

基于近红外光谱技术建立湿加松针叶黄芪苷含量的预测模型*

彭冠明¹ 吕欣欣² 毛积鹏^{1,2} 欧惠玲¹
谢诺¹ 李福明¹

(1. 台山市红岭种子园, 广东 台山 529223; 2. 华南农业大学 林学与风景园林学院 / 广东省森林植物种质创新与利用重点实验室, 广东 广州 510642)

摘要 研究使用 DA2700 型近红外光谱仪采集了 112 个湿加松 *Pinus elliottii* × *P. oaribaea* 松针粉末样本的光谱数据。结合实际测定值, 采用偏最小二乘 (PLS) 回归法并选择最佳光谱预处理方法和最佳主成分数, 建立湿加松松针黄芪苷含量的近红外快速预测模型。结果表明: 当采用一阶导数 (FD) + 标准正态变量转换法 (SNV) 对光谱数据进行预处理, 主成分数为 6, 此时模型的预测效果最好, 校正集相关系数 (R_C) 和交互验证集相关系数 (R_V) 分别为 0.808 2 和 0.710 9。校正集均方根误差 (RMSEC) 和交互验证集均方根误差 (RMSEV) 分别为 1.931 4 和 2.398 8, 说明模型的预测效果较好。利用外部验证集对模型进行验证, 得到模型的外部验证相关系数 $R=0.812 9$, 预测均方根误差 RMSEP=2.973 8。

关键词 湿加松; 黄芪苷; 近红外; 预测模型

中图分类号: S718.3 文献标志码: A 文章编号: 2096-2053 (2022) 01-0062-06

Prediction Model of Astragalin Content in *Pinus elliottii* × *P. caribaea* Needles Based on Near Infrared Spectroscopy

PENG Guanming¹ LYU Xinxin² MAO Jipeng^{1,2} OU Huiling¹
XIE Nuo¹ LI Fuming¹

(1. Taishan Hongling Seed Orchard, Jiangmen, Guangdong 529223, China; 2. Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm/College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, Guangdong 510642, China)

Abstract In this study, the DA2700 near infrared spectrometer was used to collect the spectral data of 112 wet pine needle powder samples. Combined with the actual measured value, the near infrared rapid prediction model of astragalin content in wet pine needles was established by partial least squares (PLS) regression method and selecting the best spectral pretreatment method and the best principal component fraction. The results show that the prediction effect of the model is the best using the combination of first derivative (FD) and standard normal variable transformation (SNV) method to preprocess the spectral data, and when the principal component fraction is 6, the correlation coefficient of correction set (R_C) and cross validation set (R_V) was 0.808 2 and 0.710 9 respectively. The RMSEC and RMSEV was 1.931 4 and 2.398 8, respectively, indicating that the prediction effect of the model was better. The external validation set was used to verify the model, and the correlation coefficient of

基金项目: 台山市红岭国家湿地松、杂交松良种基地 2020 年中央财政林木良种繁育补助项目。

第一作者: 彭冠明 (1970—), 男, 工程师, 主要从事林木遗传育种, E-mail: tsshlyzy@126.com。

通信作者: 李福明 (1974—), 男, 高级工程师, 主要从事林木遗传育种, E-mail: 893869588@qq.com。

external validation was $R=0.8129$ with the root mean square error of prediction was $RMSEP = 2.9738$.

Key words *Pinus elliottii* × *P. caribaea*; astragaloside; near infrared spectroscopy; prediction model

《本草纲目》中有关于松针入药的记载“细切，以水及面饮服之，或捣屑丸服，可断谷及治恶疾”。通过现代科学技术发现，松针中含有包括挥发油、色素及维生素、氨基酸、蛋白质类、黄酮类、莽草酸、可溶性糖类、木质素等多种化学成分^[1]，其中黄酮类化合物具有降血糖^[2]、调血脂^[3]、抗氧化^[4]、抑菌^[5]、消炎^[6-7]等作用。随着生活水平的提高，导致现实生活中人们被各种疾病所困扰，对具有保健功能的纯天然绿色保健品需求也越来越大，作为松树的主要副产品之一，松针在保健品中的应用前景非常广阔。黄芪苷是松针中多种黄酮类化合物，是多种中药材的重要有效成分之一^[8]，主要的提取方法是通过溶剂提取法^[9]。随着提取工艺的不断改进，超声波辅助提取法、减压内部沸腾法等方法被应用于松针黄酮类化合物的提取过程中，结果表明可以提高松针黄酮的得率^[10-11]。

传统的植物黄酮类化合物含量测定需要借助于高效液相色谱-质谱(HPLC-MS)、气相色谱-质谱(GC-MS)、紫外分光光度计法等，然而这些提取方法都有一定的局限性，如需要用到具有污染性的化学试剂、成本高、耗时长且对受测样品造成了一定的损坏。近红外光谱分析技术是一项绿色环保的分析技术，具有速度快、成本低、易操作等特点，在农业、食品等领域的研究和应用十分广泛^[12-13]，例如紫花苜蓿 *Medicago sativa* 品质的评价^[14]，桔梗中总皂苷的快速检测^[15]，食用产品玛咖粉的真假鉴定^[16]等。近红外光谱技术在植物有效成分含量的预测方面有大量的成功案例，如朱斌等^[17]利用近红外光谱分析技术建立了能够定量测定金丝桃 *Hypericum monogynum* 药材中的金丝桃苷与贯叶金丝桃素的近红外预测模型；黄倩倩等^[18]采用偏最小二乘算法建立了近红外漫反射光谱技术快速测定中药黄芩 *Scutellaria baicalensis* 中主要活性成分黄芩苷和总黄酮含量的方法。

湿加松 *Pinus elliottii* × *P. oaribaea* 是松科松属的常绿乔木，是湿地松与加勒比松的杂交后代，又称杂种松、杂交松，它综合了其双亲干型通直、

生长迅速、适应性强、产脂量高等优点，湿加松还是世界上发达国家商品林基地的主要树种之一^[19-20]。本研究旨在建立一种能够快速预测湿加松针叶中黄芪苷成分含量的近红外模型，以期为湿加松针叶黄酮类成分含量的测定提供更加快捷有效的方法，为高黄酮类含量的湿加松选育工作提供技术基础。

1 材料与方法

1.1 试验材料

以种植于广东省台山市红岭种子园的6 a 生湿加松子代测定林为研究材料，于2019年8月分别随机采取了112个生长旺盛且无病虫害单株的当年生松针组织部位鲜样约100 g。75 °C条件下烘72 h，充分粉碎后过60目筛，得112份湿加松松针粉末样本。随机选择其中94份样本作为预测集，其余18份样本作为外部验证集。

1.2 样本制备与标准品配制

各样本分别称取100 mg于10 mL EP管中，随后加入5 mL 甲醇溶液，充分混匀后在室温下超声60 min，4 °C条件下，12000 rpm离心12 min。取上清液浓缩至干后加200 μL 甲醇复溶，混匀后取上清液过0.22 μm 微孔滤膜，滤液保存于进样瓶中用于LS-MS分析。称取黄芪苷的标准品适量，并与80%的甲醇溶液混合制成母液。量取适量母液，用80%甲醇稀释至合适浓度，制成工作标准液置于-80 °C待LC-MS分析。

1.3 色谱和质谱数据采集

色谱条件：ACQUITY UPLC®BEH C18 色谱柱(2.1 × 100 mm, 1.7 μm, 美国 Waters 公司)，进样量5 μL，柱温40 °C，流动相A为0.1%甲酸，流动相B为甲醇，流速设置为0.25 mL/min。梯度洗脱条件为0~1 min, 10% B; 1~3 min, 10%~33% B; 3~10 min, 33% B; 10~15 min, 33%~50% B; 15~20 min, 50%~90% B; 20~21 min, 90% B; 21~22 min, 90%~10% B; 22~25 min, 10% B。

质谱条件：电喷雾电离(ESI)源，负离子电离模式。离子源温度500 °C，离子源电压4500 V，

碰撞气 6 psi, 气帘气 30 psi, 雾化气和辅助气均为 50 psi。采用多重反应监测 (MRM) 进行扫描。

1.4 近红外光谱数据收集

采用 DA7200 近红外光谱分析仪扫描样品, 收集近红外光谱数据, 首先将仪器室温度控制在 22 ℃ 左右, 并将样本提前于仪器室放置 24 h。开机后预热 30 min, 近红外光光斑直径为 3.5 cm, 近红外光谱扫描范围为 950~1 650 nm, 仪器分辨率为 5 nm, 光谱采集速度约为 100 条光谱 /s, 扫描次数为 100 次 /s。每一个松针粉样品扫描 3 次和重复装样 3 次, 取其平均近红外光谱作为样品的原始光谱。

1.5 近红外建模方法

通过采用偏最小二乘法回归 (Partial Least Square Regression, PLS) 法对近红外光谱数据进行分析, 能够对光散射和其它组分带来的干扰做出补偿, 比对逐个因变量做多元回归更有效。适用于复杂体系, 而且模型比较稳定、分析性能和预测精度更好。因此, 本研究采用偏最小二乘法对实验样品进行近红外建模。

1.6 模型评价方法

模型的预测准确性从两个方面评价, 一方面是相关系数, 包括校正集相关系数 (Related coefficient of calibration, R_c) 和交互验证集相关系数 (Related coefficient of validation, R_v), 校正集均方根误差 (Root Mean Square Error of Calibration, RMSEC) 和交互验证集均方根误差 (Root Mean Square Error of validation, RMSEV)。所建模型应具有较高的 R_c 和 R_v , 较低的 RMSEC 和 RMSEV, 其中主要以 R_v 和 RMSEV 作为判断标准。另一方面, 模型应该具有较高的相关系数, 即预测值与实测值的相关系数 (Related coefficient, r) 和较低的预测均方根误差 (Root Mean Square Error of prediction, RMSEP)。

1.7 数据处理

利用 Microsoft Excel 2019 和 R 3.5.1 软件对各样本中黄芪苷含量进行统计分析, 使用软件 The Unscrambler 9.7 对近红外光谱图进行处理和模型建立。

2 结果与分析

2.1 松针粉黄芪苷实测值的数据特征

利用 LC-MS 对所有样本中的黄芪苷进行定量检测, 各数据集样本黄芪苷含量的统计结果如表 1 所示。94 个校正集样本黄芪苷含量的最小值为 $3.67 \mu\text{g} \cdot \text{g}^{-1}$, 最大值为 $27.67 \mu\text{g} \cdot \text{g}^{-1}$, 平均值为 $10.24 \mu\text{g} \cdot \text{g}^{-1}$, 标准偏差为 $4.99 \mu\text{g} \cdot \text{g}^{-1}$ 。18 个验证集样本黄芪苷含量的最小值为 $4.84 \mu\text{g} \cdot \text{g}^{-1}$, 最大值为 $17.11 \mu\text{g} \cdot \text{g}^{-1}$, 平均值为 $10.74 \mu\text{g} \cdot \text{g}^{-1}$, 标准偏差为 $4.11 \mu\text{g} \cdot \text{g}^{-1}$ 。

2.2 不同光谱预处理方法的比较

本研究比较了归一化 (Normalization)、多元散射校正 (Multiplicative Scattering Correction, MSC)、标准正态变量转换法 (Standard Normal Variate Transformation, SNV)、一阶导数 (First Deviation, FD)、滤波拟合 (Savitzky golay, SG) 以及 FD+SG 与 FD+SNV 共 7 种光谱预处理方法。并以校正集相关系数、校正集均方根误差、交互验证集相关系数及交互验证集均方根误差为评价指标, 其中 R_c 在 4 个参数中起着主导作用, R_c 和 R_v 越大越好, RMSEC 和 RMSEV 越小越好^[21]。从表 2 可以看出, SNV、SG、FD+SG 预处理下的 R_c 都小于 0.8, 偏低; 在 R_c 高于 0.8 的 4 个预处理方法中, FD+SNV 的 R_c 值最大, RMSEC 和 RMSEV 均为最小, 模型最好。

对比无预处理光谱图 (图 1), 经过 FD+SNV 预处理的光谱图中 (图 2), 光谱分布更加集中, 吸收峰更加明显。这表明, 经过预处理后, 近红外光谱数据能够更好的反应样品中成分含量的近

表 1 湿加松针叶校正集和验证集样本黄芪苷含量的基本统计特征

Table 1 Descriptive statistics about content of astragaloside of calibration set and validation set of *P. elliotii* × *P. caribaea* needles

| 数据集 Data set | 样品数 Number of samples | 最小值 / ($\mu\text{g} \cdot \text{g}^{-1}$) Minimum | 最大值 / ($\mu\text{g} \cdot \text{g}^{-1}$) Maximum | 平均值 / ($\mu\text{g} \cdot \text{g}^{-1}$) Average value | 标准偏差 Standard deviation |
|-----------------|--------------------------|--|--|--|----------------------------|
| 校正集 | 94 | 3.67 | 27.67 | 10.24 | 4.99 |
| 验证集 | 18 | 4.84 | 17.11 | 10.74 | 4.11 |

表 2 不同光谱预处理方法下湿加松针叶黄芪苷含量预测模型的参数比较

Table 2 Comparison of parameters of content of astragaloside prediction model in *P. elliottii* × *P. caribaea* needles under different spectral pretreatment methods

| 预处理方法 Preprocessing methods | 主成分数 Principal component number | 校正集相关 系数 R_c | 校正集均方根误差 RMSEC | 交互验证集相关系 数 R_v | 交互验证集均方根 误差 RMSEV |
|-----------------------------------|---------------------------------------|----------------------|-------------------|------------------------|-------------------------|
| 归一化 | 11 | 0.803 0 | 2.103 7 | 0.633 5 | 2.860 2 |
| MSC | 6 | 0.802 8 | 2.170 8 | 0.756 5 | 2.455 4 |
| SNV | 6 | 0.790 2 | 2.226 0 | 0.739 5 | 2.508 1 |
| SG | 7 | 0.754 2 | 2.372 4 | 0.690 0 | 2.702 1 |
| FD | 5 | 0.804 3 | 2.180 6 | 0.745 0 | 2.543 0 |
| FD+SG | 5 | 0.785 4 | 2.102 5 | 0.725 3 | 2.398 6 |
| FD+SNV | 6 | 0.808 2 | 1.931 4 | 0.710 9 | 2.398 8 |

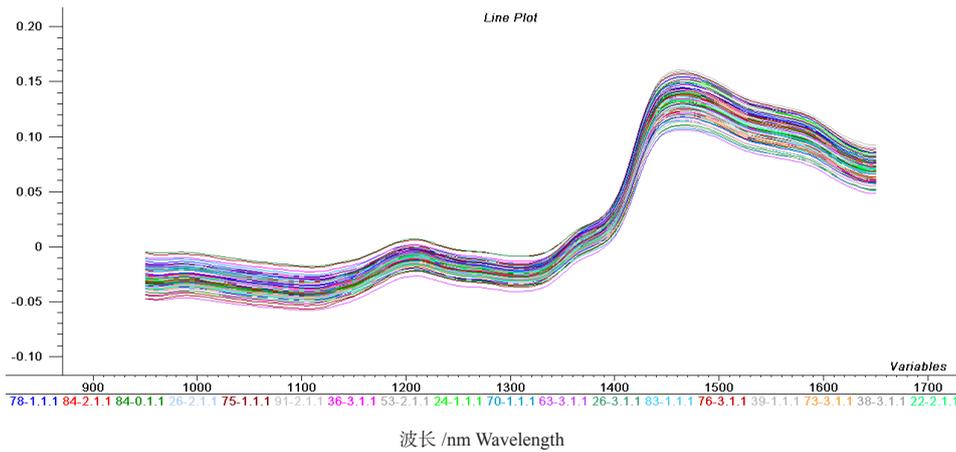


图 1 94 个湿加松针叶校正集样品近红外原始反射光谱

Figure 1 Near infrared original reflection spectrum of 94 correction set samples of *P. elliottii* × *P. caribaea* needles

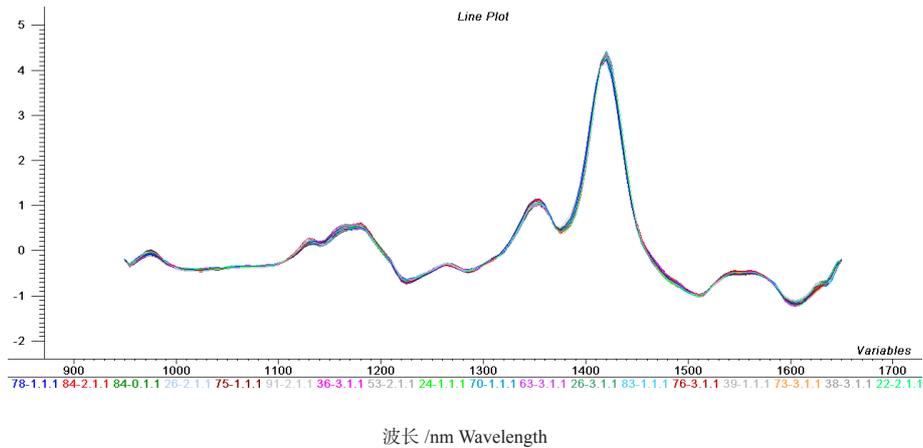


图 2 FD+SNV 法处理湿加松松针黄芪苷含量预测模型近红外光谱图

Figure 2 Near infrared spectrum of astragaloside content prediction model of wet plus pine needle treated by FD + SNV method

红外吸收特点，对于近红外预测模型的建立有更好的优化效果。

2.3 黄芪苷定量分析模型的建立

将近红外光谱数据导入软件 The Unscram-

bler 9.7, 进行光谱数据预处理和模型建立, 采用 FD+SNV 预处理, 950~1 650 nm 建模波段, 主成分数为 6, 建模结果如图 3 所示。 $R_c=0.808 2$, $RMSEC=1.931 4$, $R_v=0.710 9$, $RMSEV=2.398 8$ 。黄芪苷含量校正集相关系数最大, 校正集和交互验证集均方根误差均最低, 说明模型预测性能较好。

2.4 近红外模型的外部验证

为检验预测模型的可靠性, 本研究利用所建预测模型对 18 个验证集样本的黄芪苷含量进行了预测并与实际测定值进行比较分析(图 4)。结果表明, 验证集样本的预测值与实际测定值之间存在绝对误差, 但最大为 6.21, $R=0.812 9$, $RME-SP=2.973 8$, 接近于 $RMSEC$, 表明模型预测效果

好, 可信度高。

3 讨论

本研究基于近红外光谱技术的应用, 结合偏最小二乘法建立了湿加松松针黄芪苷含量近红外快速预测模型, 其 R_c 值为 0.808 2, 利用外部验证集来对模型进行验证, 得到 $r=0.812 9$ 。吴宪等^[22]根据尾叶桉拉丁名木材密度光谱特征值建立了尾叶桉木材密度近红外预测模型, 该模型的相关系数达到 0.83, 具有较高的检测精度。本研究与前人的研究结果相近, 说明该模型能够准确预测湿加松松针黄芪苷的含量。

湿加松的研究多集中在其产脂量的和木材性

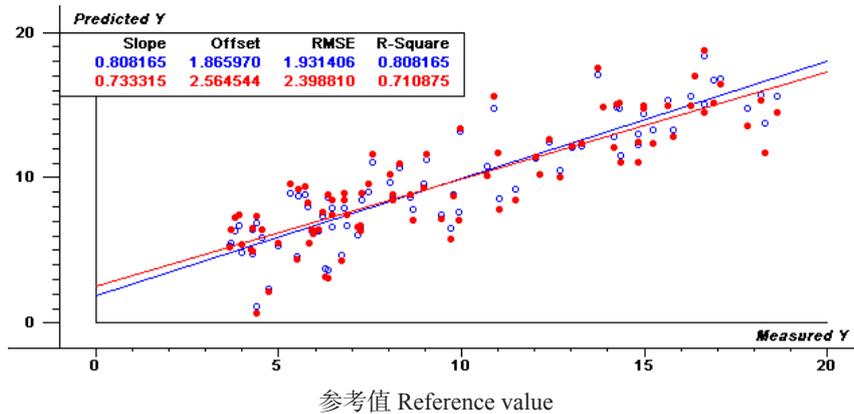


图 3 湿加松黄芪苷含量近红外预测模型

Figure 3 The near-infrared prediction model of content of astragaloside in *P. elliotii* × *P. caribaea*

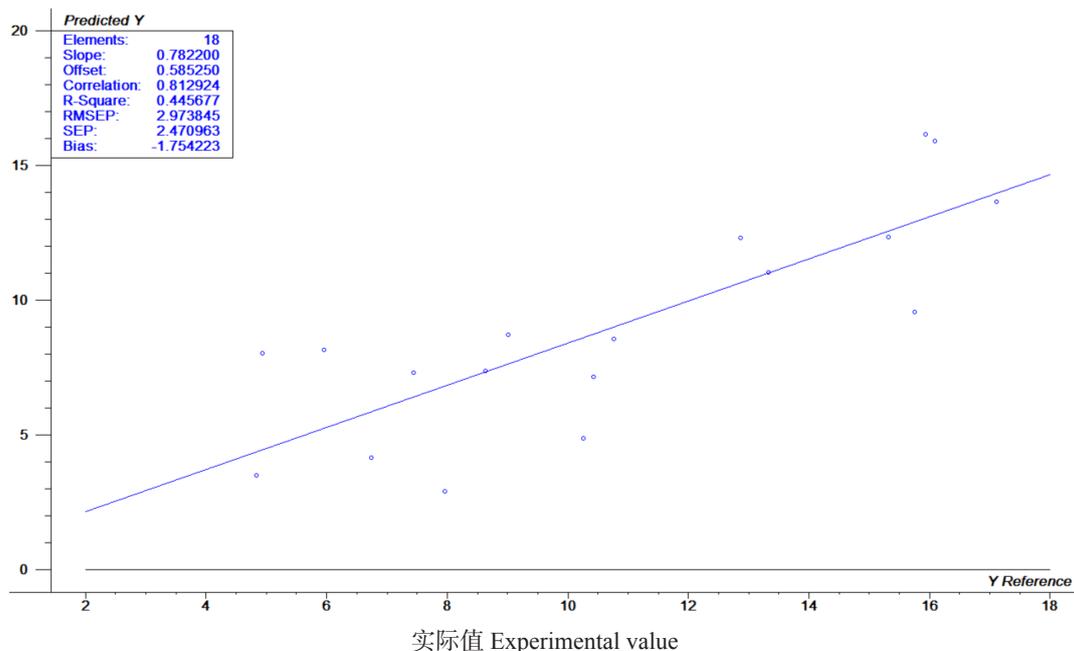


图 4 黄芪苷含量近红外预测模型的外部验证结果

Figure 4 External verification results of near infrared prediction model of astragaloside content

质的提高上, 关于湿加松针叶的生物活性成分含量研究鲜有相关报道。本研究建立的湿加松针叶黄芪苷含量近红外快速预测模型可以用来进行湿加松松针黄芪苷含量的快速预测, 在湿加松高活性成分含量的良种选育和利用近红外光谱技术预测植物的活性成分含量工作以及应用推广中起到奠定基础的作用。但是在本研究中的试验样品仅包含于台山市红岭种子园现有湿加松种质资源, 为了进一步扩大模型的适用性、提高模型的准确率, 还需要在模型中加入其他分布地的湿加松样品。

参考文献

- [1] 谢济运. 松针有效成分的提取与分析研究进展[J]. 安徽农业科学, 2011, 39(22): 13498-13500.
- [2] 김미정, 안진홍, 최강호, et al. Effects of Pine Needle Extract Oil on Blood Glucose and Serum Insulin Levels in db/db Mice[J]. Journal of the Korean Society of Food Science and Nutrition, 2006, 35(3).
- [3] 郑晓珂, 王小兰, 冯卫生. 松针提取物降血脂作用研究[J]. 中药药理与临床, 2008(3): 81-82.
- [4] 张桂芳, 张东杰, 郭希娟, 等. 三种松针活性成分含量及清除自由基能力的比较[J]. 天然产物研究与开发, 2016, 28(7): 1066-1070.
- [5] 杨书珍, 蒋丹丹, 范刚, 等. 松针提取物对柑橘青霉病菌的抑制作用及活性成分分析[J]. 现代食品科技, 2016, 32(1): 65-69.
- [6] 李丽芬, 石扣兰, 刘斌钰, 等. 复方松叶提取液的镇痛抗炎作用的实验研究[J]. 齐齐哈尔医学院学报, 2001(5): 489-490.
- [7] 华云玮, 朱凌宇, 林俊儒. 4种不同中药单体对诺如病毒感染性肠炎小鼠细胞TNF- α 、IL-1 β 、IL-6水平及PKR/pPKR的影响[J]. 辽宁中医药大学学报, 2020, 23(6): 24-27.
- [8] 李利红, 张春辉, 李荣誉, 等. 三种拮抗大肠埃希菌耐药质粒的中药研究[Z]. 郑州: 河南牧业经济学院, 2016.
- [9] BENGUO L, YONGYI Z. Extraction of flavonoids from flavonoid-rich parts in tartary buckwheat and identification of the main flavonoids[J]. J Food Eng, 2005, 78(2): 584-587.
- [10] 邹玉红, 寇小燕, 韩秋霞. 超声波辅助法提取甘草总黄酮工艺的研究[J]. 安徽农业科学, 2011, 39(14): 8345-8347.
- [11] 翁艳英, 苏秀芳. 红蓝草总黄酮的减压内部沸腾提取工艺优化[J]. 食品科技, 2016, 41(4): 200-204.
- [12] 伍观娣, 赖敏婷, 汪迎利, 等. 近红外光谱分析技术应用于植物叶片研究综述[J]. 林业与环境科学, 2020, 36(1): 118-127.
- [13] 向娜娜, 赵江涛, 陈丽, 等. 大豆油脂肪酸近红外模型的建立和品质鉴定[J]. 中国饲料, 2021(5): 72-77.
- [14] 王秀华, 王璐, 杨婉君, 等. 基于近红外光谱法的紫花苜蓿品质综合评价[J]. 中国草地学报, 2021, 43(4): 38-44.
- [15] 刘雯雯, 刘晓凤, 董航. 近红外光谱检测技术快速测定桔梗总皂苷的研究[J]. 成都大学学报(自然科学版), 2021, 40(1): 20-24.
- [16] ZENG M N, ZHENG S Y, GONZALES G F, et al. Near infrared spectroscopy combined with chemometrics to detect and quantify adulteration of maca powder[J]. J Near Infrared Spec, 2021, 29(2): 108-115.
- [17] 朱斌, 杨阳, 马淑艳, 等. 近红外漫反射光谱法定量测定金丝桃药材中的金丝桃苷与贯叶金丝桃素[J]. 药物分析杂志, 2013, 33(5): 889-893.
- [18] 黄倩倩, 潘瑞乐, 魏建和, 等. 近红外漫反射光谱法测定黄芩中总黄酮及黄芩苷的含量[J]. 光谱学与光谱分析, 2009, 29(9): 2425-2428.
- [19] 易敬林, 覃雯霞, 陈元松, 等. 17.8a生澳大利亚湿加松在桂中的生长表现[J]. 林业与环境科学, 2018, 34(6): 108-114.
- [20] 沈熙环, 黄永权. 广东省湿加松良种选育和推广进展[J]. 林业科技通讯, 2018(1): 74-75.
- [21] 蒋开彬, 牛品, 王博, 等. 火炬松木材基本密度和纤维长度近红外模型的建立与应用[J]. 福建农林大学学报(自然科学版), 2018, 47(6): 764-768.
- [22] 吴宪, 杨晓慧, 杨会肖, 等. 尾叶桉木材密度光谱特征模型研究[J]. 林业与环境科学, 2018, 34(4): 15-20.