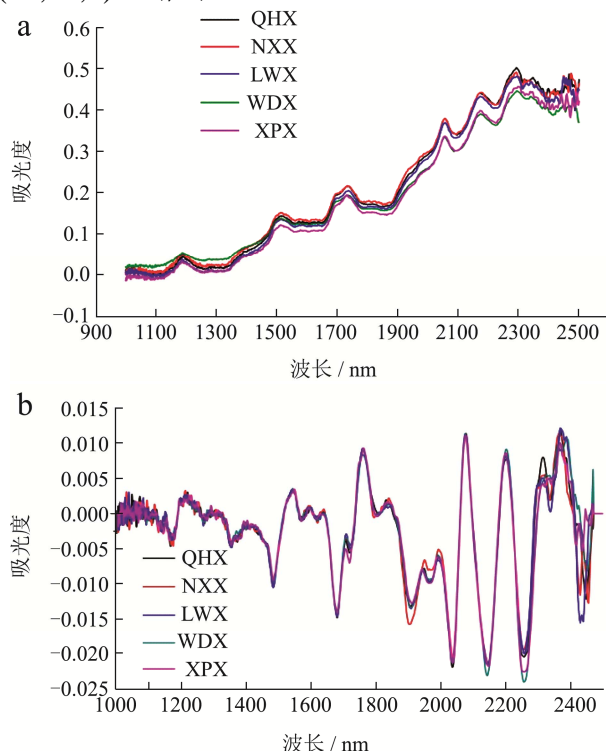


图 1 胸肌冻干粉近红外光谱图

Fig.1 NIRS of the freeze-dried breast muscles powder

注: a: 原始光谱; b: SNV+gapsegment (1#,15,7)光谱。

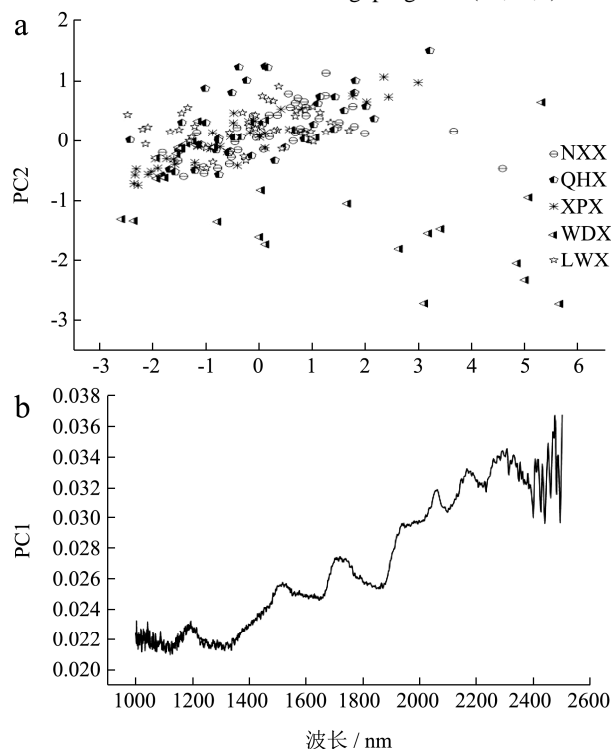


图 2 胸肌冻干粉原始光谱主成分分析

Fig.2 PCA analysis of original NIRS of the freeze-dried breast muscles powder

注: a: 得分图; b: 装载图。

由图 1a 可知,五种地方鸡胸肌冻干粉的原始光

谱相似,有6个明显的吸收峰,但在2300~2502 nm区间内表示出相对较大的差异,图1b中也表现出此特征。武定鸡在1000~1450 nm谱段的吸光度值相对较高,而在1900~2502 nm谱段武定鸡和新平黎明鸡相对其他3种鸡吸光度值较低。就整条光谱吸光度值变化来看,青花鸡和尼西鸡的吸光度值变化差异最大,武定鸡的吸光度值变化差异最小,推测吸光度值的差异与不同鸡种胸肌冻干粉中粗蛋白质的含量不同有关。研究报导,赖氨酸纯品NIRS吸收峰的谱区为1150~1250 nm、1400~1450 nm、1700~1800 nm、2100~2500 nm,强吸收峰出现在1722 nm和2122 nm,这与赖氨酸主链上存在多个 $-CH_2$ 有关^[21]。鸡腿肌冻干粉提取蛋白质的吸收峰与腿肌冻干粉的吸收峰相似,也有6个明显吸收峰^[22],与图1中的6个吸收峰1130~1275 nm、1350~1580 nm、1670~1890 nm、1900~2460 nm位置相同。因胸肌冻干粉中蛋白质含量在80%~93%之间,所以图1中冻干粉NIRS主要受蛋白质NIRS的影响。构成蛋白质的酰氨基团在1500~1530 nm附近为N-H伸缩振动的一级倍频,2050~2060 nm附近为N-H伸缩振动的组合频吸收,2168~2180 nm附近为N-H弯曲的二级倍频与C=O伸缩/N-H面内弯曲/C-N伸缩振动的组合频^[23]。 α 螺旋结构较多的蛋白质在2445 nm、2291~2281 nm、2167 nm和1738 nm区域会出现比较强的谱带, β 折叠结构为主的蛋白质则多在2463 nm、2270 nm、2210~2203 nm、2055 nm和1691~1688 nm附近出现^[24]。其中,1130~1275 nm处的吸收峰与C-H基团振动二级倍频吸收有关,1350~1580 nm处的吸收峰与2100~2500 nm与C-H基团的组合频相关^[22]。牛血清蛋白冻干粉NIRS在1570 nm、1698~1738 nm、2057 nm、2071 nm、2289~2462 nm区域也有5个明显特征吸收峰^[25]。综上,图1中1200 nm附近的吸收峰与C-H基团振动的二级倍频有关,1550 nm附近的吸收光谱主要与酰氨基团N-H伸缩振动的一级倍频、C-H基团振动的组合频有关;1700~1800 nm附近的吸收光谱主要与C-H基团振动的一级倍频有关以及蛋白质的二级结构有关;2000~2502 nm处的三个吸收峰与蛋白质的二级结构、NH基团的倍频、羧酸等相关,在2300~2502 nm谱区光谱表现尤其复杂,赖氨酸纯品的特征吸收峰均在此谱区中出现。

对胸肌冻干粉原始光谱进行PCA分析,以2个主成分为坐标建立二维得分图(图2a)。其中PC1对光谱矩阵的累积方差贡献为82%,PC2为14%,因此,样本在二维空间的分布可反应其特征。在图2a中,除武定鸡部分样品散落于得分图的下方,其余4种鸡和

一半武定鸡样品混杂为一体,聚类趋势不明显。由于PC1已包含了绝大多数分类信息,因此可根据PC1的载荷图提取近红外的特征光谱^[26]。从载荷图2b可以看出,胸肌冻干粉的特征光谱区主要在1100~1250 nm、1350~1550 nm、1650~1800 nm、1900~2100 nm、2100~2250 nm、2250~2502 nm六个光谱区间。De la haba等的研究表明混合肉糜及屠宰下脚料中氨基酸的特征谱区为1650~1794 nm和2224~2394 nm^[27]。

综合图1和图2分析结果,将1000~2502 nm波长区域划分为:1000~1299 nm、1300~1649 nm、1650~1879 nm、1880~2089 nm、2091~2229 nm、2231~2502 nm六个光谱区域,同时又根据原始光谱吸光度的高低,将全谱段划分成1600~2502 nm和1000~1599 nm两个组合区域,研究上述8个光谱区域对胸肌冻干粉赖氨酸NIRS定量模型预测精度的影响。

2.3.2 不同谱区建模效果比较

使用原始光谱和SNV+gapsegment(1#,15,7)光谱,在上述8个光谱区域使用PLS方法进行建模,建模结果如表3所示。

在表3所建的16个模型中,最优模型是1000~1599 nm谱区,采用原始光谱所建模型,其 R_{CAL}^2 为0.83, SE_{CV} 为0.5152, RPD_P 为1.70,优于表2中的最优模型(1000~2502 nm, SNV+gapsegment(1#,15,7), $R_{CAL}^2=0.83$, $SE_{CV}=0.5198$, $RPD_P=1.52$)。表3的次优模型为1000~1299 nm谱区,采用原始光谱所建模型,其 R_{CAL}^2 为0.80, SE_{CV} 为0.5399, RPD_P 为1.62。从表3可以看出,采用原始光谱所建模型均优于采用SNV+gapsegment(1#,15,7)光谱所建模型。

2.3.3 采用FiPLS、BiPLS、FBiPLS三种方法进行建模特征谱区筛选

在1000~2502 nm和1000~1599 nm上采用FiPLS、BiPLS、FBiPLS三种方法筛选建模最适特征谱区,特征谱区筛选结果如表4所示。其中,使用FiPLS方法获得7个建模谱区,BiPLS获得4个建模谱区,FBiPLS获得2个建模谱区。分别在这13个特征谱区,使用原始光谱和SNV+gapsegment(1#,15,7)光谱,应用PLS方法建立校正模型。建模结果如表5所示。

在表5中,最优模型为Lys1558-39和Lys973-97B两个特征谱区使用原始光谱所建,虽然Lys1558-155特征谱区所建模型 R_{CAL}^2 和 SE_{CV} 结果较好,但其 RPD_P 为1.41,说明模型的稳定性较差,不能作为最优模型。为更全面的研究建模谱区对模型的影响,综合考虑将Lys1558-39、Lys973-40、Lys973-97B、Lys973-40B、Lys1558-39-29B五个光谱区域定为下一步研究的建模光谱区域。

表3 胸肌冻干粉赖氨酸建模结果

Table 3 The modeling results for lysine of freeze-dried breast muscles powder

| 建模波长范围 | 光谱预处理方法 | 主成分数 | R^2_{CAL} | SE_{CAL} | R^2_{CV} | SE_{CV} | R^2_p | SE_p | RPD _p |
|--------------|-------------------------|------|-------------|------------|------------|-----------|---------|--------|------------------|
| 2231~2502 nm | 原始光谱 | 5 | 0.58 | 0.5937 | 0.45 | 0.6891 | 0.13 | 0.7680 | 1.07 |
| | SNV+gapsegment(1#,15,7) | 8 | 0.61 | 0.5743 | 0.49 | 0.6587 | 0.21 | 0.7350 | 1.12 |
| 2091~2229 nm | 原始光谱 | 4 | 0.35 | 0.7427 | 0.21 | 0.8252 | 0.13 | 0.7678 | 1.07 |
| | SNV+gapsegment(1#,15,7) | 2 | 0.16 | 0.8433 | 0.09 | 0.8819 | -0.09 | 0.8617 | 0.96 |
| 1880~2089 nm | 原始光谱 | 5 | 0.61 | 0.5757 | 0.46 | 0.6776 | 0.46 | 0.6085 | 1.36 |
| | SNV+gapsegment(1#,15,7) | 9 | 0.52 | 0.6405 | 0.30 | 0.7729 | 0.26 | 0.7101 | 1.16 |
| 1650~1879 nm | 原始光谱 | 6 | 0.69 | 0.5135 | 0.59 | 0.5922 | 0.51 | 0.5752 | 1.43 |
| | SNV+gapsegment(1#,15,7) | 4 | 0.59 | 0.5910 | 0.48 | 0.6658 | 0.46 | 0.6050 | 1.36 |
| 1300~1649 nm | 原始光谱 | 6 | 0.78 | 0.4330 | 0.67 | 0.5326 | 0.60 | 0.5201 | 1.59 |
| | SNV+gapsegment(1#,15,7) | 4 | 0.71 | 0.4969 | 0.60 | 0.5875 | 0.43 | 0.6241 | 1.32 |
| 1000~1299 nm | 原始光谱 | 6 | 0.80 | 0.4165 | 0.66 | 0.5399 | 0.62 | 0.5103 | 1.62 |
| | SNV+gapsegment(1#,15,7) | 4 | 0.69 | 0.5152 | 0.56 | 0.6115 | 0.53 | 0.5665 | 1.46 |
| 1600~2502 nm | 原始光谱 | 6 | 0.68 | 0.5172 | 0.56 | 0.6163 | 0.25 | 0.7138 | 1.16 |
| | SNV+gapsegment(1#,15,7) | 8 | 0.74 | 0.4729 | 0.60 | 0.5888 | 0.43 | 0.6206 | 1.33 |
| 1000~1599 nm | 原始光谱 | 7 | 0.83 | 0.3830 | 0.69 | 0.5152 | 0.65 | 0.4864 | 1.70 |
| | SNV+gapsegment(1#,15,7) | 4 | 0.73 | 0.4783 | 0.62 | 0.5714 | 0.55 | 0.5539 | 1.49 |

表4 FiPLS、BiPLS、FBiPLS 优选的建模光谱区域

Table 4 The optimum intervals of modeling spectra by FiPLS, BiPLS and FBiPLS

| 建模光谱 筛选方法 | 供筛选的 光谱区域/nm (波点数/个) | 优选出的建模光谱区域编号 划分的子区间数/个 每个区间的光谱数/个 | 优选出的建模光谱区域/nm (光谱数/个) |
|--------------------|----------------------------|---|--|
| FiPLS | 1000~2502 (1558) | Lys1558-155 155 10 | 2016-2002; 1970-1957; 1733-1723; 1710-1700; 1645-1636; 1604-1595; 1585-1576; 1556-1539; 1502-1486; 1476-1469; 1404-1398; 1389-1383; 1325-1319; 1286-1280; 1255-1249; 1237-1226; 1186-1181; 1082-1078; 1069-1065; 1022-1015 (250) |
| | | Lys1558-77 77 20 | 2000-1942; 1493-1428; 1397-1383; 1367-1340; 1286-1226; 1191-1181; 1159-1150; 1110-1101 (960) |
| | | Lys1558-39 39 39 | 2106-2043; 1980-1870; 1639-1600; 1562-1527; 1492-1398; 1341-1315; 1289-1219; 1175-1136; 1080-1063 (585) |
| | | Lys1558-10 10 155 | 1926-1229 (775) |
| 1000~1599 (973) | Lys973-97 97 10 | Lys973-97 97 10 | 1386-1379; 1329-1323; 1289-1283; 1276-1258; 1240-1235; 1194-1184; 1173-1168; 1142-1137; 1103-1099; 1080-1076; 1058-1054; 1012-1009; 1004-1001 (160) |
| | | Lys973-40 40 24 | 1510-1470; 1411-1359; 1325-1309; 1293-1221; 1193-1181; 1167-1155; 1064-1054 (336) |

转下页

接上页

| | | |
|------------------------------|----------------------------|---|
| | Lys973-10 10 97 | 1599-1357; 1290-1177; 1082-1040 (582) |
| 1000~2502 (1558) | Lys1558-155B 155 10 | 2481-2457; 2409-2367; 2323-2284; 2262-1313; 1305-1186; 1181-1031; 1026-1019; 1014-1003 (1450) |
| | Lys1558-39B 39 39 | 2412-2250; 2175-2108; 2041-1982; 1770-1527; 1492-1369; 1315-1290; 1265-1219; 1175-1136; 1116-1030 (936) |
| BiPLS 1000~1599 (973) | Lys973-97B 97 10 | 1599-1571; 1561-1543; 1524-1465; 1456-1409; 1401-1350; 1336-1316; 1286-1283; 1276-1235; 1189-1178; 1157-1152; 1147-1127; 1107-1099; 1084-1076; 1071-1067; 1058-1054; 1045-1041; 1028-1025; 1012-1009; 1004-1001 (540) |
| | Lys973-40B 40 24 | 1576-1309; 1293-1235; 1206-1194; 1167-1155; 1143-1131; 1085-1054 (576) |
| Lys1558-39 (585) | Lys1558-39-29B 29 20 | 2106-2074; 1979-1951; 1920-1870; 1639-1600; 1562-1527; 1492-1471; 1454-1439; 1422-1398; 1341-1323; 1271-1247; 1223-1219; 1175-1149; 1138-1136; 1080-1065 (404) |
| FBiPLS Lys973-40 (336) | Lys973-40-22B 22 15 | 1510-1498; 1487-1470; 1411-1392; 1380-1369; 1325-1309; 1293-1226; 1190-1182; 1062-1056 (240) |

2.4 异常样本的剔除

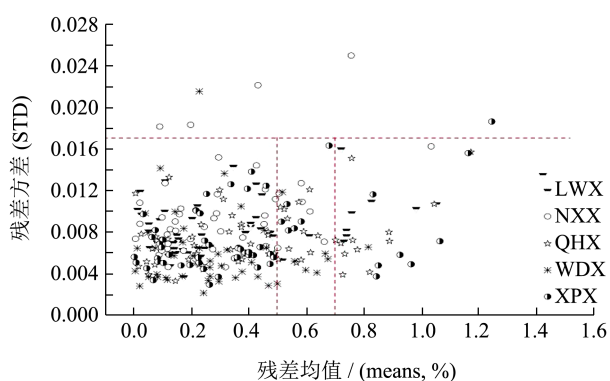


图3 胸肌冻干赖氨酸原始光谱 MCCV 法残差均值-方差图
Fig.3 The residual mean-variance scatter of MCCV based on the spectra of freeze-dried breast muscle powder and lysine values

在图3中,残差方差超过0.017以上有6个样本,其中尼西鸡样本4个,武定鸡样本1个,新平黎明鸡1个。根据残差均值和离散程序,将残差均值的判定线分别定为0.7、0.5,即异常样本集1(残差均值>0.7和方差>0.017)有36个样本,异常样本集2(残差均

值>0.5和方差>0.017)有68个样本。

2.5 综合优化建模结果

表6显示:没有删除异常样本时,使用SPXY法选择的校正样本和外部验证样本,最优模型为Lys973-97B建模谱区,使用原始光谱所建,其 $R_{CAL}^2=0.85$ 、 $SE_{CV}=0.4797$ 、 $R_p^2=0.75$ 、 $RPD_p=1.98$ 。删除异常样本集1时,3个模型结果较优,分别为1000~2502 nm使用原始光谱所建,其 $R_{CAL}^2=0.84$ 、 $SE_{CV}=0.4300$ 、 $R_p^2=0.84$ 、 $RPD_p=2.52$;Lys973-97B使用原始光谱所建,其 $R_{CAL}^2=0.87$ 、 $SE_{CV}=0.4521$ 、 $R_p^2=0.82$ 、 $RPD_p=2.39$;1000~1599 nm区域使用原始光谱所建,其 $R_{CAL}^2=0.85$ 、 $SE_{CV}=0.4626$ 、 $R_p^2=0.84$ 、 $RPD_p=2.50$ 。删除异常样本集2的建模结果也类似,最优模型为1000~2502 nm使用原始光谱所建,其 $R_{CAL}^2=0.92$ 、 $SE_{CV}=0.3284$ 、 $R_p^2=0.88$ 、 $RPD_p=2.93$;其次为Lys973-97B谱区使用原始光谱所建,其 $R_{CAL}^2=0.86$ 、 $SE_{CV}=0.4066$ 、 $R_p^2=0.86$ 、 $RPD_p=2.67$;1000~1599 nm区域使用原始光谱所建模型也较优,其 $R_{CAL}^2=0.86$ 、 $SE_{CV}=0.4114$ 、 $R_p^2=0.87$ 、 $RPD_p=2.74$ 。

表5 胸肌冻干粉赖氨酸 FiPLS、BiPLS、FBiPLS 特征光谱区的建模结果

Table 5 The modeling results for lysine of freeze-dried breast muscle powder by FiPLS, BiPLS and FBiPLS

| 光谱筛选方法 | 建模波长范围 | 建模光谱数 | 光谱预处理方法 | 主成分数 | R_{CAL}^2 | SE_{CAL} | R_{CV}^2 | SE_{CV} | R_p^2 | SE_p | RPD_p | |
|---------------|--------------|----------------|--------------------------|--------------------------|-------------|------------|------------|-----------|---------|--------|---------|------|
| Fipls | Lys1558-155 | 250 | 原始光谱 | 10 | 0.90 | 0.2841 | 0.82 | 0.3980 | 0.50 | 0.5854 | 1.41 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.72 | 0.4832 | 0.63 | 0.5641 | 0.50 | 0.5857 | 1.41 | |
| | Lys1558-77 | 960 | 原始光谱 | 8 | 0.82 | 0.3932 | 0.69 | 0.5129 | 0.61 | 0.5171 | 1.60 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.79 | 0.4245 | 0.68 | 0.5268 | 0.56 | 0.5453 | 1.51 | |
| | Lys1558-39 | 585 | 原始光谱 | 9 | 0.88 | 0.3165 | 0.75 | 0.4602 | 0.61 | 0.5157 | 1.60 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.77 | 0.4371 | 0.65 | 0.5501 | 0.60 | 0.5244 | 1.57 | |
| | Lys1558-10 | 775 | 原始光谱 | 6 | 0.79 | 0.4218 | 0.72 | 0.4899 | 0.56 | 0.5500 | 1.50 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.80 | 0.4095 | 0.71 | 0.4962 | 0.52 | 0.5733 | 1.44 | |
| | Lys973-97 | 160 | 原始光谱 | 7 | 0.82 | 0.3900 | 0.73 | 0.4787 | 0.56 | 0.5465 | 1.51 | |
| | | | SNV+gapsegment (1#,15,7) | 9 | 0.69 | 0.5138 | 0.50 | 0.6568 | 0.46 | 0.6079 | 1.36 | |
| | Lys973-40 | 336 | 原始光谱 | 8 | 0.84 | 0.3702 | 0.74 | 0.4753 | 0.61 | 0.5163 | 1.60 | |
| | | | SNV+gapsegment (1#,15,7) | 6 | 0.74 | 0.4673 | 0.59 | 0.5945 | 0.38 | 0.6513 | 1.27 | |
| | Lys973-10 | 582 | 原始光谱 | 7 | 0.83 | 0.3840 | 0.72 | 0.4905 | 0.59 | 0.5271 | 1.57 | |
| | | | SNV+gapsegment (1#,15,7) | 4 | 0.73 | 0.4741 | 0.63 | 0.5593 | 0.51 | 0.5801 | 1.42 | |
| Bipls | Lys1558-155B | 1450 | 原始光谱 | 8 | 0.83 | 0.3814 | 0.73 | 0.4836 | 0.58 | 0.5354 | 1.54 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.78 | 0.4291 | 0.67 | 0.5299 | 0.51 | 0.5778 | 1.43 | |
| | Lys1558-39B | 936 | 原始光谱 | 7 | 0.84 | 0.3715 | 0.72 | 0.4940 | 0.53 | 0.5672 | 1.45 | |
| | | | SNV+gapsegment (1#,15,7) | 6 | 0.79 | 0.4199 | 0.68 | 0.5239 | 0.51 | 0.5779 | 1.43 | |
| | Lys973-97B | 540 | 原始光谱 | 9 | 0.88 | 0.3188 | 0.76 | 0.4530 | 0.58 | 0.5322 | 1.55 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.77 | 0.4447 | 0.64 | 0.5574 | 0.56 | 0.5446 | 1.51 | |
| | Lys973-40B | 576 | 原始光谱 | 6 | 0.81 | 0.4000 | 0.72 | 0.4887 | 0.64 | 0.4930 | 1.54 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.77 | 0.4436 | 0.64 | 0.5518 | 0.55 | 0.5538 | 1.49 | |
| | FBipls | Lys1558-39-29B | 404 | 原始光谱 | 9 | 0.88 | 0.3191 | 0.76 | 0.4494 | 0.58 | 0.5374 | 1.54 |
| | | | | SNV+gapsegment (1#,15,7) | 5 | 0.73 | 0.4775 | 0.61 | 0.5778 | 0.61 | 0.5182 | 1.59 |
| Lys973-40-22B | | 240 | 原始光谱 | 7 | 0.83 | 0.3822 | 0.74 | 0.4684 | 0.60 | 0.5250 | 1.57 | |
| | | | SNV+gapsegment (1#,15,7) | 5 | 0.71 | 0.4939 | 0.59 | 0.5958 | 0.34 | 0.6696 | 1.23 | |

表 6 胸肌冻干粉赖氨酸 NIRS 定量预测模型评价结果

Table 6 The evaluate results of NIRS quantitative prediction model for the lysine of the freeze-dried breast muscle powder

| 删除的异常样本集 | 建模样本数 | 验证样本数 | 建模波长 | 光谱预处理方法 | 主成分数 | R^2_{CAL} | SE_{CAL} | R^2_{CV} | SE_{CV} | R^2_p | SE_p | RPD_p | | | |
|----------------|-------------------------|-------|----------------|-------------------------|------|--------------|-------------------------|------------|-----------|---------|--------|---------|------|--------|------|
| 无 | 210 | 53 | 1000~2502 nm | 原始光谱 | 7 | 0.78 | 0.4249 | 0.67 | 0.5255 | 0.66 | 0.5068 | 1.71 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 8 | 0.82 | 0.3809 | 0.68 | 0.5165 | 0.67 | 0.5017 | 1.73 | | | |
| | | | 1000~1599 nm | 原始光谱 | 6 | 0.78 | 0.4296 | 0.67 | 0.5240 | 0.68 | 0.4898 | 1.77 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 4 | 0.73 | 0.4750 | 0.61 | 0.5710 | 0.70 | 0.4761 | 1.82 | | | |
| | | | Lys1558-39 | 原始光谱 | 8 | 0.85 | 0.3532 | 0.72 | 0.4846 | 0.74 | 0.4428 | 1.96 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 5 | 0.75 | 0.4555 | 0.62 | 0.5611 | 0.66 | 0.5072 | 1.71 | | | |
| | | | Lys973-97B | 原始光谱 | 8 | 0.85 | 0.3544 | 0.72 | 0.4797 | 0.75 | 0.4378 | 1.98 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 5 | 0.76 | 0.4468 | 0.62 | 0.5599 | 0.67 | 0.5003 | 1.74 | | | |
| | | | Lys1558-39-29B | 原始光谱 | 8 | 0.84 | 0.3675 | 0.72 | 0.4817 | 0.71 | 0.4675 | 1.86 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 5 | 0.72 | 0.4795 | 0.60 | 0.5759 | 0.62 | 0.5383 | 1.61 | | | |
| | | | 异常样本集 1 | 182 | 45 | 1000~2502 nm | 原始光谱 | 6 | 0.84 | 0.3580 | 0.77 | 0.4300 | 0.84 | 0.3359 | 2.52 |
| | | | | | | | SNV+gapsegment(1#,15,7) | 5 | 0.79 | 0.4040 | 0.70 | 0.4881 | 0.76 | 0.4151 | 2.04 |
| 1000~1599 nm | 原始光谱 | 7 | | | | 0.85 | 0.3386 | 0.73 | 0.4626 | 0.84 | 0.3391 | 2.50 | | | |
| | SNV+gapsegment(1#,15,7) | 4 | | | | 0.75 | 0.4382 | 0.64 | 0.5361 | 0.78 | 0.3988 | 2.13 | | | |
| Lys1558-39 | 原始光谱 | 7 | | | | 0.82 | 0.3744 | 0.73 | 0.4600 | 0.82 | 0.3640 | 2.33 | | | |
| | SNV+gapsegment(1#,15,7) | 4 | | | | 0.77 | 0.4287 | 0.66 | 0.5168 | 0.71 | 0.4595 | 1.85 | | | |
| Lys973-97B | 原始光谱 | 8 | | | | 0.87 | 0.3214 | 0.74 | 0.4521 | 0.82 | 0.3554 | 2.39 | | | |
| | SNV+gapsegment(1#,15,7) | 4 | | | | 0.75 | 0.4395 | 0.64 | 0.5339 | 0.78 | 0.3987 | 2.13 | | | |
| Lys1558-39-29B | 原始光谱 | 7 | | | | 0.82 | 0.3759 | 0.74 | 0.4564 | 0.79 | 0.3902 | 2.17 | | | |
| | SNV+gapsegment(1#,15,7) | 4 | | | | 0.74 | 0.4544 | 0.64 | 0.5320 | 0.67 | 0.4875 | 1.74 | | | |
| 异常样本集 2 | 156 | 39 | | | | 1000~2502 nm | 原始光谱 | 7 | 0.92 | 0.2451 | 0.85 | 0.3284 | 0.88 | 0.3197 | 2.93 |
| | | | | | | | SNV+gapsegment(1#,15,7) | 6 | 0.88 | 0.2956 | 0.78 | 0.4019 | 0.85 | 0.3638 | 2.57 |
| | | | 1000~1599 nm | 原始光谱 | 6 | 0.86 | 0.3155 | 0.77 | 0.4114 | 0.87 | 0.3408 | 2.74 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 4 | 0.83 | 0.3553 | 0.73 | 0.4485 | 0.84 | 0.3721 | 2.51 | | | |
| | | | Lys1558-39 | 原始光谱 | 5 | 0.84 | 0.3422 | 0.79 | 0.3962 | 0.78 | 0.4393 | 2.13 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 4 | 0.81 | 0.3727 | 0.69 | 0.4806 | 0.82 | 0.3992 | 2.34 | | | |
| | | | Lys973-97B | 原始光谱 | 6 | 0.86 | 0.3182 | 0.78 | 0.4066 | 0.86 | 0.3507 | 2.67 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 4 | 0.82 | 0.3674 | 0.71 | 0.4642 | 0.84 | 0.3718 | 2.52 | | | |
| | | | Lys1558-39-29B | 原始光谱 | 7 | 0.88 | 0.3020 | 0.80 | 0.3814 | 0.78 | 0.4355 | 2.15 | | | |
| | | | | SNV+gapsegment(1#,15,7) | 4 | 0.77 | 0.4094 | 0.67 | 0.4977 | 0.79 | 0.4325 | 2.16 | | | |

对于光谱预处理的作用,从表2、表3、表6的建模结果可以看出,在没有删除异常样本时,在1000~2502 nm 建模谱区,使用 SNV+gapsegment (1#,15,7)光谱所建模型优于原始光谱所建;在1000~1599 nm 建模谱区,使用原始光谱所建模型优于 SNV+gapsegment (1#,15,7)光谱所建。在表6中,删除异常样本集1和集2后,原始光谱在5个建模谱区上所建模型的结果均优于 SNV+gapsegment (1#,15,7)光谱所建模型。

对于异常样本剔除的作用,从表6中可以看出,删除异常样本集2后在5个谱区上所建模型均优于删除异常样本集1所建模型,删除异常样本集1和集2的建模效果均优于没有删除异常样本集,尤其表现在外部验证集中。

对于建模样本选取方法的作用,展晓日等^[28]比较了随机取样法, Kennard-Stone(KS)、duplex、SPXY 四种训练集取样法对橘叶橙皮苷 NIRS 模型的影响,结果表明, SPXY 法选取的训练集建立的模型优于其他三种方法。在本研究中,从表2、表3、表5、表6的建模结果可以看出,删除异常样本后, SPXY 取样法优于随机取样法(鸡种取样)。但在没有删除异常样本时,鸡种选择法的校正建模结果和交互验证结果均优于 SPXY 法的建模效果,但 SPXY 法外部验证和模型稳健性的结果要优于鸡种选择法的建模结果。因此,异常样本的剔除还需做深入研究。

对于特征光谱的作用,从表3和表5可以看出,使用 FiPLS 在全谱段筛选的建模谱区所建模型的校正建模结果优于按光谱特征筛选的建模谱区,但外部验证和模型稳健性的结果低于按光谱特征筛选的建模谱区。在按光谱特征筛选的建模谱区上使用 BiPLS 法筛选的建模谱区要优于 FiPLS 法。结合表6可以看出, Lys973-97B 建模谱区相对于其他谱区建模结果的波动性最小,综合考虑到模型建模的难易度、精确性和适用性等方面,认为胸肌赖氨酸的建模特征谱区为 Lys973-97B。Qu 等^[29]使用 FiPLS、BiPLS、FBiPLS 三种方法建立水中乙醇含量的定量预测模型,发现 FBiPLS 能改善 FiPLS 与 BiPLS 贪婪搜索的特性,对特征波段的选取更高效、更具代表性。陶琳丽等^[22]使用 FiPLS、BiPLS、FBiPLS 三种方法筛选鸡腿肌冻干粉粗蛋白质 NIRS 模型的建模谱段,结果也表明 FBiPLS 法优选出的建模谱段建模结果最优。上述两项研究结果与本研究不一致。而仇逊超等^[30]利用 BiPLS 和无信息变量消除法选取特征波段,建立全波段和特征波段下的带壳松子和松仁蛋白质近红外预测模型,结果表明 BiPLS 法筛选的谱段建模结果最优,与本研

究结果一致。

3 结论

根据以上研究可以看出,鸡肉赖氨酸近红外预测模型的建立受异常样本和建模样本集的选择影响较大,特征光谱的选择对模型的适用性影响较大,光谱预处理仅在全谱段建模不做其它处理时表现出较好的效果。

(1) 在本研究中,最优胸肌冻干粉赖氨酸 NIRS 模型为使用 156 个校正样本,39 个外部验证样本,在 1000~2502 nm 使用原始光谱所建模型,其 R_{CAL}^2 为 0.92、 SE_{CV} 为 0.3284、 R_p^2 为 0.88、 RPD_p 为 2.93。

(2) 没有进行异常样本剔除、特征建模谱区选择等处理时,胸肌冻干粉赖氨酸全谱段 NIRS 建模光谱需进行 SNV+gapsegment (1#,15,7)方法预处理。

(3) 建模谱区为 Lys973-97B (540 个光谱数据)时,使用原始光谱所建模型的适用性最强。

参考文献

- [1] 林太凤,赵梦柯,王妍鸣,等.氨基酸检测技术研究进展[J].安徽农业科学,2015,43(18):16-19
LIN Tai-feng, ZHAO Meng-ke, WANG Yan-ming, et al. Research progress on the detection technology of amino acids [J]. Journal of Anhui Agricultural Science, 2015, 43(18): 16-19
- [2] 吴琼,孙艳发,夏李意,等.不同雉鸡肌肉营养成分分析和评价[J].动物营养学报,2020,32(1):155-166
WU Qiong, SUN Yan-fa, XIA Li-yi, et al. Analysis and evaluation of nutritional components in muscle of different pheasants [J]. Chinese Journal of Animal Nutrition, 2020, 32(1): 155-166
- [3] 梅凤艳,曹志勇,杨秀娟,等.不同性别武定鸡腿肌和胸肌中氨基酸含量比较分析[J].中国家禽,2016,38(21):11-15
MEI Feng-yan, CAO Zhi-yong, YANG Xiu-juan, et al. Comparative analysis of amino acid content between leg muscle and breast muscle in Wuding chicken [J]. China Poultry, 2016, 38(21): 11-15
- [4] Fernández-Novales J, Garde-Cerdán T, Tardáguila J, et al. Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening [J]. Talanta, 2019, 199: 244-253
- [5] González-Martín I, Álvarez-García N, González-Cabrera J M. Near-infrared spectroscopy (NIRS) with a fibre-optic probe for the prediction of the amino acid composition in animal feeds [J]. Talanta, 2006, 69(3): 706-710
- [6] Lin H, Chen Q S, Zhao J W, et al. Determination of free amino

- acid content in *Radix pseudostellariae* using nearinfrared (NIR) spectroscopy and different multivariate calibrations [J]. Journal of Pharmaceutical and Biomedical Analysis, 2009, 50: 803-808
- [7] 何鸿举,朱亚东,王慧,等.近红外光谱技术在生鲜禽肉质量检测中应用的研究进展[J].食品科学,2019,40(21):325-331
HE Hong-ju, ZHU Ya-dong, WANG Hui, et al. Recent advances in application of near-infrared spectroscopy for quality detection in fresh poultry meat [J]. Food Science, 2019, 40(21): 325-331
- [8] 马兰,夏俊芳,张战锋,等.光谱预处理对近红外光谱无损检测番茄可溶性固形物含量的影响[J].华中农业大学学报,2008, 27(5):672-675
MA Lan, XIA Jun-fang, ZHANG Zhan-feng, et al. Effects of spectral pretreatment on nondestructive evaluation of soluble solids content of tomatoes with near infrared spectroscopy [J]. Journal of Huazhong Agricultural University, 2008, 27(5): 672-675
- [9] Zou X B, Zhao J W, Malcolm J W P, et al. Variables selection methods in near-infrared spectroscopy [J]. Analytica Chimica Acta, 2010, 667: 14-32
- [10] 刘翠玲,胡玉君,吴胜男,等.近红外光谱奇异样本剔除方法研究[J].食品科学技术学报,2014,32(5):74-79
LIU Cui-ling, HU Yu-jun, WU Sheng-nan, et al. Outlier sample eliminating methods for building calibration model of near infrared spectroscopy analysis [J]. Journal of Food Science and Technology, 2014, 32(5): 74-79
- [11] 李彦周,闵顺耕,刘霞.主成分分析在近红外定量分析校正集样本优选中的应用[J].分析化学研究简报,2007,35(9):1331-1334
LI Yan-zhou, MIN Shun-geng, LIU Xia. Application of principal component analysis in calibration sample selection of near-infrared quantitative model [J]. Chinese Journal of Analytical Chemistry, 2007, 35(9): 1331-1334
- [12] 岳中花,张景然,张建柱,等.高效液相色谱法测定饲料中的氨基酸[J].饲料工业,2012,23:44-47
YUE Zhong-hua, ZHANG Jing-ran, ZHANG Jian-zhu, et al. Determination of amino acids in feedstuffs by high performance liquid chromatography (HPLC) [J]. Feed Industry, 2012, 23: 44-47
- [13] 皮付伟,王加华,孙旭东,等.基于聚乙烯膜包装奶酪成分的NIRS检测研究[J].光谱学与光谱分析,2008,10:2321-2324
PI Fu-wei, WANG Jia-hua, SUN Xu-dong, et al. Study on component detection of cheese wrapped by polyethylene film based on NIRS [J]. Spectroscopy and Spectral Analysis, 2008, 10: 2321-2324
- [14] 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,2004,16(4):528-542
CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique [J]. Progress in Chemistry, 2004, 16(4): 528-542
- [15] Norgaard L, Saudland A, Wagner J, et al. Interval partial leastsquares regression (iPLS): a comparative chemometric study with an example from near - infrared spectroscopy [J]. Applied Spectroscopy, 2000, 54(3): 413-419
- [16] Leardi R, Norgaard L. Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions [J]. Journal of Chemometrics, 2004, 18(11): 486-497
- [17] 尹宝全,史银雪,孙瑞志,等.近红外多组分分析中异常样本识别方法[J].农业机械学报,2015,46, (S1):122-127
YIN Bao-quan, SHI Yin-xue, SUN Rui-zhi, et al. Outlier samples detection method for NIR multicomponent analysis [J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46, (S1): 122-127
- [18] Galvão R K H, Araujo M C U, José G E, et al. A method for calibration and validation subset partitioning [J]. Talanta, 2005, 67(4): 736-740
- [19] 陆艳婷,张小明,叶胜海,等.粳稻七种必需氨基酸含量近红外漫反射光谱分析技术研究[J].核农学报,2007,21(5):478-482
LU Yan-ting, ZHANG Xiao-ming, YE Sheng-hai, et al. Measuring of seven by near amino acid contented in Japonica rice infrared spectroscopy technique [J]. Journal of Nuclear Agricultural Sciences, 2007, 21(5): 478-482
- [20] 牛智有,韩鲁佳.鱼粉中氨基酸近红外光谱定量分析[J].农业机械学报,2007,38(5):114-117
NIU Zhi-you, HAN Lu-jia. Quantitative analysis of amino acid in fishmeal by nearinfrared reflectance spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(5): 114-117
- [21] 陶琳丽,黄伟,杨秀娟,等.20种氨基酸近红外光谱及其分子结构的相关性[J].光谱学与光谱分析,2016,36(9):2766-2773
TAO Lin-li, HUANG Wei, YANG Xiu-juan, et al. Correlations between near infrared spectra and molecular structures of 20 standard amino acids [J]. Spectroscopy and Spectral Analysis, 2016, 36(9): 2766-2773
- [22] 陶琳丽,杨秀娟,邓君明,等.基于光谱波段优化的鸡肉冻干粉粗蛋白近红外定量预测模型研究[J].现代食品科技,2019,35 (8):236-246

- TAO Lin-li, YANG Xiu-juan, DENG Jun-ming, et al. Quantitative prediction model of the crude protein content in the chicken freeze-dried powder based on the optimizing spectral region of near infrared spectroscopy [J]. *Modern Food Science and Technology*, 2019, 35(8): 236-246
- [23] Workman J, Weyer L. 褚小立, 许育鹏, 田高友, 译. 近红外光谱解析实用指南[M]. 北京: 化学工业出版社, 2009
- Workman J, Weyer L. Translated by CHU Xiao-li, XU Yu-peng, TIAN Gao-you. *Practical Guide to Interpretive Near-Infrared Spectroscopy* [M]. Beijing: Chemical Industry Press, 2009
- [24] 江艳, 武培怡. 近红外光谱在蛋白质和含酰胺基团聚合物研究中的应用[J]. *化学进展*, 2008, 20(12): 2021-2033
- JIANG Yan, WU Pei-yi. Application of near infrared spectroscopy in the study of protein and polymers with amide group [J]. *Progress in Chemistry*, 2008, 20(12): 2021-2033
- [25] Izutsu K, Fjjimaki Y, Kuwabara A, et al. Near-infrared analysis of protein secondary structure in aqueous solutions and freeze-dried solids [J]. *Journal of Pharmaceutical Sciences*, 2006, 95(4): 781-789
- [26] 孙潇, 史岩. 近红外光谱技术对加工后鸡肉产地溯源的研究[J]. *现代食品科技*, 2015, 31(6): 315-321
- SUN Xiao, SHI Yan. Determining the geographic origin of cooked chicken based on near-infrared spectroscopy [J]. *Modern Food Science and Technology*, 2015, 31(6): 315-321
- [27] De la Haba M J, Garrido-Varo A, Guerrero-Ginel J E, et al. Near-infrared reflectance spectroscopy for predicting aminoacids content in intact processed animal proteins [J]. *Journal of Agricultural and Food Chemistry*, 2006, 54: 7703-7709
- [28] 展晓日, 朱向荣, 史新元, 等. SPXY 样本划分法及蒙特卡罗交叉验证结合近红外光谱用于橘叶中橙皮苷的含量测定[J]. *光谱学与光谱分析*, 2009, 29(4): 964-968
- ZHAN Xiao-ri, ZHU Xiang-rong, SHI Xin-yuan, et al. Determination of hesperidin in tangerine leaf by near-infrared spectroscopy with SPXY algorithm for sample subset partitioning and Monte Carlo cross validation [J]. *Spectroscopy and Spectral Analysis*, 2009, 29(4): 964-968
- [29] Qu F F, Ren D, Hou J J, et al. The characteristic spectral selection method based on forward and backward interval partial least squares [J]. *Spectroscopy and Spectral Analysis*, 2016, 36(2): 593-598
- [30] 仇逸超, 曹军. 近红外光谱波段优化在东北松子蛋白质定量检测中的应用[J]. *现代食品科技*, 2016, 32(11): 303-309
- QIU Xun-chao, CAO Jun. Application of band optimization of near-infrared spectra for quantitative detection of proteins in northeastern pine nuts [J]. *Modern Food Science and Technology*, 2016, 32(11): 303-309

(上接第 219 页)

- [22] 胡坪华, 邱莉苹, 宁方红, 等. 氨基表面修饰有序介孔硅对槲皮素的载药性能及药物释放影响研究[J]. *中国药学杂志*, 2013, 48(11): 894-898
- HU Ping-hua, QIU Li-ping, NING Fang-hong, et al. Inclusion of quercetin in amino-group modified ordered mesoporous silica: drug loading and release properties [J]. *Chinese Pharmaceutical Journal*, 2013, 48(11): 894-898
- [23] 荣家闵, 张仁文. 槲皮素体外抗氧化活性研究[J]. *科学技术创新*, 2019, 5: 43-44
- RONG Jia-min, ZHANG Ren-wen. Study on the antioxidation activity of quercetin *in vitro* [J]. *Scientific and Technological Innovation*, 2019, 5: 43-44
- [24] 聂鲁美, 张俊计, 陈积世, 等. 单分散二氧化硅微球的制备及反应机理[J]. *陶瓷学报*, 2010, 31(1): 75-78
- NIE Lu-mei, ZHANG Jun-ji, CHEN Ji-shi, et al. Preparation and formation mechanisms of monodispersed silicon dioxide microspheres [J]. *Journal of Ceramics*, 2010, 31(1): 75-78
- [25] 刘建红, 杨儒, 刘家祥, 等. 有序介孔材料的合成及表征方法[J]. *材料导报*, 2002, 16(9): 52-55
- LIU Jian-hong, YANG Ru, LIU Jia-xiang, et al. Syntheses and characterization of ordered mesoporous materials [J]. *Materials Reports*, 2002, 16(9): 52-55
- [26] Yang H, Zheng K, Zhang Z. Adsorption and protection of plasmid DNA on mesoporous silica nanoparticles modified with various amounts of organosilane [J]. *Journal of Colloid and Interface Science*, 2012, 369(1): 317-322
- [27] Lee G, Lee S, Jeong S, et al. Antioxidative and anti-inflammatory activities of quercetin-loaded silica nanoparticles [J]. *Colloid and Surfaces B Biointerfaces*, 2016, 143: 511-517