

DOI:10.11686/cyxb2025054

http://cyxb.magtech.com.cn

吕博文, 温馨, 李军乔, 等. 基于近红外光谱分析技术的蕨麻产地判别及定量检测模型评价. 草业学报, 2026, 35(1): 140—153.

LYU Bo-wen, WEN Xin, LI Jun-qiao, et al. Development of a model based on near-infrared spectral data to evaluate the origin and quality of *Potentilla anserina* materials. Acta Prataculturae Sinica, 2026, 35(1): 140—153.

# 基于近红外光谱分析技术的蕨麻产地判别及定量检测模型评价

吕博文<sup>1,2,3</sup>, 温馨<sup>4</sup>, 李军乔<sup>1,2,3\*</sup>, 王聪<sup>1</sup>

(1. 青海民族大学生态环境与资源学院, 青海 西宁 810007; 2. 青海省特色经济植物高值化利用重点实验室, 青海 西宁 810007; 3. 青海民族大学青藏高原蕨麻产业研究院, 青海 西宁 810007; 4. 青海卫生职业技术学院, 青海 西宁 810000)

**摘要:** 本试验旨在利用近红外光谱技术, 建立蕨麻产地判别模型和营养成分快速检测的近红外预测模型。以32个采样点蕨麻块根为研究对象, 参考国家标准和行业标准分别测定蕨麻中淀粉、蛋白质、多糖、鞣质和总皂苷5个关键质量属性的含量, 并采集其全反射红外光谱(ATR)和近红外光谱(NIR)数据。利用ATR和NIR光谱数据结合建模方法、校正方法、图谱类型为因素进行3因素3水平的正交试验, 选取蕨麻样品分为校正集430个, 预测集215个构建并验证产地判别模型的优劣。经对比分析, ATR模型中, 以建模方法为扩散模型(DM), 校正方法为标准正态变量校正(SNV), 谱图类型为原谱图的组合为最优建模条件, 优化后, 其识别率为99.07%, 预测率为97.21%, 判别效果较好。在此基础上, 通过优化预处理方法、建模波段等建模条件, 建立了蕨麻5种有效营养成分的定量检测模型。其中, 鞣质最优模型为PCR+MSC+D1+Norris平滑(5,5), 建模波段为6148~5379 cm<sup>-1</sup>, 其预测相关系数(R<sub>p</sub>)为0.9393, 外部验证相对分析误差(RPD)为2.86, >2.00; 多糖的预测模型效果次之, 其最优模型为PCR+MSC+spectrum+Norris平滑(5,5), 建模波段为7000~4173 cm<sup>-1</sup>, 其R<sub>p</sub>为0.8470, RPD为1.68, >1.40。近红外光谱技术结合化学计量学可实现蕨麻产地判别及多种营养成分综合质量的快速准确检测, 为蕨麻快速综合质量评价模型的建立奠定了基础。

**关键词:** 蕨麻; 红外光谱; 产地判别模型; 定量检测模型

## Development of a model based on near-infrared spectral data to evaluate the origin and quality of *Potentilla anserina* materials

LYU Bo-wen<sup>1,2,3</sup>, WEN Xin<sup>4</sup>, LI Jun-qiao<sup>1,2,3\*</sup>, WANG Cong<sup>1</sup>

1. College of Ecological Environment and Resources, Qinghai Nationalities University, Xining 810007, China; 2. Key Laboratory of High Value Utilization of Characteristic Economic Plants of Qinghai Province, Xining 810007, China; 3. Tibetan Plateau *Potentilla anserina* L. Industry Research Institute, Qinghai Nationalities University, Xining 810007, China; 4. Qinghai College of Health Professions and Technology, Xining 810000, China

**Abstract:** In this study, we established a model based on near-infrared data to predict the origin of *Potentilla anserina* tubers and rapidly detect their nutrient status. Tubers of *P. anserina* from 32 sampling sites were analyzed to quantify five key quality attributes, namely amyllum farina, protein, polysaccharide, ellagitannin, and total saponin contents. These analyses were conducted according to the national standards and industry standards, and attenuated

收稿日期: 2025-02-25; 改回日期: 2025-04-28

基金项目: 国家自然科学基金项目(U23A20152)资助。

作者简介: 吕博文(1999—), 男, 河北邢台人, 硕士。E-mail: 17731812526@163.com

\* 通信作者 Corresponding author. E-mail: ljqlily2002@126.com

total reflection (ATR) and near infrared spectroscopy (NIR) infrared spectroscopic data were also collected for each material. The ATR and NIR spectral data combined with a modeling method, optical range type, and map type was used to conduct a three-factor, three-level orthogonal test. Selected *P. anserina* samples were divided into 430 correction sets and 215 prediction sets to construct and validate the origin discrimination model, and then the strengths and weaknesses of the model were evaluated. Comparative analyses were conducted to establish the optimal conditions for the ATR model, which were as follows: modeling method, diffusion model, light range type, standard normal variate, spectrogram type, original spectrogram. After optimization, the recognition rate of the ATR model was 99.07% and its prediction rate was 97.21%, indicating that it had a better discriminatory effect. On this basis, we established models for the quantitative detection of five compounds in *P. anserina* tubers. These models were established by optimizing the pre-processing method, modeling band, and other modeling conditions. The optimal model, which was for tannins, had the following conditions: PCR+MSC+D1+Norris smoothing (5, 5); modeling band, 6148–5379  $\text{cm}^{-1}$ . The predictive correlation coefficient ( $R_p$ ) of the tannins model was 0.9393, and the ratio of standard deviation of the validation set to standard error of prediction (RPD) was 2.86 ( $>2.00$ ). The prediction model for polysaccharides was the second most effective. The optimal conditions for the polysaccharide model were as follows: PCR+MSC+spectrum+Norris smoothing (5, 5); modeling band, 7000–4173  $\text{cm}^{-1}$ . The  $R_p$  was 0.8470 and the RPD was 1.68 ( $>1.40$ ). Our study documents the development and optimization of models incorporating NIR spectroscopy data and chemometric data for the detection of the origin and quality of *P. anserina* materials. These results laid a foundation for the establishment of a rapid and reliable method for evaluating *P. anserina* based on NIR spectral data.

**Key words:** *Potentilla anserina*; infrared spectroscopy; origin discrimination model; quantitative detection model

蕨麻(*Potentilla anserina*)属于蔷薇科(Rosaceae)委陵菜属(*Potentilla*),多年生草本,是典型的匍匐茎克隆植物,具有生长速度快、耐贫瘠、耐干旱、耐盐碱、耐践踏等生物学特性,是青藏高原特有的珍稀植物资源,俗称“人参果”“延寿果”,富含淀粉、蛋白质、膳食纤维、皂苷、总黄酮等多种营养及药用成分,药食同源,具有抗缺氧、抗疲劳、抗肿瘤、抗氧化、保肝护肝、提高免疫力的功效<sup>[1-2]</sup>,还有生津止渴、健脾益胃、收敛止血、止咳利痰、益气补血等作用,是青藏高原及周边地区人民群众喜爱的保健食品;《本草纲目拾遗》<sup>[3]</sup>中记载“蕨麻产于青藏高原,其味微涩而甘,特益于老人,而婴儿先天不足者,尤为上药,惜乎南方罕得也”。胡庭俊等<sup>[4]</sup>研究发现蕨麻多糖能提高小鼠脾指数,提高免疫抑制小鼠脾和胸腺中谷胱甘肽水平,降低丙二醇水平,表明蕨麻多糖能不同程度地拮抗环玲酰胺引起的免疫抑制,并影响机体免疫系统的功能。张新全等<sup>[5]</sup>报道了蕨麻素具有提高肝糖原合成和提高血清蛋白含量的作用,可提高血清和肝过氧化物酶活力。由于蕨麻在营养和药用方面的重要价值,长期以来深受青藏高原人民的喜爱。但是目前缺乏快速有效的鉴别手段,分类分级上缺少严格标准,导致市场上的蕨麻良莠不齐,对蕨麻产业化可持续发展极为不利,为加强蕨麻产品的规范化、标准化、分级化,打造具有青藏高原特色的品牌,需要建立低成本、简单、快速的产地判别模型。

近红外光谱技术(near infrared spectroscopy, NIRS)是利用物质分子对特定波长的红外光吸收而引起分子中振动能级和转动能级的跃迁,通过测量物质发射、吸收或散射光的波长和强度,反映样品中有机物的组分和含量的信息,进行物质定性和定量分析的技术<sup>[6]</sup>。其具有高效、简单、无损和重现性好等特点<sup>[7-8]</sup>,已广泛应用于天麻(*Gastrodia elata*)<sup>[9]</sup>、山里红(*Crataegus pinnatifida*)<sup>[10]</sup>、茯苓(*Poria cocos*)<sup>[11]</sup>、枸杞(*Lycium barbarum*)<sup>[12]</sup>、土茯苓(*Rhizoma smilacis*)<sup>[13]</sup>等产地识别及质量评价。目前已有蕨麻品质、营养及化学成分评价的研究<sup>[4-5]</sup>,但未见蕨麻产地判别方面的研究。本研究以青海 32 个采样点的 645 份蕨麻样品为对象,利用近红外光谱技术对其进行红外光谱特征检测及分析,建立蕨麻的 NIRS 图谱及研究蕨麻产地判别模型,并结合化学计量学方法,建立蕨麻多种营养成分含量的快速检测模型,以期建立快速、准确、可靠的蕨麻产地判别方法,为蕨麻的质量评价提供科学依据。

## 1 材料与方 法

### 1.1 仪器与试剂

傅里叶变换红外光谱分析仪(Nicolet IS 50, ThermoFisher, 美国), 含衰减全反射(attenuated total reflection, ATR)模块和积分球漫反射模块; 烘箱(上海一恒科学仪器有限公司, 中国), 压片机(GS03940, Specac, 英国), 电子天平(ME104, 0.0001 g, Mettler Toledo, 瑞士), 粉碎机(FW135, 天津市泰斯特有限公司), 样品筛(0.149 mm, 浙江上虞市道墟张兴纱筛厂)等。

D-无水葡萄糖[C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> 纯度: ≥98% (HPLC)], 甲基红、盐酸、酒石酸钠、盐酸标准溶液(0.1 mol·L<sup>-1</sup>)、硫酸铜(CuSO<sub>4</sub>·5H<sub>2</sub>O)、硫酸钾(K<sub>2</sub>SO<sub>4</sub>)、硼酸(H<sub>3</sub>BO<sub>3</sub>)、Sevag试剂(三氯甲烷:正丁醇=4:1)、苯酚、乙醇、硫酸、钨酸钠-钼酸钠混合溶液、碳酸钠溶液(75 g·L<sup>-1</sup>)、没食子酸标准储备液(1000 mg·L<sup>-1</sup>)、人参皂苷Re对照品、冰醋酸、高氯酸, 香草醛等。

### 1.2 试验材料

本试验所用样品均为蕨麻块根, 共32个样点, 具体分布见表1。样品于2017年10至11月进行采集, 完成采集后用纯水清洗干净, 自然晾干, 共计645份样品, 样品经粉碎后, 过0.149 mm筛, 存于干燥器中备用。进行红外光谱分析前, 放入烘箱60℃烘干至恒重, 待分析用。

### 1.3 试验方法

**1.3.1 ATR FTIR数据的采集** 使用傅里叶变换红外光谱分析仪的衰减全反射模块对蕨麻样品进行ATR-IR一维红外光谱采集, 扫描次数为32次, 多次扫描通过信号平均降低随机噪声。32次扫描是经验值, 可使信噪比(signal to noise ratio, SNR)提升约5.7倍, 同时避免仪器漂移。分辨率为4 cm<sup>-1</sup>, 光谱采集范围为4000~400 cm<sup>-1</sup>, 每个样品重复测试3次, 取平均值用于分析, 以便减少试验误差, 得到其平均图谱( $n=3$ )。

**1.3.2 FT-NIR数据的采集** 使用傅里叶变换红外光谱分析仪的积分球漫反射模块对蕨麻样品进行近红外光谱采集, 扫描次数为32次, 分辨率为6 cm<sup>-1</sup>, 光谱采集范围为10000~4000 cm<sup>-1</sup>, 每个样品重复测试3次, 取平均值用于分析, 以便减少试验误差, 得到其平均图谱( $n=3$ )。32次扫描可覆盖10000~4000 cm<sup>-1</sup>范围, 兼顾数据质量和效率。

**1.3.3 产地判别模型的建立** 1) ATR定性模型的建立。模型的建立需要具有光谱数据和其他定性数据的校正集、预测集, 故将样品的红外均谱( $n=3$ )导入TQ Analyst软件, 运用随机法对样本集进行随机划

表1 蕨麻样品信息

Table 1 Information of *P. anserina*

采样地区 Sampling area	编号 Number	样品数量 Number of samples
海北藏族自治州门源回族自治县 Menyuan Huizu Autonomous County, Haibei Tibetan Autonomous Prefecture	门源 01 Menyuan 01	20
	门源 02A Menyuan 02A	21
	门源 02B Menyuan 02B	20
	门源 03 Menyuan 03	19
	门源 04 Menyuan 04	23
	门源 05 Menyuan 05	16
门源野生 Menyuan wild	01 Menyuan wild 01	20
	02 Menyuan wild 02	17
	03 Menyuan wild 03	19
	04 Menyuan wild 04	24
	05 Menyuan wild 05	21
	06 Menyuan wild 06	22
西宁市湟源县 Huangyuan County, Xining City	湟源 01 Huangyuan 01	21
	湟源 03 Huangyuan 03	18
	湟源 04 Huangyuan 04	18
	湟源 05 Huangyuan 05	21
	湟源 06 Huangyuan 06	22
	湟源 07 Huangyuan 07	20
	海西蒙古族藏族自治州 Haixi Mongol and Tibetan Autonomous Prefecture	海西 01 Haixi 01
海西 02 Haixi 02		23
海西 02B Haixi 02B		20
海西 03A Haixi 03A		19
海西 03B Haixi 03B		22
玉树藏族自治州囊谦县 Nangqian County, Yushu Tibetan Autonomous Prefecture	囊谦 01 Nangqian 01	20
	囊谦野生 01 Nangqian wild 01	20
黄南藏族自治州泽库县 Zeku County, Huangnan Tibetan Autonomous Prefecture	泽库 01 Zeku 01	23
	泽库 02 Zeku 02	18
	泽库 03 Zeku 03	20
	泽库 04 Zeku 04	17
海南藏族自治州贵南县 Guinan County, Hainan Tibetan Autonomous Prefecture	贵南 01 Guinan 01	20
海南藏族自治州 Hainan Tibetan Autonomous Prefecture	海南 01 Hainan 01	22
果洛藏族自治州 Golog Tibetan Autonomous Prefecture	果洛 01 Golog 01	20

分, 共计 645 份样品, 2/3 为校正集, 1/3 为预测集, 即校正集为 430 个, 预测集为 215 个。模型优化利用正交试验进行, 考虑其可能影响模型准确性的主要因素(表 2), 设计 3 因素 3 水平的正交试验(表 3), 其中因素一为 2 水平, 共 18 组试验。以蕨麻样品分布的各行政区为单位进行分类, 根据识别率和预测率计算公式选出实际最优组合。判别分析(discriminant analysis, DA)用于建立分类模型, 通过提取数据特征区分不同样本类别; 扩散模型(diffusion models, DM)指代数据模型(描述数据结构与关系)或基于深度学习的扩散模型(用于数据生成与优化)。DA 常用于光谱分类与标志物筛选, 而 DM 则侧重于数据优化(如生成合成光谱或降噪处理), 故本试验建模方法选择两者进行对比分析。恒定光程(Constant): 指理想状态下光传播路径固定不变的情况, 例如样品表面完全均匀时, 光程可视为恒定。多元散射校正(multiplicative signal correction, MSC): 通过建立样品光谱与平均光谱的线性关系, 消除颗粒散射、光程变化和基线漂移, 适用于固体颗粒不均的样品(如粉末或谷物)。标准正态变量校正(standard normal variate, SNV): 通过单条光谱的标准化消除光程差异和散射干扰。SNV 不依赖参考光谱, 独立处理每条数据, 适用于复杂基质或填充条件多变的场景, 可有效减少粒径差异对光谱的影响, 需注意引入人为误差。Constant 是理想化假设, 而 MSC 和 SNV 是实际应用中校正光程差异的核心方法。MSC 基于参考光谱的线性关系, 适合系统性散射校正; SNV 通过独立标准化处理, 灵活应对复杂样品。其中建模方法中选取 DM 与 DA 两种方法; 光谱数据预处理方法主要尝试采用 Constant、MSC 和

SNV; 图谱处理主要采用原谱图、一阶导数谱图(D1)、二阶导数谱图(D2)3 种。原谱图是未经数学处理的原始数据, 直接反映样品对光的吸收、发射或散射特性。D1 是对原谱图进行一阶修正后的谱图, 通常用于消除噪声、基线漂移或增强特征峰分辨率。D2 是在 D1 基础上进行高阶修正或多因素交互分析的谱图, 用于提取精细结构或解决复杂干扰。原谱图不经过平滑处理, D1、D2 均经 Norris 平滑处理, 有效位数为 5, 有效位间隔为 5。

通过样品的近红外光谱信息初步知道样品的成分信息和结构组成, 样品的近红外光谱信息与其结构性质参数相关, 经过算法处理后, 建立数学矢量模型, 经优化处理, 即可获得产地判别模型。好的模型要求预测结果与标准方法测量结果一致, 能预测未知样本的归属, 才具有推广意义。目前常用的模型精度评价指标主要有偏差( $d$ )与极差( $e$ )、相关系数(R)、交叉验证均方根误差(root mean square error of cross validation, RMSECV)、预测均方差(root mean square error of prediction, RMSEP)等。

$$\text{识别率} = (\text{校正集总数} - \text{校正集误判数}) / \text{校正集总数} \times 100\%$$

$$\text{预测率} = (\text{预测集总数} - \text{预测集误判数}) / \text{预测集总数} \times 100\%$$

2) NIR 定性模型的建立。将 645 个样品的近红外均谱( $n=3$ )导入 TQ Analyst 软件, 运用随机法对样本集进

表 2 正交试验因素水平

Table 2 Factor-level of orthogonal test

水平 Level	建模方法 Modeling method (A)	校正方法 Calibra- tion method (B)	谱图类型 Spectral type (C)
L <sub>1</sub>	DM	Constant	原谱图
L <sub>2</sub>	DA	MSC	D1
L <sub>3</sub>	—	SNV	D2

DM: 扩散模型 Diffusion models; DA: 判别分析 Discriminant analysis; Constant: 恒定光程; MSC: 多元散射校正 Multiplicative signal correction; SNV: 标准正态变量校正 Standard normal variate; 原谱图: Original spectrogram; D1: 一阶导数谱图 Spectrogram of first-order derivatives; D2: 二阶导数谱图 Spectrogram of second-order derivatives. —: 缺失 Deletion. 下同 The same below.

表 3 不同行政区域判别模型正交试验

Table 3 Orthogonal tests of discriminant models for different administrative regions

试验 序号 Experi- ment num- ber	建模 方法 Model- ing method (A)	校正 方法 Cali- bration meth- od (B)	谱图 类型 Spec- tral type (C)	试验 序号 Experi- ment num- ber	建模 方法 Model- ing method (A)	校正 方法 Cali- bration meth- od (B)	谱图 类型 Spec- tral type (C)
1	L <sub>1</sub>	L <sub>1</sub>	L <sub>1</sub>	10	L <sub>2</sub>	L <sub>1</sub>	L <sub>1</sub>
2	L <sub>1</sub>	L <sub>1</sub>	L <sub>2</sub>	11	L <sub>2</sub>	L <sub>1</sub>	L <sub>2</sub>
3	L <sub>1</sub>	L <sub>1</sub>	L <sub>3</sub>	12	L <sub>2</sub>	L <sub>1</sub>	L <sub>3</sub>
4	L <sub>1</sub>	L <sub>2</sub>	L <sub>1</sub>	13	L <sub>2</sub>	L <sub>2</sub>	L <sub>1</sub>
5	L <sub>1</sub>	L <sub>2</sub>	L <sub>2</sub>	14	L <sub>2</sub>	L <sub>2</sub>	L <sub>2</sub>
6	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	15	L <sub>2</sub>	L <sub>2</sub>	L <sub>3</sub>
7	L <sub>1</sub>	L <sub>3</sub>	L <sub>1</sub>	16	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub>
8	L <sub>1</sub>	L <sub>3</sub>	L <sub>2</sub>	17	L <sub>2</sub>	L <sub>3</sub>	L <sub>2</sub>
9	L <sub>1</sub>	L <sub>3</sub>	L <sub>3</sub>	18	L <sub>2</sub>	L <sub>3</sub>	L <sub>3</sub>

行划分,2/3为校正集,1/3为预测集,即校正集为430个,预测集为215个。模型优化、各因素及试验设计同上。

**1.3.4 蕨麻营养成分的测定** 淀粉含量采用酶水解法<sup>[14]</sup>测定;蛋白质含量采用凯氏定氮法<sup>[15]</sup>测定;多糖含量采用紫外分光光度法<sup>[16]</sup>测定;鞣质含量的测定参考NY/T 1600-2008<sup>[17]</sup>;总皂苷含量采用分光光度法<sup>[18]</sup>测定。试验重复测量3次取平均值进行分析。

#### 1.4 数据处理

**1.4.1 多元统计分析** 采用OMNIC软件采集不同产地蕨麻的红外图谱,对其共有峰高进行归一化处理,得到其共有峰的相对峰高,并进行分析,定性模型采用TQ Analyst软件进行。试验数据以Excel进行预处理,使用SPSS 22.0进行单因素方差分析。

**1.4.2 定量模型的建立** 光谱预处理:通过TQ Analyst 9.0软件,采用主成分回归法(principal component regression, PCR)为定量校正算法。采用Constant、MSC、SNV、D1、D2及Norris平滑(其中segment lengths选择5, gap between segment选择5)等方法对光谱进行预处理,得到关联系数较高的数据模型。以校正均方差(relative root mean square error, RMSEC)、校准相关系数(calibration correlation coefficient,  $R_c$ )、预测相关系数(predictive correlation coefficient,  $R_p$ )、预测均方差(RMSEP)及验证集标准偏差与预测集标准偏差的比值(ratio of standard deviation of the validation set to standard error of prediction, RPD)对定量模型进行综合评价,考察不同预处理方法对模型建立的影响<sup>[19-20]</sup>。分别建立蕨麻中淀粉、多糖、蛋白质、总皂苷、鞣质的营养物质含量快速检测模型。

## 2 结果与分析

### 2.1 光谱数据采集

**2.1.1 ATR光谱数据** 利用OMNIC软件,在扫描次数为32次,分辨率为 $4\text{ cm}^{-1}$ 的条件下,分析红外光谱的 $4000\sim 500\text{ cm}^{-1}$ 波段,蕨麻各样点ATR平均光谱如图1所示。

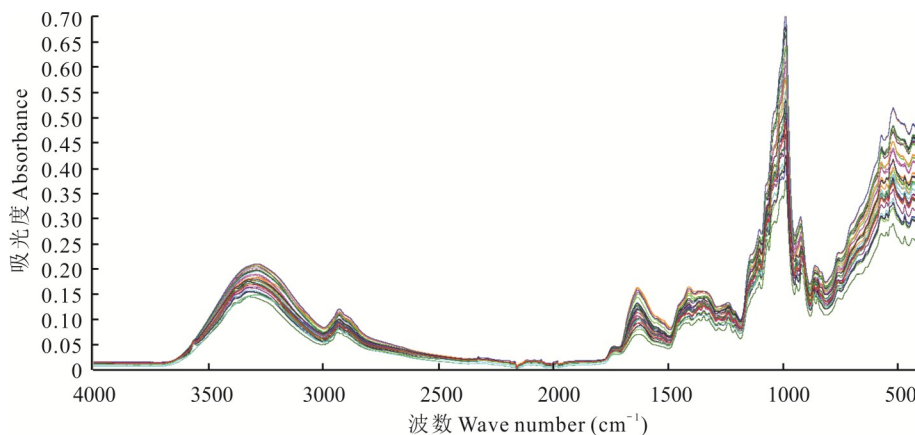


图1 蕨麻各样点ATR光谱

Fig. 1 ATR spectra of ferns at each sample site

对32个不同蕨麻样点的645份样品进行ATR-IR一维红外光谱扫描,以灵敏度80、阈值0.01为标准标峰,得到不同样点近红外光谱吸收数据。经数据处理及分析可知,32个不同蕨麻样点的共有峰有2931、1629、1409、1236、1103、989、921、866、572和 $522\text{ cm}^{-1}$ 共10个系列(表4)。各样点红外共有峰及吸光度平均值分析结果表明,由于每个样点的样品分别被随机分成了若干等份,在光谱数据采集中,基础数据较多,所以对各样点红外共有峰及吸光度的值进行均值化处理。

**2.1.2 NIR光谱数据** 利用OMNIC软件,在扫描次数32次、分辨率为 $6\text{ cm}^{-1}$ 条件下,分析红外光谱的 $10000\sim 4000\text{ cm}^{-1}$ 波段,蕨麻各样点NIR平均光谱如图2所示。

表 4 蕨麻各样点 ATR 吸收数据

Table 4 Data of *P. anserina* ATR uptake at each sample site

样点 Site	共有峰 Common peaks ( $\text{cm}^{-1}$ )									
	2931	1629	1409	1236	1103	989	921	866	572	522
贵南 01 Guinan 01	0.094	0.104	0.120	0.119	0.225	0.521	0.242	0.178	0.364	0.400
果洛 01 Golog 01	0.093	0.105	0.121	0.111	0.210	0.508	0.238	0.173	0.349	0.381
海南 01 Hainan 01	0.087	0.092	0.111	0.112	0.208	0.487	0.223	0.165	0.340	0.368
海西 01 Haixi 01	0.108	0.149	0.151	0.120	0.245	0.578	0.257	0.180	0.407	0.451
海西 02 Haixi 02	0.080	0.083	0.104	0.105	0.200	0.439	0.217	0.165	0.310	0.339
海西 02B Haixi 02B	0.114	0.152	0.156	0.132	0.263	0.634	0.272	0.193	0.437	0.483
海西 03A Haixi 03A	0.087	0.087	0.109	0.113	0.216	0.504	0.233	0.177	0.346	0.378
海西 03B Haixi 03B	0.075	0.069	0.090	0.090	0.171	0.374	0.191	0.146	0.269	0.287
湟源 01 Huangyuan 01	0.097	0.110	0.124	0.121	0.232	0.549	0.243	0.175	0.380	0.413
湟源 03 Huangyuan 03	0.089	0.100	0.112	0.114	0.192	0.406	0.216	0.161	0.293	0.329
湟源 04 Huangyuan 04	0.082	0.097	0.112	0.114	0.196	0.410	0.191	0.146	0.301	0.331
湟源 05 Huangyuan 05	0.080	0.080	0.098	0.098	0.185	0.423	0.204	0.150	0.300	0.321
湟源 06 Huangyuan 06	0.093	0.101	0.115	0.113	0.218	0.514	0.235	0.170	0.355	0.385
湟源 07 Huangyuan 07	0.098	0.101	0.125	0.115	0.237	0.594	0.282	0.187	0.389	0.432
门源 01 Menyuan 01	0.098	0.101	0.125	0.115	0.237	0.610	0.260	0.176	0.400	0.433
门源 02A Menyuan 02A	0.097	0.126	0.132	0.118	0.225	0.533	0.245	0.181	0.374	0.415
门源 02B Menyuan 02B	0.111	0.160	0.155	0.127	0.249	0.573	0.247	0.176	0.410	0.451
门源 03 Menyuan 03	0.093	0.123	0.124	0.115	0.217	0.468	0.223	0.165	0.341	0.374
门源 04 Menyuan 04	0.086	0.097	0.112	0.113	0.209	0.461	0.221	0.168	0.325	0.357
门源 05 Menyuan 05	0.081	0.116	0.117	0.094	0.200	0.531	0.219	0.142	0.362	0.394
门源野生 01 Menyuan wild 01	0.105	0.131	0.140	0.117	0.251	0.686	0.289	0.188	0.434	0.484
门源野生 02 Menyuan wild 02	0.091	0.105	0.118	0.108	0.211	0.490	0.229	0.171	0.332	0.374
门源野生 03 Menyuan wild 03	0.107	0.130	0.139	0.122	0.253	0.676	0.282	0.190	0.436	0.481
门源野生 04 Menyuan wild 04	0.110	0.123	0.140	0.128	0.261	0.671	0.288	0.197	0.440	0.482
门源野生 05 Menyuan wild 05	0.119	0.160	0.162	0.130	0.278	0.701	0.297	0.199	0.463	0.515
门源野生 06 Menyuan wild 06	0.120	0.158	0.160	0.130	0.276	0.705	0.302	0.206	0.464	0.518
囊谦 01 Nangqian 01	0.091	0.098	0.117	0.108	0.212	0.499	0.229	0.163	0.344	0.369
囊谦野生 01 Nangqian wild 01	0.109	0.143	0.145	0.125	0.259	0.640	0.281	0.189	0.430	0.474
泽库 01 Zeku 01	0.078	0.097	0.106	0.092	0.192	0.553	0.223	0.147	0.345	0.387
泽库 02 Zeku 02	0.099	0.111	0.128	0.116	0.241	0.601	0.262	0.179	0.379	0.438
泽库 03 Zeku 03	0.098	0.119	0.127	0.125	0.238	0.534	0.245	0.179	0.375	0.414
泽库 04 Zeku 04	0.090	0.101	0.119	0.116	0.217	0.480	0.233	0.175	0.339	0.373

利用 OMNIC 软件,采集  $10000\sim 4000\text{ cm}^{-1}$  的近红外光谱,以灵敏度 80、阈值 0.01 为标准标峰,分析蕨麻各样点近红外吸收峰波数(K)及对应吸光度(V),基础数据较多,对各样点红外共有峰及吸光度的值进行处理,得到不同样点蕨麻近红外 NIR 吸收数据(表 5),32 个不同蕨麻样点的 NIR 吸收波数为 8332、6790、5672、5168、4763、4309  $\text{cm}^{-1}$ ,共 6 个系列。

## 2.2 产地判别模型研究

### 2.2.1 ATR 模型

根据极差分析可知,A 因素在 1 水平效果较好,B 因素在 3 水平效果较好,C 因素在 2 水平效果较好,故理论最优组合为  $A_1B_3C_2$ ,即建模方法为 DM,校正方法为 SNV,谱图类型为 D1(表 6)。根据 R 值可知,建模方法是影响判别模型的主要因素,其次为谱图类型,校正方法对模型影响效果较小。理论最优组合 No. 8

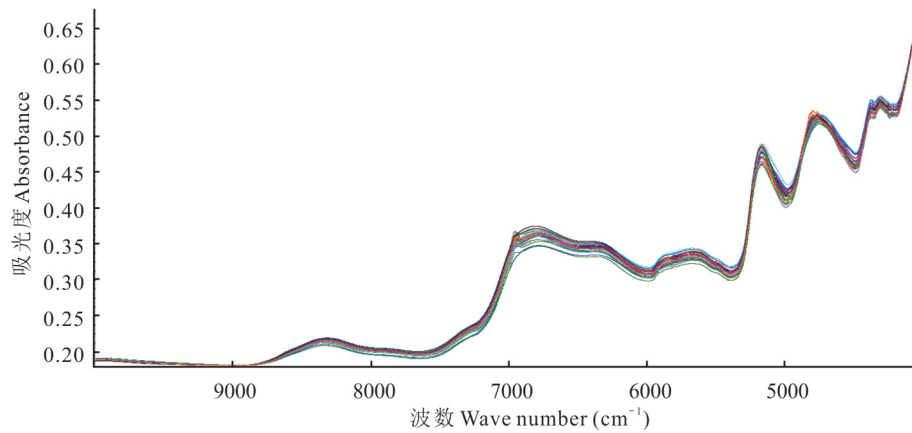


图2 蕨麻各样点 NIR 光谱

Fig. 2 NIR spectra of ferns at each site

表5 不同样点蕨麻近红外 NIR 吸收数据

Table 5 Near-infrared (NIR) absorption data of *P. anserina* with different sites ( $\text{cm}^{-1}$ )

样点 Site	波数 Wave number	样点 Site	波数 Wave number
贵南 01 Guinan 01	8337, 6794, 5674, 5168, 4763, 4310	门源 03 Menyuan 03	8344, 6780, 5674, 5168, 4763, 4310
果洛 01 Golog 01	8337, 6779, 5674, 5163, 4763, 4310	门源 04 Menyuan 04	8350, 6779, 5674, 5168, 4763, 4310
海南 01 Hainan 01	8325, 6794, 5674, 5170, 4763, 4310	门源 05 Menyuan 05	8319, 6807, 5674, 5168, 4742, 4304
海西 01 Haixi 01	8321, 6780, 5674, 5163, 4761, 4310	门源野生 01 Menyuan wild 01	8331, 6781, 5674, 5168, 4763, 4310
海西 02 Haixi 02	8350, 6779, 5674, 5168, 4796, 4310	门源野生 02 Menyuan wild 02	8343, 6780, 5674, 5168, 4763, 4310
海西 02B Haixi 02B	8337, 6781, 5674, 5163, 4761, 4310	门源野生 03 Menyuan wild 03	8314, 6809, 5668, 5168, 4748, 4306
海西 03A Haixi 03A	8344, 6786, 5674, 5174, 4794, 4310	门源野生 04 Menyuan wild 04	8320, 6810, 5668, 5168, 4756, 4306
海西 03B Haixi 03B	8350, 6775, 5674, 5168, 4794, 4312	门源野生 05 Menyuan wild 05	8320, 6781, 5668, 5163, 4748, 4306
湟源 01 Huangyuan 01	8319, 6809, 5674, 5170, 4761, 4310	门源野生 06 Menyuan wild 06	8325, 6794, 5674, 5168, 4761, 4306
湟源 03 Huangyuan 03	8344, 6780, 5674, 5170, 4763, 4310	囊谦 01 Nangqian 01	8319, 6794, 5674, 5168, 4761, 4310
湟源 04 Huangyuan 04	8314, 6809, 5668, 5168, 4742, 4306	囊谦野生 01 Nangqian wild 01	8319, 6807, 5674, 5168, 4754, 4306
湟源 05 Huangyuan 05	8325, 6786, 5674, 5170, 4761, 4310	泽库 01 Zeku 01	8319, 6809, 5668, 5168, 4748, 4306
湟源 06 Huangyuan 06	8325, 6786, 5674, 5170, 4761, 4310	泽库 02 Zeku 02	8319, 6807, 5674, 5168, 4761, 4310
湟源 07 Huangyuan 07	8314, 6809, 5639, 5168, 4761, 4310	泽库 03 Zeku 03	8350, 6781, 5674, 5168, 4763, 4310
门源 01 Menyuan 01	8350, 6779, 5674, 5168, 4763, 4310	泽库 04 Zeku 04	8350, 6780, 5674, 5168, 4794, 4310
门源 02A Menyuan 02A	8350, 6775, 5674, 5168, 4794, 4310	均值 Averages	8332, 6790, 5672, 5168, 4763, 4309
门源 02B Menyuan 02B	8325, 6781, 5674, 5163, 4738, 4306		

识别率和预测率分别为 85.35%、65.58%，实际最优组合 No. 7 识别率和预测率分别为 82.33%、80.47%，因实际最优组合的识别率和预测率相差较小，且预测率高于理论最优组合，最终确定以 No. 7 组合为建模条件，即建模方法为 DM，校正方法为 SNV，谱图类型为原谱图；同时利用 TQ Analyst 9.0 软件自带的 Regions 进行建模波段的选择，最终确定建模波段为 3928~472  $\text{cm}^{-1}$ 。按优化的建模条件及建模波段进行建模，其识别率为 99.07%，预测率为 97.21%，各样本较为分散，可以应用于产地模型判定。

**2.2.2 NIR 模型** 正交试验结果(表 7)表明，A 因素在 1 水平效果较好，B 因素在 3 水平效果较好，C 因素在 2 水平效果较好，故理论最优组合为  $A_1B_3C_2$ ，即 No. 8。根据 R 值可知，谱图类型是影响判别模型的主要因素，其次为建模方法，校正方法对模型影响效果较小。理论最优组合为 No. 8，实际最优组合为 No. 2，No. 8 与 No. 2 两种组合的识别率分别是 81.16% 和 80.00%，相差较小，No. 8 预测率为 63.26%，No. 2 预测率为 68.84%，相差较为明显。最终确定以 No. 2 组合为建模条件，即建模方法为 DM，校正方法为 Constant，谱图类型为 D1 谱图；同时利用 TQ Analyst 9.0 软件自带的 Regions 进行建模波段的选择，最终确定建模波段为 9879~4119  $\text{cm}^{-1}$ 。

表 6 ATR 区域判别模型正交试验结果

Table 6 Results of orthogonal tests of the ATR regional discriminant model

试验序号 Experiment number	建模方法 Modeling method (A)	校正方法 Calibra- tion meth- od (B)	谱图 类型 Spectral type (C)	识别率 Recogni- tion rate (%)	预测率 Predic- tion rate (%)	总和 Total (%)	试验序号 Experiment number	建模方法 Modeling method (A)	校正方法 Calibra- tion meth- od (B)	谱图 类型 Spectral type (C)	识别率 Recogni- tion rate (%)	预测率 Predic- tion rate (%)	总和 Total (%)
1	L <sub>1</sub>	L <sub>1</sub>	L <sub>1</sub>	44.19	47.44	45.27	10	L <sub>2</sub>	L <sub>1</sub>	L <sub>1</sub>	64.42	65.12	64.65
2	L <sub>1</sub>	L <sub>1</sub>	L <sub>2</sub>	83.02	57.21	74.42	11	L <sub>2</sub>	L <sub>1</sub>	L <sub>2</sub>	63.02	61.86	62.64
3	L <sub>1</sub>	L <sub>1</sub>	L <sub>3</sub>	87.44	51.16	75.35	12	L <sub>2</sub>	L <sub>1</sub>	L <sub>3</sub>	46.74	44.65	46.05
4	L <sub>1</sub>	L <sub>2</sub>	L <sub>1</sub>	82.56	79.53	81.55	13	L <sub>2</sub>	L <sub>2</sub>	L <sub>1</sub>	66.51	65.58	66.20
5	L <sub>1</sub>	L <sub>2</sub>	L <sub>2</sub>	85.12	65.58	78.60	14	L <sub>2</sub>	L <sub>2</sub>	L <sub>2</sub>	66.05	68.37	66.82
6	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	82.79	58.60	74.73	15	L <sub>2</sub>	L <sub>2</sub>	L <sub>3</sub>	45.81	43.26	44.96
7	L <sub>1</sub>	L <sub>3</sub>	L <sub>1</sub>	82.33	80.47	81.71	16	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub>	66.74	66.98	66.82
8	L <sub>1</sub>	L <sub>3</sub>	L <sub>2</sub>	85.35	65.58	78.76	17	L <sub>2</sub>	L <sub>3</sub>	L <sub>2</sub>	66.05	69.30	67.13
9	L <sub>1</sub>	L <sub>3</sub>	L <sub>3</sub>	83.02	58.60	74.88	18	L <sub>2</sub>	L <sub>3</sub>	L <sub>3</sub>	46.05	43.26	45.12

表 7 NIR 区域判别模型正交试验结果

Table 7 Results of the orthogonal experiment for the NIR region discriminant model

试验序号 Experiment number	建模方法 Modeling method (A)	校正方法 Calibra- tion meth- od (B)	谱图 类型 Spectral type (C)	识别率 Recogni- tion rate (%)	预测率 Predic- tion rate (%)	总和 Total (%)	试验序号 Experiment number	建模方法 Modeling method (A)	校正方法 Calibra- tion meth- od (B)	谱图 类型 Spectral type (C)	识别率 Recogni- tion rate (%)	预测率 Predic- tion rate (%)	总和 Total (%)
1	L <sub>1</sub>	L <sub>1</sub>	L <sub>1</sub>	43.02	36.28	40.78	10	L <sub>2</sub>	L <sub>1</sub>	L <sub>1</sub>	73.26	68.37	71.63
2	L <sub>1</sub>	L <sub>1</sub>	L <sub>2</sub>	80.00	68.84	76.28	11	L <sub>2</sub>	L <sub>1</sub>	L <sub>2</sub>	71.40	71.16	71.32
3	L <sub>1</sub>	L <sub>1</sub>	L <sub>3</sub>	73.02	45.12	63.72	12	L <sub>2</sub>	L <sub>1</sub>	L <sub>3</sub>	65.81	67.44	66.36
4	L <sub>1</sub>	L <sub>2</sub>	L <sub>1</sub>	74.42	70.70	73.18	13	L <sub>2</sub>	L <sub>2</sub>	L <sub>1</sub>	74.19	73.49	73.95
5	L <sub>1</sub>	L <sub>2</sub>	L <sub>2</sub>	81.16	63.26	75.19	14	L <sub>2</sub>	L <sub>2</sub>	L <sub>2</sub>	66.98	64.65	66.20
6	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	75.12	51.63	67.29	15	L <sub>2</sub>	L <sub>2</sub>	L <sub>3</sub>	57.21	61.40	58.60
7	L <sub>1</sub>	L <sub>3</sub>	L <sub>1</sub>	73.26	71.63	72.71	16	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub>	73.72	73.49	73.64
8	L <sub>1</sub>	L <sub>3</sub>	L <sub>2</sub>	81.16	63.26	75.19	17	L <sub>2</sub>	L <sub>3</sub>	L <sub>2</sub>	67.44	65.58	66.82
9	L <sub>1</sub>	L <sub>3</sub>	L <sub>3</sub>	75.12	51.63	67.29	18	L <sub>2</sub>	L <sub>3</sub>	L <sub>3</sub>	57.21	61.40	58.60

对比 ATR 和 NIR 模型的结果可知,建模方法为 DM,校正方法为 SNV,谱图类型为原谱图;建模波段为 3928~472 cm<sup>-1</sup>;优化建模后,其识别率为 99.07%,预测率为 97.21%,表明各个样本间的差异性较为明显,即可区分开,说明该建模方法对蕨麻的产地分布的判定极为敏感。以同样的采集处理和试验条件,以及本次 32 个蕨麻产地为基准分布点,可以基本判定新样品的产地。

### 2.3 不同产地蕨麻营养成分的测定

对 32 个样点的蕨麻样品进行 5 种营养成分含量的测定(表 8),其中,淀粉含量为 412~507 mg·g<sup>-1</sup>,平均值为 448 mg·g<sup>-1</sup>,湟源 07 的淀粉含量最高;蛋白质含量为 85.1~155.0 mg·g<sup>-1</sup>,平均值为 111.6 mg·g<sup>-1</sup>,门源 02B 的蛋白质含量最高;多糖含量为 32.2~162.9 mg·g<sup>-1</sup>,平均值为 94.0 mg·g<sup>-1</sup>,门源 02B 的多糖含量最高;鞣质含量为 8.7~36.6 mg·kg<sup>-1</sup>,平均值为 20.33 mg·kg<sup>-1</sup>,湟源 03 的鞣质含量最高;总皂苷含量为 14.2~41.1 mg·g<sup>-1</sup>,平均值为 27.1 mg·g<sup>-1</sup>,湟源 03 的总皂苷含量最高。

表8 蕨麻中5种营养成分含量测定结果

Table 8 Determination result of five nutrient contents in *P. anserina* (mg·g<sup>-1</sup>)

样品编号 Sample number	淀粉 Amylum	蛋白质 Protein	多糖 Polysaccharide	鞣质 Tannin	总皂苷 Total saponin
泽库01 Zeku 01	442±6.9	92.6±1.9	63.6±2.4	18.6±0.21	14.2±0.6
果洛01 Golog 01	447±6.7	97.8±2.1	32.2±1.5	12.5±0.32	20.1±1.1
泽库04 Zeku 04	422±6.2	92.8±1.7	111.2±1.1	20.8±0.34	27.6±1.3
囊谦野生01 Nangqian wild 01	414±1.1	125.0±1.5	122.5±8.8	12.1±0.41	20.3±0.6
囊谦01 Nangqian 01	412±5.0	85.1±1.5	38.5±3.0	19.5±0.27	25.8±0.6
贵南01 Guinan 01	453±8.3	125.0±2.8	108.5±6.4	25.4±0.73	24.0±1.2
海西03A Haixi 03A	439±6.1	90.1±1.2	62.4±3.4	27.0±1.03	25.7±0.9
海西02A Haixi 02A	433±6.7	101.0±2.5	152.1±10.9	17.9±0.24	22.6±0.9
泽库03 Zeku 03	420±5.7	90.1±2.0	132.7±3.8	24.4±0.91	25.6±0.9
门源01 Menyuan 01	422±6.5	95.5±1.6	123.4±0.6	19.4±0.25	24.4±0.9
门源03 Menyuan 03	450±4.0	135.0±3.0	152.8±11.2	24.2±0.97	27.4±0.2
门源02A Menyuan 02A	416±1.7	124.0±0.8	70.2±4.5	14.0±0.55	24.9±1.5
门源04 Menyuan 04	448±6.9	110.0±2.8	103.2±5.3	22.2±1.12	28.5±1.7
门源野生03 Menyuan wild 03	444±1.8	100.0±1.4	102.5±4.9	18.7±0.77	25.7±0.7
海南01 Hainan 01	448±6.5	125.0±1.8	32.6±1.7	30.1±1.10	24.6±1.2
湟源04 Huangyuan 04	487±7.8	91.6±1.6	56.4±3.9	23.3±1.01	19.8±1.0
海西01 Haixi 01	418±4.0	138.0±1.2	128.3±6.6	13.2±0.39	32.9±1.5
泽库02 Zeku 02	418±3.3	104.0±2.0	83.6±6.4	15.9±0.32	31.8±1.7
湟源07 Huangyuan 07	507±8.9	121.0±2.7	48.0±3.6	23.1±0.83	22.6±0.9
海西02B Haixi 02B	429±6.5	128.0±1.9	146.0±7.1	20.2±0.20	25.5±1.1
门源野生01 Menyuan wild 01	456±2.8	110.0±2.6	153.4±6.9	10.6±0.43	25.8±1.3
湟源01 Huangyuan 01	446±5.6	103.0±2.0	64.4±4.7	32.9±1.41	31.4±1.3
海西03B Haixi 03B	503±8.7	94.3±2.0	110.0±4.3	17.5±0.60	29.0±1.1
湟源06 Huangyuan 06	469±5.3	121.0±2.8	61.7±2.7	30.6±1.21	28.3±0.9
门源02B Menyuan 02B	432±5.0	155.0±1.5	162.9±8.2	18.2±0.39	27.1±1.5
门源野生04 Menyuan wild 04	469±6.2	90.9±2.2	64.6±4.5	20.2±0.31	30.2±0.7
门源野生06 Menyuan wild 06	474±4.7	123.0±2.5	104.5±6.4	11.1±0.40	31.7±1.8
门源野生05 Menyuan wild 05	432±7.2	150.0±3.3	95.4±2.6	15.2±0.68	30.1±1.2
湟源05 Huangyuan 05	480±5.8	99.4±1.4	41.4±3.0	28.6±0.60	33.6±2.0
湟源03 Huangyuan 03	420±0.4	105.0±2.4	78.9±2.3	36.6±1.13	41.1±1.7
门源野生02 Menyuan wild 02	494±7.4	103.0±1.4	72.3±2.5	18.0±0.76	32.8±1.4
门源05 Menyuan 05	482±3.2	145.0±1.8	128.8±10.2	8.7±0.29	33.0±1.4

## 2.4 定量检测模型的建立

使用 TQ Analyst 9.0 进行定量检测模型的建立,以海南01、海西02B、湟源01、湟源05、门源01、门源03、门源野生01和门源野生04为验证集,其余样品为建模集,进行淀粉单因素定量检测模型优化试验,具体模型参数如表9所示。其单因素筛选出最优模型为 PCR+MSC+D1+Norris 平滑(5,5),建模波段为 6140~5390 cm<sup>-1</sup>,所建淀粉模型的 R<sub>p</sub> 为 0.6895,RPD 为 1.38,<1.40,模型效果稍低;以海南01、海西02B、湟源01、湟源05、门源01、门源03、门源野生01、门源野生04、门源野生05和泽库02为验证集,其余样品为建模集,进行蛋白质单因素定量检测模型优化试验,筛选出其最优模型为 PCR+SNV+D2+Norris 平滑(5,5),建模波段为 5234~4129 cm<sup>-1</sup>,所建蛋

表 9 蕨麻多指标关键质量属性单因素近红外光谱模型参数

Table 9 Parameters of one-factor near-infrared spectral model for key quality attributes of *P. anserina* multi-indicators

成分 Component	校正方法 Calibration method	谱图类型 Spectral type	潜变量 Latent variable	建模波段 Modeling band (cm <sup>-1</sup> )	RMSEC	R <sub>c</sub>	RMSEP	R <sub>p</sub>	RPD
淀粉 Starch	Constant	spectrum	3	5280~4382	2.650	0.3626	1.460	0.5790	1.31
		D1(NS)	3	6140~5390	2.240	0.6155	1.480	0.6959	1.29
		D2(NS)	3	6140~5390	1.920	0.7374	1.630	0.5260	1.17
	MSC	spectrum	3	5280~4382	2.420	0.5280	1.460	0.6676	1.31
		D1(NS)	3	6140~5390	2.190	0.6394	1.380	0.6895	1.38
		D2(NS)	3	6140~5390	2.020	0.7030	1.670	0.4880	1.14
	SNV	spectrum	3	5280~4382	2.410	0.5305	1.480	0.6640	1.29
		D1(NS)	3	6140~5390	2.190	0.6387	1.380	0.6908	1.38
		D2(NS)	3	6140~5390	2.020	0.7027	1.660	0.4887	1.15
蛋白质 Protein	Constant	spectrum	6	5221~4144	1.480	0.6348	1.250	0.7771	1.55
		D1(NS)	6	5234~4129	1.310	0.7309	1.250	0.7335	1.55
		D2(NS)	6	5234~4129	1.270	0.7485	1.210	0.7561	1.60
	MSC	spectrum	6	5221~4144	1.280	0.7455	1.380	0.6866	1.41
		D1(NS)	6	5234~4129	1.270	0.7506	1.170	0.7726	1.66
		D2(NS)	6	5234~4129	1.260	0.7546	1.120	0.7965	1.73
	SNV	spectrum	6	5221~4144	1.280	0.7453	1.370	0.6898	1.42
		D1(NS)	6	5234~4129	1.270	0.7508	1.170	0.7723	1.66
		D2(NS)	6	5234~4129	1.260	0.7547	1.120	0.7966	1.73
多糖 Polysaccharides	Constant	spectrum	6	7000~4173	1.930	0.8154	2.700	0.7762	1.64
		D1(NS)	10	6940~4490	1.790	0.8439	3.220	0.6527	1.37
		D2(NS)	10	6940~4490	1.540	0.8870	2.900	0.7726	1.52
	MSC	spectrum	6	7000~4173	2.140	0.7660	2.630	0.8470	1.68
		D1(NS)	10	6940~4490	1.620	0.8730	2.940	0.7312	1.50
		D2(NS)	10	6940~4490	1.490	0.8938	2.940	0.7677	1.50
	SNV	spectrum	6	7000~4173	2.140	0.7657	2.640	0.8449	1.67
		D1(NS)	10	6940~4490	1.630	0.8719	2.940	0.7310	1.50
		D2(NS)	10	6940~4490	1.490	0.8939	2.940	0.7682	1.50
鞣质 Tannins	Constant	spectrum	4	5215~4144	0.311	0.8443	0.451	0.7251	1.44
		D1(NS)	4	6148~5379	0.222	0.9243	0.312	0.8928	2.08
		D2(NS)	4	6148~5379	0.192	0.9438	0.270	0.9080	2.41
	MSC	spectrum	4	5215~4144	0.395	0.7331	0.519	0.6414	1.25
		D1(NS)	4	6148~5379	0.240	0.9103	0.227	0.9393	2.86
		D2(NS)	4	6148~5379	0.230	0.9184	0.263	0.9240	2.47
	SNV	spectrum	4	5215~4144	0.399	0.7274	0.521	0.6341	1.25
		D1(NS)	4	6148~5379	0.241	0.9098	0.228	0.9392	2.85
		D2(NS)	4	6148~5379	0.230	0.9185	0.263	0.9240	2.47
总皂苷 Total saponin	Constant	spectrum	3	7015~4392	0.506	0.4396	0.439	0.2780	0.66
		D1(NS)	3	4381~4362	0.341	0.7953	0.319	0.6732	0.91
		D2(NS)	3	4381~4362	0.509	0.4272	0.350	0.1327	0.83
	MSC	spectrum	3	7015~4392	0.546	0.2407	0.369	0.5327	0.79
		D1(NS)	3	4381~4362	0.358	0.7714	0.349	0.6466	0.83
		D2(NS)	3	4381~4362	0.479	0.5252	0.448	0.2873	0.65

续表 Continued Table

成分	校正方法	谱图类型	潜变量	建模波段	RMSEC	R <sub>c</sub>	RMSEP	R <sub>p</sub>	RPD
Component	Calibration method	Spectral type	Latent variable	Modeling band (cm <sup>-1</sup> )					
总皂苷	SNV	spectrum	3	7015~4392	0.546	0.2421	0.370	0.5322	0.78
Total saponin		D1(NS)	3	4381~4362	0.358	0.7715	0.349	0.6467	0.83
		D2(NS)	3	4381~4362	0.480	0.5225	0.445	0.2829	0.65

RMSEC: 校正均方差 Relative root mean square error; RMSEP: 预测均方差 Root mean square error of prediction; R<sub>c</sub>: 校准相关系数 Calibration correlation coefficient; R<sub>p</sub>: 预测相关系数 Predictive correlation coefficient; RPD: 验证集标准偏差与预测集标准偏差的比值 Ratio of standard deviation of the validation set to standard error of prediction.

白质模型的 R<sub>p</sub> 为 0.7966, RPD 为 1.73, >1.40, 模型效果较为可靠; 以海南 01、海西 02B、湟源 01、湟源 05、门源 01、门源 03、门源野生 03、门源野生 04、囊谦 01 和泽库 02 为验证集, 其余样品为建模集, 进行多糖单因素定量检测模型优化试验, 筛选出其最优模型为 PCR+MSC+spectrum+Norris 平滑(5,5), 建模波段为 7000~4173 cm<sup>-1</sup>, 所建多糖模型的 R<sub>p</sub> 为 0.8470, RPD 为 1.68, >1.40, 模型效果较为可靠; 以海南 01、海西 02B、湟源 05、湟源 06、门源 01、门源 03、门源野生 01、门源野生 04、囊谦 01 和泽库 02 为验证集, 其余样品为建模集, 进行鞣质单因素定量检测模型优化试验, 筛选出其最优模型为 PCR+MSC+D1+Norris 平滑(5,5), 建模波段为 6148~5379 cm<sup>-1</sup>, 所建鞣质模型的 R<sub>p</sub> 为 0.9393, RPD 为 2.86, >2.00, 模型效果具有较高的可靠性; 以海南 01、海西 02B、湟源 01、门源 01、门源 03、门源野生 03、囊谦 01 和泽库 02 为验证集, 其余样品为建模集, 进行总皂苷单因素定量检测模型优化试验, 筛选出其最优模型为 PCR+constant+D1+Norris 平滑(5,5), 建模波段为 4381~4362 cm<sup>-1</sup>, 所建总皂苷模型的 R<sub>p</sub> 为 0.6732, RPD 为 0.91, <1.40, 模型效果较为不可靠。

本试验以 32 个不同产地的蕨麻块根为样本, 使用近红外光谱技术对其 5 种营养成分(淀粉、蛋白质、多糖、鞣质、总皂苷)含量进行了定量分析, 初步建立了其营养成分的定量预测模型。以上结果表明, 所初步建立的营养成分的近红外预测模型整体效果较好。其中, 鞣质的预测模型效果最好, 多糖的预测模型效果次之, 而淀粉与蛋白质的预测模型效果较好, 而总皂苷的预测效果较差, 其预测模型的性能有待进一步提高。

### 3 讨论

#### 3.1 近红外技术优势

NIRS 作为近年来快速发展的新型分析手段, 凭借非破坏性、高效率等优势, 在食品和药品检测中广泛应用。其通过物质分子振动倍频与合频吸收效应, 避免了传统方法中的复杂预处理, 即可直接获取液态或固态样本的光谱响应信号, 有效简化了操作流程, 在检测时效性和样本完整性要求严苛的领域具有独特优势<sup>[21]</sup>。国家标准中淀粉、蛋白质、多糖含量和行业标准中鞣质和总皂苷含量的测定方法虽更为准确, 但相较于近红外光谱分析, 其样品的预处理较为复杂, 耗时较长, 影响整体试验的效率。但其在高水分或复杂基质样品应用中存在一定程度的局限性, 如高水分样品(如土壤、中药材)因水分子 O-H 键在 1450 和 1940 nm 附近的强吸收峰会掩盖其他成分信号, 导致模型预测精度下降; 复杂基质(如深色药材、多组分混合物)则因光散射效应和背景干扰(如颜色吸光或颗粒不均)导致信噪比降低, 需通过波段筛选(1500~2300 nm)和 SNV 预处理削弱干扰。目前主要通过光谱预处理(MSC、SNV)、算法优化(非线性模型、特征波长提取)及样品前处理(干燥、粉碎)缓解问题, 但未来需结合短波近红外传感器开发、多技术联用(如拉曼光谱)和动态建模以突破现有瓶颈, 进一步提升 NIRS 在食品、医药等复杂场景的适用性。

#### 3.2 模型局限性

在光谱采集过程中, 仪器响应稳定性与环境参数波动是影响数据质量的关键因素<sup>[22]</sup>。ATR-FTIR 通过高折射率晶体与样品表面的倏逝波耦合作用获取化学信息, 相较于传统透射法(NIR)可有效避免溴化钾压片导致的样品破坏和吸湿干扰<sup>[23]</sup>。常规的透射式红外光谱法以透过样品的干涉辐射所携带的物质信息来分析该样品, 该方法要求样品具有很好的红外线通透性。ATR 是基于光内反射, 从光源发出的红外光经过折射率大的晶体再投

射到折射率小的试样表面上,当入射角大于临界角时,入射光线就会产生全反射。红外光并不是全部被反射回来,而是穿透到样品表面内一定深度后再返回。在该过程中,试样在入射光频率区域内有选择吸收,反射光强度减弱,产生与透射吸收相似的谱图,从而获得样品表层化学成份的结构信息<sup>[24]</sup>。本试验分别进行了两种光谱技术的数据采集及分析,运用主成分回归法、原始光谱、多元散射校正、标准正态变换、一阶导数、二阶导数及 Norris 平滑等方法对光谱进行预处理,初步建立 ATR 和 NIR 产地判别模型,并对其进行数据验证。

市场上蕨麻良莠不齐,混乱发展,故产地判别则可以为其提供有效的地域信息,供其择优选择。韩沅汐等<sup>[25]</sup>收集西藏自治区 4 个地区的风干牦牛肉样品 240 份,优化其近红外光谱数据,建立了 MC-BOSS-SVM 模型对风干牦牛肉进行产地判别,预测集准确率为 91.67%。崔萌等<sup>[10]</sup>收集了 6 个省份的山里红样本 91 份,采集其 NIRS,应用多种机器学习算法对其进行产地溯源分析及判别,建立了 NIRS 结合 ANN 模型的山里红产地溯源的方法。杨承恩等<sup>[26]</sup>利用 15 个产区的 740 份黑果腺肋花楸(*Aronia melanocarpa*)果实为样品,运用中红外光谱技术及化学计量学对其进行产地鉴别,确定最佳识别模型为 SPA-SVM,其训练集与测试集识别率均达到 100%,且所需波长点仅为 16 个,可有效、准确鉴别黑果腺肋花楸果实产地。本试验所用蕨麻材料的 32 个样点包含了西宁市、海北藏族自治州、海西蒙古族藏族自治州、海南藏族自治州、黄南藏族自治州、玉树藏族自治州、果洛藏族自治州等青海省大部分地区,其中西宁市、海北藏族自治州样点数量较多,分别为 12 和 6 个,主要是两地的蕨麻分布较多较广。青海省的蕨麻产地分布极为广泛,32 个样本点位无法全面覆盖,因此未来研究将扩大样本范围、数量,增加样本多样性,完善样品库,通过更为丰富的样品识别算法及模型训练,同时加强其他地区或不同时期的蕨麻样品的收集及分析,以有效提高模型的普适性、稳定性和可预测性。

近红外光谱定量分析只需建立近红外光谱参数及样品待测成分的校准曲线,即可建立特定成分的定量预测模型,优化其模型,便可对未知样品待测成分进行预测。裴莉昕等<sup>[27]</sup>收集不同产地的 15 份葛根(*Pueraria lobata*)样品,采用紫外分光光度法、HPLC 法及烘干法对其 4 种主要成分进行定量分析,结合红外光谱技术及化学计量学建立了其有效成分快速鉴别的红外预测模型。乔璐等<sup>[28]</sup>则采集 4 省份 16 区(县)的地黄(*Rehmannia glutinosa*)样品,利用红外光谱结合高效液相色谱法对其进行快速无损鉴别,并建立起红外特征图谱。本试验分别以酶水解法和凯氏定氮法测定不同产地蕨麻中淀粉和蛋白质含量,以分光光度法测定不同产地蕨麻中多糖、鞣质、总皂苷含量,结合近红外光谱技术所采集的光谱数据分别进行单因素试验,以确定其主要营养成分的最佳定量检测模型。

#### 4 结论

本试验以 32 个采样点蕨麻块根为研究对象,利用红外光谱技术,分别建立了青海省 32 个产地蕨麻块根的 ATR 和 NIR 红外光谱基础数据库并对其进行分析,结合化学计量学建立青海蕨麻产地判别模型和淀粉、多糖、蛋白质、总皂苷、鞣质等营养物质含量快速检测预测模型,探究利用 NIRS 对蕨麻质量进行快速、全面评价的可行性。经正交试验筛选对比,ATR 模式下,采取建模方法为 DM,校正方法为 SNV,谱图类型为原谱图;建模波段为 3928~472  $\text{cm}^{-1}$ ,其识别率可达 99.07%;结合样品分布点位图,各样品点位分布较为分散,边界清晰,此建模方法可以作为产地判别模型来应用。NIR 模式下,建模方法为 DM,校正方法为 Constant,谱图类型为一阶导数谱图;建模波段为 9879~4119  $\text{cm}^{-1}$ ;其识别率可达 81.6%。蕨麻不同营养成分的最佳建模方法不同,其预测模型显示,鞣质的模型评价指标最优, $R_p$  为 0.9393,RPD 值为 2.86;多糖的预测模型效果次之, $R_p$  为 0.8470,RPD 为 1.68;淀粉与蛋白质的预测模型效果下降了 50% 左右,而总皂苷的预测效果较差。整体而言,所建模型对蕨麻样品的预测能力较强。因此,利用近红外光谱技术结合湿化学试验的方式可以简便、准确、快速地对蕨麻产地进行判别及营养成分的检测,以期为蕨麻快速综合质量评价模型的建立奠定一定基础,为蕨麻质量评价标准的完善与生产实践的快速应用提供新的研究思路和分析手段。

**参考文献 References:**

- [1] Li J Q, Cai G M, Li L Z. The Chinese Juema. Beijing: Science Press, 2020.  
李军乔, 蔡光明, 李灵芝. 中国蕨麻. 北京: 科学出版社, 2020.
- [2] Committee on Flora of China, Chinese Academy of Sciences. Flora of China (Vol. 37, Rosaceae). Beijing: Science Press, 1985.  
中国科学院中国植物志委员会. 中国植物志(第37卷蔷薇科). 北京: 科学出版社, 1985.
- [3] Zhao X M. Gleanings from the compendium of materia medica. Beijing: People's Health Publishing House, 1963.  
赵学敏. 本草纲目拾遗. 北京: 人民卫生出版社, 1963.
- [4] Hu T J, Chen J R, Cheng F S, *et al.* Effect of *Potentilla anserine* polysaccharide on levels of IL-6, IFN- $\gamma$  and TNF- $\alpha$  in murine serum. Chinese Veterinary Science, 2005(8): 653–656.  
胡庭俊, 陈昊然, 程富胜, 等. 蕨麻多糖对小鼠血清中三种细胞因子水平的影响. 中国兽医科技, 2005(8): 653–656.
- [5] Zhang X Q, Zhao Y L, Shan L M, *et al.* Study on protective mechanism of JMS on chemical liver injury. Pharmaceutical Journal of Chinese People's Liberation Army, 2004(4): 259–261.  
张新全, 赵艳玲, 山丽梅, 等. 蕨麻素对化学性肝损伤保护作用机制的研究. 解放军药理学学报, 2004(4): 259–261.
- [6] Chu X L, Shi Y Y, Chen P, *et al.* Research and application progresses of near infrared spectroscopy analytical technique in China in past five years. Journal of Instrumental Analysis, 2019, 38(5): 603–611.  
褚小立, 史云颖, 陈瀑, 等. 近五年我国近红外光谱分析技术研究与应用进展. 分析测试学报, 2019, 38(5): 603–611.
- [7] Zhang Y, Wang D, Li X, *et al.* Research progress of near infrared spectroscopy based geographical origin traceability of agricultural products. Journal of Food Safety & Quality, 2018, 9(23): 6161–6166.  
张勇, 王督, 李雪, 等. 基于近红外光谱技术的农产品产地溯源研究进展. 食品安全质量检测学报, 2018, 9(23): 6161–6166.
- [8] Xia Z Z, Zheng D, Xia H, *et al.* Rapid identification of shiitake geographical origins based on near infrared spectroscopy. Journal of Instrumental Analysis, 2020, 39(11): 1371–1377.  
夏珍珍, 郑丹, 夏虹, 等. 基于近红外光谱分析技术的栽培香菇产地快速鉴别. 分析测试学报, 2020, 39(11): 1371–1377.
- [9] Bai Q X, Hou Y, Yang P P, *et al.* Identification method of the production site of *Gastrodia elata* Blume based on near infrared spectroscopy. Journal of West China Forestry Science, 2021(3): 124–130.  
白庆旭, 候英, 杨盼盼, 等. 基于近红外光谱技术的天麻产地鉴别方法. 西部林业科学, 2021(3): 124–130.
- [10] Cui M, Duan B Z, Cheng L, *et al.* Geographical origin traceability of *Crataegus pinnatifida* var. *major* based on near-infrared spectroscopy combined with chemometrics. Chinese Traditional and Herbal Drugs, 2024, 55(14): 4897–4906.  
崔萌, 段宝忠, 程蕾, 等. 基于近红外光谱结合化学计量学方法的山里红产地溯源分析. 中草药, 2024, 55(14): 4897–4906.
- [11] Li J Y, Yu M, Zheng Y, *et al.* Nondestructive identification of *Poria cocos* blocks from different origins based on near infrared spectroscopy. Chinese Journal of Analysis Laboratory, 2021(12): 1381–1386.  
李嘉仪, 余梅, 郑郁, 等. 基于近红外光谱技术的不同产地茯苓块无损鉴别. 分析实验室, 2021(12): 1381–1386.
- [12] Xu R, Ao D M, Xu X, *et al.* Study on the identification method of *Lycium barbarum* cultivars in Ningxia based on infrared spectrum and cluster analysis. Spectroscopy and Spectral Analysis, 2024, 44(5): 1386–1391.  
徐荣, 敖冬梅, 徐鑫, 等. 基于红外光谱与聚类分析法的宁夏产地枸杞子品种鉴别研究. 光谱学与光谱分析, 2024, 44(5): 1386–1391.
- [13] Lü M Y, Wan X Y, Shuai J H, *et al.* A processing-trajectory strategy for rapid quality assessment of *Rhizoma smilacis* Glabrae by NIR combined with PLS modeling. Journal of Yangzhou University (Agricultural and Life Science Edition), 2024(3): 82–89.  
吕蒙莹, 万夏芸, 帅锦浩, 等. 近红外技术结合参数轨迹策略快速评价土茯苓质量. 扬州大学学报(农业与生命科学版), 2024(3): 82–89.
- [14] National Health Commission of the People's Republic of China. National standard for food safety determination of amyllum C<sub>6</sub>H<sub>10</sub>O<sub>5</sub> in food: GB 5009.9-2023. Beijing: State Administration for Market Regulation, 2023.  
中华人民共和国国家卫生健康委员会. 食品安全国家标准 食品中淀粉的测定: GB 5009.9-2023. 北京: 国家市场监督管理总局, 2023.
- [15] National Health and Family Planning Commission of the People's Republic of China, China Food and Drug Administration. National standard for food safety determination of protein in food: GB 5009.5-2016. Beijing: China Food and Drug Administration, 2016.

- 中华人民共和国国家卫生和计划生育委员会, 国家食品药品监督管理总局. 食品安全国家标准 食品中蛋白质的测定: GB 5009.5-2016. 北京: 国家食品药品监督管理总局, 2016.
- [16] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Wolfberry: GB/T 18672-2014. Beijing: Standardization Administration of the People's Republic of China, 2014.  
中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 枸杞: GB/T 18672-2014. 北京: 中国国家标准化管理委员会, 2014.
- [17] Ministry of Agriculture of the People's Republic of China. Determination of tannin content in fruits, vegetables and derived product—Spectrophotometry method: NY/T 1600-2008. Beijing: Ministry of Agriculture of the People's Republic of China, 2008.  
中华人民共和国农业部. 水果、蔬菜及其制品中单宁含量的测定 分光光度法: NY/T 1600-2008. 北京: 中华人民共和国农业部, 2008.
- [18] Ministry of Health of the People's Republic of China. Technical standards for testing & assessment of health food. Beijing: Standards Press of China, 2003.  
中华人民共和国卫生部. 保健食品检验与评价技术规范. 北京: 中国标准出版社, 2003.
- [19] Du C Z, Dai S Y, Zhao A B, *et al.* Optimization of PLS modeling parameters via quality by design concept for *Gardenia jasminoides* Ellis using online NIR sensor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 222: 117267.
- [20] Luo L, Wang J Y, Xu Z J, *et al.* Geographic origin discrimination of wood using NIR spectroscopy combined with machine learning techniques. *Spectroscopy and Spectral Analysis*, 2023, 43(11): 3372–3379.  
骆立, 王静仪, 徐兆军, 等. 基于近红外光谱技术建立木材产地鉴别模型. *光谱学与光谱分析*, 2023, 43(11): 3372–3379.
- [21] Liu C L. Study on chemical fingerprint and geographic origins traceability of *Panax notoginseng*. Kunming: Yunnan University of Chinese Medicine, 2023.  
刘春路. 三七化学指纹图谱及产地溯源研究. 昆明: 云南中医药大学, 2023.
- [22] Han C S. Application of exosomes derived from mesenchymal stem cell modified by MiR-675 and hydrogel. Suzhou: Soochow University, 2019.  
韩超珊. MiR-675 和水凝胶对间充质干细胞外泌体的改造和应用研究. 苏州: 苏州大学, 2019.
- [23] Signe V, Anu T, Pillerin P, *et al.* ATR-FT-IR spectral collection of conservation materials in the extended region of 4000~80  $\text{cm}^{-1}$ . *Analytical and Bioanalytical Chemistry*, 2016, 408(13): 3373–3379.
- [24] Yu Z. Preparation and properties of antioxidant active packaging film based on metal ions chelation principle. Wuxi: Jiangnan University, 2021.  
于振. 基于金属离子螯合原理的抗氧化活性包装膜制备及性能研究. 无锡: 江南大学, 2021.
- [25] Han Y X, Hu Y, Zhang E H, *et al.* Place of origin identification for air-dried yak meat from Tibet based on portable near-infrared spectroscopy. *Food and Fermentation Industries*, 2025, 51(10): 336–343.  
韩沅汐, 胡云, 张二豪, 等. 基于便携式近红外光谱的西藏风干牦牛肉产地鉴别研究. *食品与发酵工业*, 2025, 51(10): 336–343.
- [26] Yang C E, Li M, Wang T C, *et al.* Identification of *Aronia melanocarpa* fruits from different areas by mid-infrared spectroscopy. *Spectroscopy and Spectral Analysis*, 2024, 44(4): 991–996.  
杨承恩, 李萌, 王天赐, 等. 红外光谱的不同产地黑果腺肋花楸果实鉴别. *光谱学与光谱分析*, 2024, 44(4): 991–996.
- [27] Pei L X, He J L, Wang K L, *et al.* Rapid quantitative study of *Pueraria lobata* in traditional Chinese medicine based on Fourier transform infrared spectroscopy combined with multivariate statistical analysis. *Chemical Research and Application*, 2024, 36(9): 1988–1996.  
裴莉昕, 何江龙, 王锴乐, 等. 基于傅里叶变换红外光谱技术与多元统计分析相结合的中药葛根快速定量研究. *化学研究与应用*, 2024, 36(9): 1988–1996.
- [28] Qiao L, Yin Z Y, Duan H H, *et al.* Rapid and nondestructive identification of Dihuang (*Rehmannia glutinosa*) of different origins based on infrared spectroscopy and high performance liquid chromatography. *Chinese Archives of Traditional Chinese Medicine*, 2025, 43(3): 42–46, 261–267.  
乔璐, 阴紫钰, 段浩瀚, 等. 基于中红外光谱技术结合 HPLC 对不同产地地黄快速无损鉴别. *中华中医药学刊*, 2025, 43(3): 42–46, 261–267.