

基于非标准物质及小样本的近红外光谱设备选型方法研究

张宏^{1,2}, 徐芳芳^{1,2,3,4*}, 韩嘉丽³, 候化蕊³, 张欣^{1,3,4},
张永超³, 王振中^{1,2,3,4*}

(1. 中药制药过程控制与智能制造技术全国重点实验室(江苏康缘药业股份有限公司/南京中医药大学), 江苏 连云港 222047; 2. 南京中医药大学 康缘中药学院, 江苏 南京 210023; 3. 江苏康缘药业股份有限公司, 江苏 连云港 222047; 4. 中成药智能制造技术国家地方联合工程研究中心, 江苏 连云港 222047)

摘要: 针对制药工业近红外光谱(NIRS)设备试机期内标准物质匮乏及建模样本量受限的决策瓶颈, 该研究提出一套基于“4S”物料筛选原则与多准则综合评价模型(MCDA)的“吸光度动态重复性+小样本建模潜力”双重评价策略。选取3台不同分光原理的近红外光谱仪(傅里叶变换型A, 光栅型E与I), 以甘露醇、硬脂酸、聚乙烯3种非标准物质为载体评价设备吸光度精密密度; 同时以桂枝茯苓胶囊材水分定量为目标, 利用留一交叉验证法(LOOCV)构建小样本偏最小二乘(PLS)模型, 对比各设备的应用建模潜力。结果表明, 3台设备性能差异显著: 设备A展现出最优的吸光度精密密度与建模精度(RMSECV=0.1923, R^2_{cv} =0.9469), 其MCDA综合选型得分(92.93)显著领先, 体现了高信噪比硬件的性能优势; 设备E的吸光度重复性显著逊于其他设备, 其固有的硬件稳定性欠佳限制了模型潜力; 设备I在热敏性样品(硬脂酸)测试中暴露了显著的温控系统缺陷, 导致测量失效。该策略无需昂贵标准品, 且能在小样本条件下精准识别设备硬件隐患, 为制药企业的设备选型与风险评估提供了一种经济、高效且具备计量学可信度的决策参考。

关键词: 近红外光谱; 设备评价与选型; 吸光度重复性; 偏最小二乘法; 留一交叉验证

中图分类号: O657.3 **文献标识码:** A **文章编号:** 1004-4957(2026)09-1974-14

Methodology for NIR Spectrometer Selection Using Proxy Materials and Limited Sample Sets

ZHANG Hong^{1,2}, XU Fang-fang^{1,2,3,4*}, HAN Jia-li³, HOU Hua-rui³, ZHANG Xin^{1,3,4},
ZHANG Yong-chao³, WANG Zhen-zhong^{1,2,3,4*}

(1. State Key Laboratory of Technologies for Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture(Jiangsu Kanion Pharmaceutical Co., Ltd/Nanjing University of Chinese Medicine), Lianyungang 222047, China; 2. Kanion School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China; 3. Jiangsu Kanion Pharmaceutical Co., Ltd, Lianyungang 222047, China; 4. National & Local Joint Engineering Research Center of TCM Intelligent Manufacturing Technology, Lianyungang 222047, China)

Abstract: To address decision-making bottlenecks caused by scarce standard reference materials and limited sample sizes during near-infrared spectroscopy(NIRS) equipment trials in the pharmaceutical industry, this study proposes a dual evaluation strategy combining “absorbance dynamic repeatability” and “small-sample modeling potential”. This strategy is based on the “4S principles” for material screening and a multicriteria decision analysis(MCDA) model. Three NIRS instruments with different spectroscopic principles—Fourier transform(A) and grating-based(E and I)—were selected to evaluate absorbance precision using three nonstandard materials (mannitol, stearic acid, and polyethylene) as carriers. Simultaneously, targeting the moisture quantification of Guizhi Fuling

收稿日期: 2026-03-12; **修回日期:** 2026-04-01

基金项目: 国家长三角科技创新共同体联合攻关项目(2023CSJGG1700); 国家工信部产业基础再造和制造业高质量发展专项(TC2308068); 中药制药过程控制与智能制造技术全国重点实验室开放基金课题(SK12023D02003)

* **通讯作者:** 徐芳芳, 博士, 副主任药师, 研究方向: 中药智能制造研究, E-mail: 879164331@qq.com
王振中, 博士, 研究员, 研究方向: 中药新药创制与过程控制研究, E-mail: kyyywzz@163.com

网络首发日期: 2026-08-03

Capsules soft material, small-sample partial least-squares models were constructed via leave-one-out cross-validation to compare the application modeling potential of each instrument. The results revealed significant performance disparities: Instrument A exhibited optimal absorbance precision and modeling accuracy ($RMSECV=0.192\ 3$, $R^2_{cv}=0.946\ 9$), with its integrated MCDA score (92.93) leading significantly, thereby reflecting the hardware advantages of a high signal-to-noise ratio. Instrument E exhibited significantly inferior absorbance repeatability compared with the other devices, with its inherent hardware instability constraining its modeling potential. Instrument I revealed significant defects in its temperature control system during tests on a thermosensitive sample (stearic acid), thus resulting in measurement failure. This strategy requires no expensive standard reference materials and accurately identifies hardware risks using small sample sizes, thus providing an economical, efficient, and metrologically credible decision-making reference for equipment selection and risk assessment in pharmaceutical enterprises.

Key words: near-infrared spectroscopy; instrument evaluation and selection; absorbance repeatability; partial least squares; leave-one-out cross-validation

近红外光谱(NIRS)技术作为过程分析技术(PAT)的关键组成部分,已深度融入现代制药的检测与质控等方面。在制药企业或科研单位进行设备选型时,通常存在两种主流的评价方式。其一是基于《中华人民共和国药典》、T/CIS 17006-2022等标准的性能指标评估。这些指标偏向于波长准确度、信噪比等“出厂规格”的静态物理参数^[1-2],反映了仪器在理想条件下的理论性能上限。其二,是基于实际样品的应用建模潜力评估,通过对比不同设备对特定组分(如药物活性成分、水分)的预测能力进行优劣判定。例如,Czaja、王冬及Yan等^[3-18]分别利用超过百个样本集,成功比较了不同类型光谱仪(如台式、便携式)在特定应用(如食品成分、药物活性成分分析)中的预测能力,证明了该方法在反映设备系统适应性方面的优势。

然而,在通常仅为数周的设备试机期内,上述方式均面临实操困境。首先,标准物质(如SRM 2035b)不仅采购成本高昂且物流周期较长、维护条件苛刻,更关键的是设备静态指标往往难以揭示设备在连续运行下的动态稳定性及对特定物料的适应性。其次,依赖大样本建模的应用潜力评估方法,往往缺乏可行性,用户通常难以在短时间内获得大量具有代表性的样品。若将样本机械划分为训练集和验证集,则会导致样本量过小,模型极易产生过拟合^[19-21],其评估结果的稳健性和可靠性将大打折扣。因此用户陷入了一个两难困境:既无法在短期内有效复现官方规范的完整性能验证流程以及完全依赖设备厂商提供的性能指标,也因数据量难以进行可靠的模型潜力评估。

为应对上述“试机期短、样本量少、标准品难求”的决策挑战,本研究旨在建立一套以应用为导向的补充性评价策略。本研究的核心创新在于系统性地提出并验证了“非标准评价物料选取四原则(4S principles)”以及基于分层加权的“多准则综合评价模型(MCDA)”,其包含两个核心评估维度:其一,基于“4S原则”的动态重复性考察。选取聚乙烯、硬脂酸与甘露醇3种具有代表性的非标准物质,在模拟负载下量化评估动态稳定性。其二,基于留一交叉验证(LOOCV)的小样本建模评估。以少量生产物料(桂枝茯苓胶囊软材细粉)为对象,旨在最大化利用有限样本信息,从而对设备的真实建模潜力做出初步但相对可靠的判断。最终将吸光度精密度与建模潜力指标通过MCDA模型转化为直观、量化的选型分值。本策略力求为制药领域的设备选型提供一种经济、高效且具备计量学可信度的决策参考。

1 实验部分

1.1 仪器与试剂

DHG-9145A 鼓风干燥箱(上海一恒科学仪器有限公司),电子天平(ME104E,梅特勒-托利多仪器上海有限公司)。

本研究选取3台具有代表性的近红外光谱仪,分别为傅里叶变换型(设备A)及便携式仪器(设备E、设备I)。为保护相关方商业隐私,文中隐匿其具体型号,设备核心技术参数详见表1,均由江苏康缘药业股份有限公司提供。

表1 实验中所使用的3台近红外设备基础信息

Table 1 Specifications of three near-infrared instruments used in current study

Instrument parameter	Instrument A	Instrument E	Instrument I
近红外类型	傅里叶变换型	光栅扫描型	光栅扫描型
光谱测量范围	4 000~10 000 cm^{-1}	1 000~1 800 nm	900~1 700 nm
光谱分辨率	8 cm^{-1}	10 nm	12 nm
检测器类型	铟镓砷检测器	铟镓砷检测器	铟镓砷检测器
光源类型	卤素灯	卤钨灯	卤钨灯
采样附件	样品杯	光纤、探头	样品杯
采集方式	下照式漫反射	光纤漫反射	下照式漫反射

sample cup of Instrument A: diameter approx. 60 mm, material: quartz. sample cup of Instrument I: effective width 45 mm, length approx. 120 mm, material: quartz(设备A样品杯的尺寸: 直径约为60 mm, 材质: 石英; 设备I样品杯的尺寸: 有效宽度45 mm, 长度约120 mm, 材质: 石英)

甘露醇(国药集团化学试剂有限公司, 批号20120828), 硬脂酸(湖州展望药业有限公司, 批号20230705), 口服固体药用高密度聚乙烯瓶盖(安徽迎驾贡酒股份有限公司金冠容器分公司, 批号210913), 桂枝茯苓胶囊软材细粉(江苏康缘药业股份有限公司, 批号: 240801~240807、240809~240810、241001~241015、241101~241108、241201、241203~241215、250101~250106、250201~250207、250301~250310、250402~250411、250413)。

1.2 实验方法

1.2.1 水分参考值测定方法

采用第二法(烘干法)测定待测样本水分。使用鼓风干燥箱进行实验, 温度设置为105 $^{\circ}\text{C}$ 。精密称取约2.0 g样品, 均匀平铺于已干燥至恒重的称量瓶中。将开启瓶盖的称量瓶置于105 $^{\circ}\text{C}$ 的鼓风干燥箱中干燥5 h。随后, 盖好称量瓶盖, 转移至干燥器中冷却30 min后精密称重。再次于上述温度下干燥1 h, 放冷、称重, 直至连续两次称量结果的差异不超过5 mg。根据减少的重量计算样品中的含水量^[1]。

1.2.2 光谱扫描方法与数据采集

所有实验均在恒温恒湿实验室中完成(温度: 20~25 $^{\circ}\text{C}$, 相对湿度: 35%~45%, 无强光照射), 为确保设备运行状态稳定, 3台近红外光谱仪开机后均按照操作规范预热30 min。具体采集参数设定遵循设备工程师推荐的应用方案:

设备A: 采用积分球漫反射测量模式。以空气为背景, 每小时复测一次背景光谱。光谱扫描范围为4 000~10 000 cm^{-1} , 分辨率8 cm^{-1} , 扫描次数64次。增益由系统根据物料响应自动优化。每批样品取约8 g置于专用旋转样品杯中采集。

设备E: 采用探头直接接触式测量。光谱范围1 000~1 800 nm, 光谱平均次数30次, 通道增益设定为1。每个样品取约20 g, 探头紧贴样品表面进行采样。

设备I: 采用样品杯漫反射测量。光谱范围900~1 700 nm。采集模式设定为“一参多样”, 即每扫描一次参比后连续测量多个样品。每批样品取约20 g置于配套样品杯中, 光谱平均次数90次, 子样数设为3。

硬脂酸、甘露醇、聚乙烯分别连续采集50张光谱, 用以评估吸光度重复性; 枝茯苓胶囊软材细粉采集3张光谱, 取其平均光谱用于后续水分定量模型的构建与性能评价。

本研究未对3台设备的采集参数(扫描次数)进行机械化统一, 而是采用厂商针对药用粉末推荐的最优参数设置。此举旨在模拟实际工业试机场景: 用户通常在设备预设的最优参数下评估其性能。此外, 各设备分光元件及探测器响应特性的差异决定了其参数设置的差异性, 采用推荐方案更能客观反映该系统在实际应用中的性能上限。

1.2.3 奇异样本识别与筛选

在光谱数据采集过程中, 受环境噪声或样品物理状态等非化学组成因素影响, 光谱数据易产生干扰建模性能的奇异样本。为提升模型稳健性, 本研究采用蒙特卡洛偏最小二乘法(MCPLS)进行异常样本的识别。该方法的原理是通过蒙特卡洛交叉验证构建大量随机子模型, 并计算每个样本在所有模型中的预测残差分布。依据残差分布的均值与标准差, 可有效区分并同时识别出异常样本^[22-24]。以均值

误差和标准差误差为维度构建二维误差统计特征空间,为充分考虑预测误差均值与标准差之间的协方差结构及其统计相关性,本文引入马氏距离作为判定依据。在二维特征空间中,该阈值表现为以数据中心为原点、向数据分布主成分方向延伸的自适应置信椭圆。为合理确定奇异值识别阈值,本文采用梯度策略(99%、97.5%、95%)考察了置信水平。落于置信椭圆外的样本初步识别为潜在奇异样本。奇异样本的最终剔除遵循性能导向原则。通过对比原始模型与剔除异常值后模型内部交叉验证均方根误差的变化来决定剔除方案,确保剔除操作在显著优化模型精度的同时,不损害样本集的代表性^[25]。

1.2.4 光谱预处理与特征变量筛选

桂枝茯苓软材细粉的物理特性(如颗粒大小、分布、形状、堆积密度等)会导致光谱的基线漂移和散射效应,这些物理干扰会掩盖化学信息,影响模型的稳健性和预测精度。为了消除或减弱由于物理因素引起的光谱变异,从而增强与待测成分相关的化学信息^[26-27],文中使用以下方法对原光谱进行预处理:归一化法、移动平均平滑法、Savitzky-Golay平滑法、Savitzky-Golay导数法、去趋势校正法、多元散射校正法(MSC)、标准正态变换法(SNV)、基线校正法。

光谱数据中存在大量冗余和无用信息,利用合适的特征波长筛选方法能够进一步提高模型稳健度与准确性^[28]。为确保为每台仪器均构建最优模型,本实验在筛选出最佳光谱数据预处理方法的基础上,采用以下5种特征变量选择方法进行变量筛选^[22]:竞争自适应重加权采样(CARS)、连续投影算法(SPA)、反向区间偏最小二乘(BiPLS)、正向区间偏最小二乘(FiPLS)、递归加权偏最小二乘(rPLS)^[29]。

所有预处理及特征变量筛选方案的优选均遵循“最小RMSECV与最大 R^2_{cv} ”原则,旨在为3台不同近红外设备构建各自最优模型。

1.2.5 性能评价指标

1.2.5.1 吸光度精密度评价指标 为考察仪器在连续运行下的动态重复性及精密度演变规律,对硬脂酸、甘露醇、聚乙烯分别连续采集50张光谱后,对同一样品在每个波长点上,根据公式(1)计算出前6张光谱、前10张光谱、前25张光谱以及全部光谱在不同波长下的标准差(STD)。

$$STD_{\lambda} = \sqrt{\frac{\sum_{i=1}^n (A_{i,\lambda} - \bar{A}_{\lambda})^2}{n-1}} \quad (1)$$

$$STD_{Avg} = \frac{1}{m} \sum_{\lambda=1}^m STD_{\lambda} \quad (2)$$

其中, STD_{λ} 为在某一特定波长点上的吸光度标准差; n 为用于计算标准差的光谱张数; $A_{i,\lambda}$ 为该波长点第*i*张测量的吸光度值; $\bar{A}_{\lambda}$ 为该波长点*n*张测量的吸光度平均值。随后根据公式(2),计算全波段*m*个波长点的平均标准差(STD_{Avg}),作为评价仪器整体吸光度精密度水平的核心指标。

1.2.5.2 小样本建模评价指标 本研究建立模型的数据量仅80个,为最大化利用样本信息并客观评估模型稳健性,采用留一交叉验证法构建水分定量模型,并选取以下指标评价模型性能:

交叉验证均方根误差(RMSECV):反映模型对未知样本预测的准确性。其值越小且与校正集均方根误差(RMSEC)越接近,表明模型的泛化能力越强,过拟合风险越低。

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{i,-i})^2} \quad (3)$$

式中, n 为校正集样本总数, y_i 为第*i*个样本的实测值(如水分含量), $y_{i,-i}$ 为第*i*个样本在不含该样本的交叉验证模型中的预测值。

交叉验证决定系数(R^2_{cv}):衡量模型预测值对观测值变异的解释程度。其越接近1,表明模型对于水分含量的拟合优度越高,线性相关性越强。

$$R^2_{cv} = 1 - \frac{\sum_{i=1}^n (y_i - y_{i,-i})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

式中,分子为交叉验证预测残差平方和,分母为实测值总离差平方和。

变量保留率(μ):用于评价特征波长筛选的效果及模型的稀疏性。适当的保留率能在剔除冗余信

息、降低模型复杂度的同时,避免丢失关键化学特征信息^[30]。计算公式如下:

$$\mu = \frac{k}{n} \times 100\% \quad (5)$$

式中, μ 为变量保留率, k 为筛选后保留的特征变量个数, n 为原始全光谱的总变量个数。

以上指标不仅用于优选最佳预处理方法和特征变量,亦作为评价 3 台设备在实际应用场景下性能差异的最终依据。

1.2.5.3 多准则综合评价 为量化设备硬件性能与应用潜力的综合差异,本研究提出一套分层加权评价矩阵。最终综合得分(S_{Total})通过吸光度精密度分(S_{STD})与小样本建模潜力分(S_{Model})加权计算得出。

$$S_{\text{Total}} = W_1 \cdot S_{\text{STD}} + W_2 \cdot S_{\text{Model}} \quad (6)$$

式中, W_1 与 W_2 分别为吸光度精密度与小样本建模潜力权重系数(取值范围在 0~1),本研究分别取值为 0.4 与 0.6,主要基于以下考量:吸光度精密度是设备可靠运行的“基石”,若设备在试机期表现出明显的动态不稳定性,其后续运行的可靠性将无法保障;建模潜力作为“核心目标”,直接反映了设备对软材细粉水分化学特征的提取能力与定量精度。

其中 S_{STD} 为 K 种非标准物质在吸光度动态重复性测试过程中的综合得分均值:

$$S_{\text{STD}} = \frac{1}{K} \sum_{k=1}^K S_{\text{mat},k} \quad (7)$$

$$S_{\text{mat},k} = \frac{1}{4} \sum_{n \in (6, 10, 25, 50)} f(\text{STD}_{n-\text{avg},k}) \quad (8)$$

S_{Model} 综合考虑了预测精度与相关性:

$$S_{\text{Model}} = w_R \cdot f(\text{RMSECV}) + w_D \cdot f(R_{\text{cv}}^2) \quad (9)$$

为确保预测精度作为首要判据,本研究中 w_R 与 w_D 分别为 0.7 与 0.3。

考虑到各指标量级及优劣趋势的差异,引入归一化处理 $f(x_i)$ 函数,该函数采用相对评分法进行计算,对于“越小越好”指标采用公式(10),相反则采用公式(11)。

$$f(x_i) = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \times 100 \quad (10)$$

$$f(x_i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \times 100 \quad (11)$$

其中 x_i 为当前设备测量值, $x_{\max}$ 与 $x_{\min}$ 分别为 3 台设备中该指标的最大值与最小值。本评价策略的标准化操作流程(SOP)图如图 1 所示。

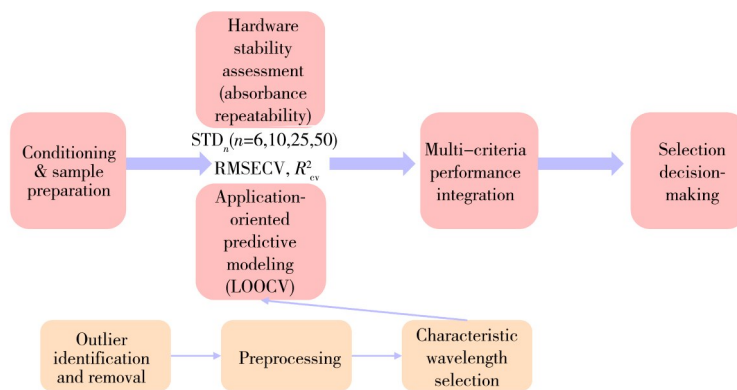


图 1 评价策略的标准化操作流程(SOP)图

Fig. 1 Standard operating procedure(SOP) diagram of the proposed evaluation strategy

文中吸光度精密度计算、异常值筛选与特征波长提取使用 Matlab 2024b 软件(MathWorks, 美国),光谱数据预处理与模型构建使用 Unscrambler 10. X 软件(Camo Analytics, 挪威),图像绘制使用 Origin 软件。

2 结果与讨论

2.1 原始光谱与奇异样本筛选

桂枝茯苓胶囊软材细粉在3台设备(A、E、I)上的原始光谱如图2所示。由图可知,由于软材细粉性状导致的光谱散射效应,3台设备的光谱均存在不同程度漂移。在1 450 nm及1 900 nm附近(对应FT-NIR的7 000 cm^{-1} 与5 200 cm^{-1} 区域)可观察到明显的O—H合频与倍频吸收带。

为提高建模稳健性,采用MCPLS法对80个样本进行奇异样本识别,根据梯度策略考察置信水平,结果发现设备A与E选择99%置信水平即可识别出奇异样本,而设备I使用较为宽松的99%阈值会出现漏选现象,更加严苛的95%阈值会导致部分包含核心化学信息的正常样本被错误划定。因此,最终使用97.5%置信水平构建判定椭圆。将识别出的潜在奇异样本进行剔除验证,结果见表2。

由表2可知,在设备I中单独剔除241101样本,RMSECV不降反升,表明该样本虽处于统计边界边缘,但仍包含关键的建模信息,故予以保留。最终设备A选择剔除样本241209,设备E剔除样本241014与240807,设备I剔除样本241209、241014与241204。在剔除奇异样本的过程中, R^2_{c} 与 R^2_{cv} 并不能展示剔除奇异样本后的性能变化,例如设备A剔除奇异样本后, R^2_{cv} 由0.911 6变为0.890 5。这一现象源于被剔除的奇异样本通常处于浓度空间分布的边缘,其存在会夸大数据集的总变异,从而导致剔除前 R^2 的计算值虚高。因此,本研究为确保剔除奇异样本后的模型稳健性与精度,选择使用RMSECV作为指标确定需要剔除的样本。

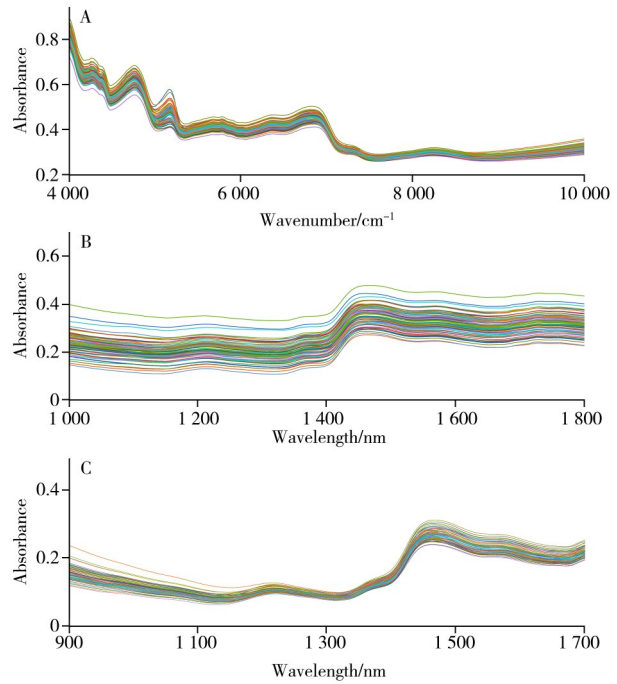


图2 3台不同设备采集的原始近红外光谱
Fig. 2 Raw NIR spectra acquired from three different instruments
A. Instrument A; B. Instrument E; C. Instrument I

表2 使用MCPLS方法进行奇异样本剔除前后,不同仪器PLS模型的结果对比

Table 2 Comparison of PLS models before and after singular sample elimination using MCPLS method

Instrument	Selected sample	RMSEC	RMSECV	R^2_{c}	R^2_{cv}	LV
A	Raw	0.226 8	0.278 3	0.939 8	0.911 6	7
	241209	0.223 0	0.276 1	0.926 7	0.890 5	7
E	Raw	0.358 6	0.436 5	0.849 4	0.782 6	9
	241014	0.355 1	0.433 6	0.853 1	0.786 6	9
	240807	0.340 0	0.404 8	0.865 5	0.814 2	9
	241014, 240807	0.338 0	0.404 5	0.867 8	0.815 5	9
I	Raw	0.426 5	0.497 7	0.787 1	0.717 4	5
	241209	0.402 2	0.467 2	0.761 4	0.686 2	5
	241101	0.427 9	0.501 8	0.787 7	0.714 6	5
	241014	0.400 9	0.470 1	0.805 3	0.749 1	5
	241204	0.416 9	0.491 7	0.779 7	0.701 4	5
	241209, 241014, 241204	0.376 7	0.436 6	0.766 9	0.695 1	5

2.2 模型优化与建立

2.2.1 光谱预处理

针对软材细粉复杂的光学物理背景,本研究对比了原始光谱及经9种预处理算法处理后的PLSR模型性能,结果见表3。实验发现,不同硬件原理的设备对预处理算法的敏感度存在显著差异。设备A采用归一化法处理后,模型性能提升最为显著,其RMSECV降至0.236 5,且 R^2_{cv} 较原始模型有明显提高。这表明归一化法有效消除了FT-NIR在积分球漫反射测量时,由于粉体样本装填时引起的光程差异及光谱强度波动^[27]。对于设备E,最佳预处理方法是移动平移平均法,这说明通过平滑处理可有效滤除样本表面不均带来的高频随机噪声,提高光谱信号的平稳性^[27]。对于设备I,去趋势法处理表现最优,由于设备I采用样品杯漫反射测量,去趋势法能有效校正所受到样本颗粒间隙与紧实度差异导致的非线性基线漂移^[27],从而显著优化了交叉验证集的稳定性和预测精度。

表3 不同设备使用不同光谱预处理方法下水分含量的PLS模型性能对比

Table 3 Comparison of PLS model performance for moisture-content prediction using different spectral preprocessing methods across various instruments

Instrument	Preprocessing method	RMSEC	RMSECV	R^2_c	R^2_{cv}	LV
A	None	0.223 0	0.276 1	0.926 7	0.890 5	7
	Moving average smoothing	0.223 3	0.276 0	0.926 4	0.890 5	7
	Savitzky-Golay smoothing	0.223 5	0.275 8	0.926 3	0.890 6	7
	Savitzky-Golay derivative	0.193 1	0.260 0	0.945 0	0.902 8	5
	Normalization	0.188 8	0.236 5	0.947 4	0.919 6	8
	Baseline	0.231 1	0.275 0	0.921 2	0.891 3	5
	SNV	0.198 3	0.242 0	0.942 0	0.915 8	6
	De-trend	0.256 2	0.287 7	0.903 2	0.881 0	3
	MSC	0.259 8	0.292 4	0.900 5	0.877 1	4
E	None	0.338 0	0.404 5	0.867 8	0.815 5	9
	Moving average smoothing	0.339 1	0.403 7	0.866 8	0.816 2	9
	Savitzky-Golay smoothing	0.338 6	0.404 5	0.867 3	0.815 5	9
	Savitzky-Golay derivative	0.271 7	0.411 2	0.914 6	0.809 4	4
	Normalization	0.337 2	0.439 9	0.868 3	0.781 8	10
	Baseline	0.347 5	0.412 4	0.860 2	0.808 2	8
	SNV	0.389 3	0.459 9	0.824 6	0.761 5	7
	De-trend	0.363 5	0.420 6	0.847 1	0.800 5	6
	MSC	0.338 0	0.404 5	0.867 8	0.815 5	9
I	None	0.376 7	0.436 6	0.766 9	0.695 1	5
	Moving average smoothing	0.378 3	0.438 1	0.764 9	0.693 0	5
	Savitzky-Golay smoothing	0.376 9	0.436 7	0.766 7	0.695 0	5
	Savitzky-Golay derivative	0.397 9	0.447 3	0.739 9	0.679 9	2
	Normalization	0.393 0	0.447 4	0.746 3	0.679 7	4
	Baseline	0.390 1	0.440 0	0.750 1	0.690 2	5
	SNV	0.390 6	0.428 3	0.749 4	0.706 6	3
	De-trend	0.350 9	0.408 5	0.797 8	0.733 0	5
	MSC	0.396 5	0.432 0	0.741 8	0.701 5	3

2.2.2 水分特征变量筛选

在光谱预处理的基础上,研究对比了3类特征变量筛选算法对模型性能的影响,结果见表4。结果表明,特征变量筛选显著降低了模型维度,剔除了大量与水分无关的冗余信息。对于高分辨率的设备A, CARS算法表现出最优的变量保留率与模型性能提升程度,不仅将设备A变量数从全波段缩减至30个,且RMSECV达到了最低值,这表明该算法在处理复杂的高维数据时具有更强的特征提取能力。相比之下,设备I在应用FiPLS算法后模型RMSECV降低与 R^2_{cv} 增加最为显著,但是此时其潜变量数却提升至7,这对于只有80个样本的小型数据集,潜变量数过高意味着模型抓取了噪音等特征^[31],而非水分的共性化学信息。这会导致模型在预测未知样品时,误差显著增大,同时也不利于模型向未知样品的迁移应用^[32]。图3展示了各算法筛选出的特征波长在原始光谱上的分布。水分在近红外区的吸收受O—H键一级倍频(1 440~1 550 nm)与合频(1 800~2 100 nm)主导。设备A的筛选结果完美覆盖了这两处强吸收区;值得注意的是,受波长范围限制(1 000~1 800 nm与900~1 700 nm)的便携式设备E与I,其5种筛选方法均能一致地锁定1 440~1 550 nm范围内的特征带,这论证了便携式光栅设备在桂枝茯苓胶囊软材细粉样品水分监测任务中的适用性。

表4 不同设备使用不同光谱变量筛选方法下水分含量的PLS模型性能对比

Table 4 Comparison of PLS model performance for moisture-content prediction using different variable selection methods across various instruments

Instrument	Wavelength selection method	RMSEC	RMSECV	R^2_c	R^2_{cv}	LV	Number of selected variables	Variable retention rate/%
A	None	0.188 8	0.236 5	0.947 4	0.919 6	8	1 556	
	FiPLS	0.203 6	0.255 1	0.938 8	0.906 5	8	604	38.82
	BiPLS	0.186 5	0.230 3	0.948 7	0.923 8	10	228	14.65
	rPLS	0.179 3	0.207 2	0.952 6	0.938 3	7	12	0.77
	CARS	0.165 9	0.192 3	0.959 4	0.946 9	6	30	0.39
	SPA	0.192 1	0.244 9	0.945 6	0.913 8	8	76	4.88
E	None	0.339 1	0.403 7	0.866 8	0.816 2	9	800	
	FiPLS	0.349 8	0.406 8	0.858 4	0.813 4	8	126	15.75

(续表4)

Instrument	Wavelength selection method	RMSEC	RMSECV	R^2_c	R^2_{cv}	LV	Number of selected variables	Variable retention rate/%
I	BiPLS	0.332 2	0.391 7	0.872 2	0.827 0	9	382	47.75
	rPLS	0.349 4	0.399 9	0.858 7	0.819 6	7	9	1.13
	CARS	0.332 8	0.384 3	0.871 8	0.833 5	8	16	2.00
	SPA	0.340 4	0.405 9	0.865 9	0.814 2	9	71	8.86
	None	0.350 9	0.408 5	0.797 8	0.733 0	5	801	
	FiPLS	0.326 2	0.374 4	0.825 3	0.775 8	7	140	17.48
	BiPLS	0.360 6	0.411 3	0.786 4	0.729 3	4	577	72.03
	rPLS	0.355 28	0.393 8	0.792 7	0.752 0	4	69	8.61
	CARS	0.351 4	0.392 2	0.797 2	0.759 8	4	89	11.11
	SPA	0.362 5	0.411 1	0.784 1	0.729 6	4	64	7.99

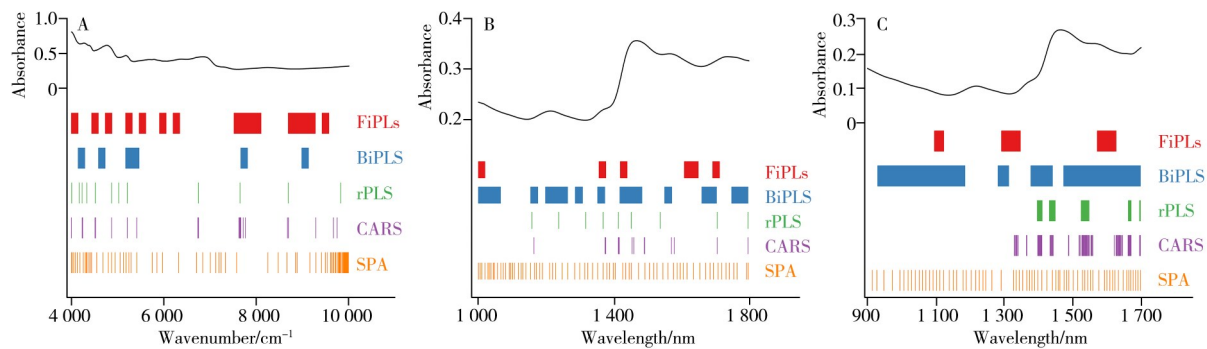


图3 不同设备原始光谱上不同变量筛选方法选出的特征波长分布

Fig. 3 Distribution of characteristic variables selected using different variable selection methods on raw spectra of different instruments

A. Instrument A; B. Instrument E; C. Instrument I

2.3 性能评价与设备选型

2.3.1 3台设备吸光度精密度比较

3种非标准物质在不同采样张数($n=6, 10, 25, 50$)下的全波段STD曲线如图4~图6的B、D、F所示, 平均吸光度标准差统计结果见表5, 本研究系统评价了3台设备的动态重复性与精密度演变趋势。

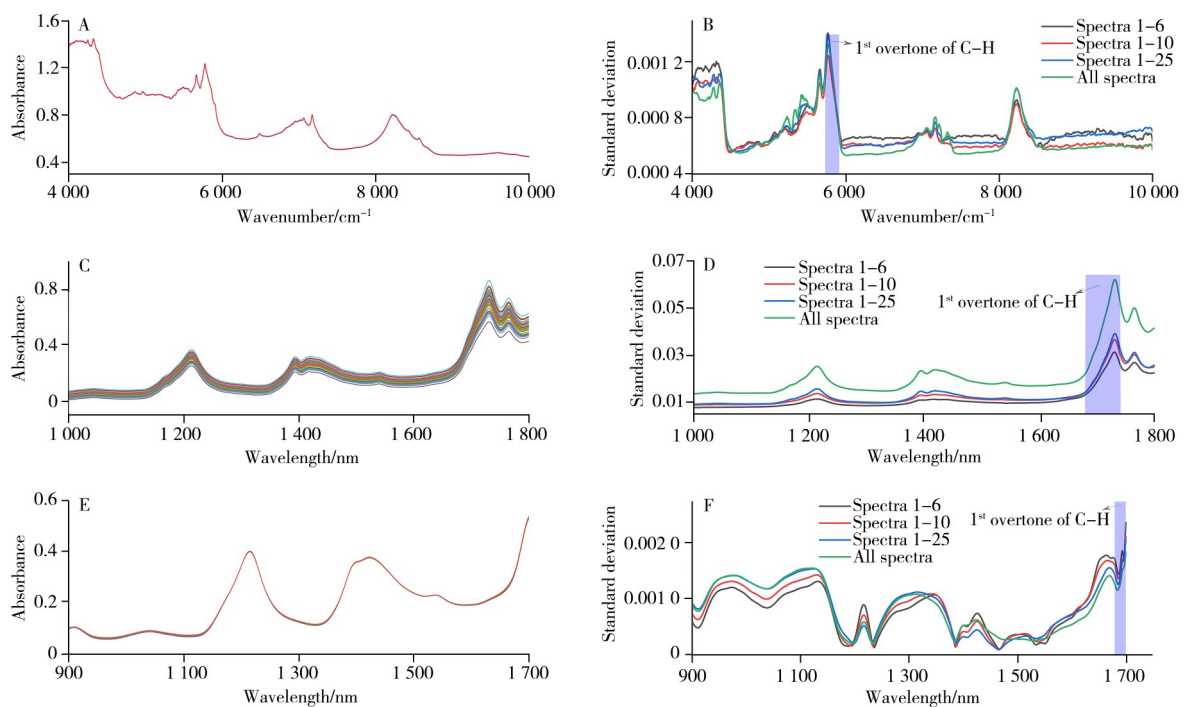


图4 聚乙烯吸光度重复性测试结果

Fig. 4 Repeatability test results of polyethylene absorbance

A, C, E: raw absorbance spectra; B, D, F: full-spectrum standard deviation(STD) curves; A, B correspond to Instrument A; C, D to Instrument E; and E, F to Instrument I(A, C, E: 原始吸光度光谱; B, D, F: 全波段STD曲线; 设备A: A、B, 设备E: C、D, 设备I: E、F)

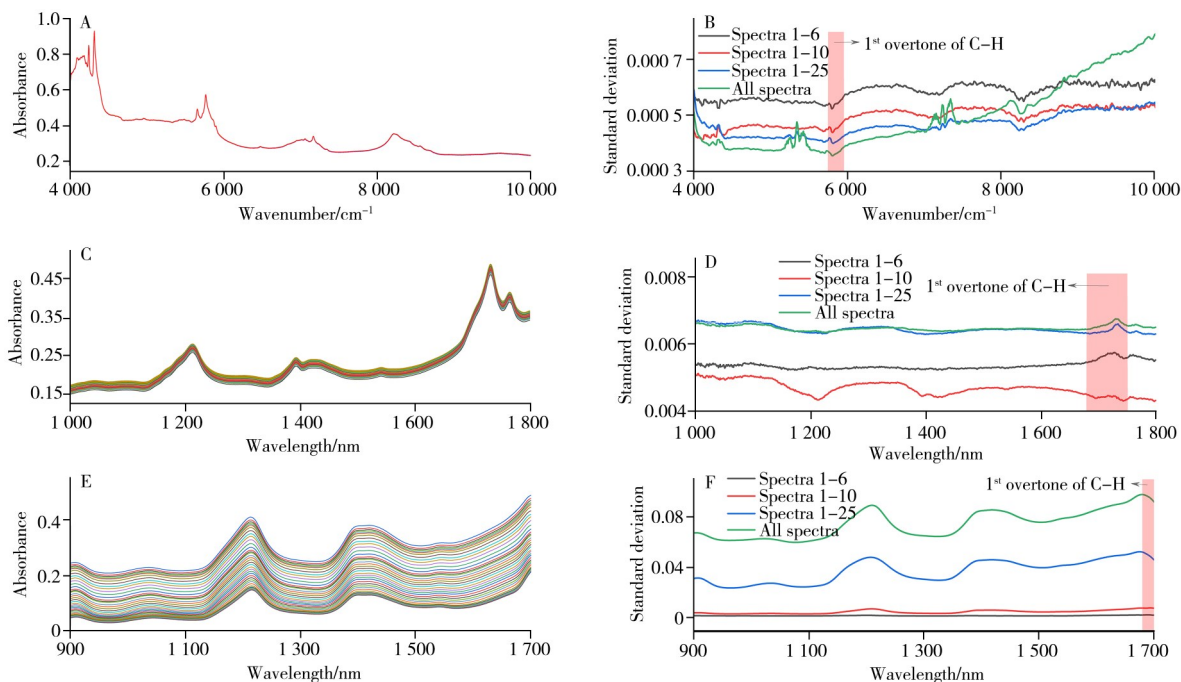


图 5 硬脂酸吸光度重复性测试结果

Fig. 5 Repeatability test results of stearic acid absorbance

A-F are the same as those in Fig. 4

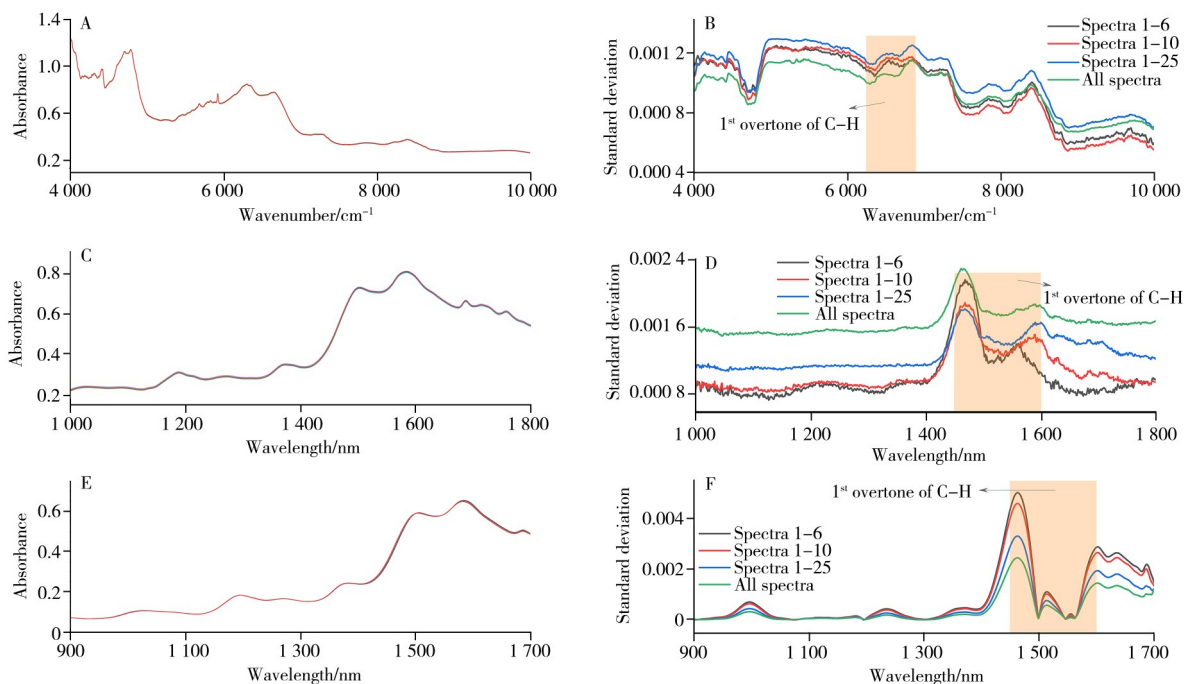


图 6 甘露醇吸光度重复性测试结果

Fig. 6 Repeatability test results of mannitol absorbance

A-F are the same as those in Fig. 4

2.3.1.1 仪器精密度与稳定性对比 实验结果显示, 设备 A 在面对不同物理性状(熔点、晶体结构)的物料时, 其 STD 曲线始终保持在极低水平($< 1 \times 10^{-3}$)且随波长分布平稳。这表明傅里叶变换光谱仪由于具备较高的光通量和稳定性, 能够提供极高的基准信噪比, 受物料特性干扰较小。需要指出的是, 针对 3 台设备扫描参数设定不一致的情况(A 为 64, E 为 30, I 为 90), 理论上随机噪音会随着次数 n 的增加而呈现 $\sqrt{n}$ 倍的降低。然而表 5 数据显示, 设备 A 吸光度标准差的差异远远超过了次数参数设定的增益。例如在聚乙烯的测试中, 设备 A 的 STD_{50-avg} (6.75×10^{-4}) 比扫描次数较少的设备 E 的 STD_{50-avg} (2.14×10^{-2}) 低约两个数量级。这表明设备 A 吸光度精密度的优势来源于其作为干涉仪的高通量以及傅里叶变

换的多路复用优势，而并非单纯的次数增加。

2.3.1.2 极端工作负荷下的热稳定性评估 在硬脂酸重复性测试中，设备I展现出严重的热稳定性缺陷。由图7可见，连续扫描次数增加至第39张，由于光源辐射热量持续累积且散热设计不佳，导致接触点处的硬脂酸出现熔化现象。这一现象直接触发了原始光谱吸光度的急剧变化和全波段不规则波动(见图5C)，使STD值呈指数级急剧上升。该发现证明了：在高频次采集场景下，设备的光源散热管理直接决定了测量结果的真实性与可复现性。

2.3.1.3 物料散射特性与光谱量程的交互影响 面对强散射晶体物质(甘露醇)，设备E与I在其特征吸收范围的STD均出现峰值波动，这反映了便携式分光系统在处理强散射信号时，光电转换的误差增大。值得注意的是，设备I在聚乙烯测试中的STD表现看似与设备A持平(见图4B、F)，但结合其光谱响应范围(900~1 700 nm)分析可知，聚乙烯在1 680~1 750 nm区域(C—H键伸缩振动的一级倍频区)的核心特征信息已被该设备的硬件量程切断。这种盲区掩盖了设备在特征区的真实表现，进一步凸显了在设备选型时，必须结合物料特征波段与设备有效量程进行综合考量的必要性，有效量程所带来的化学信息可用性是影响设备性能的最重要因素之一^[33]。



图7 硬脂酸在设备I样品盘中出现熔化现象

Fig. 7 Melting phenomenon of stearic acid observed in sample cell of Instrument I

表5 平均吸光度标准差统计结果

Table 5 Statistical results of mean absorbance standard deviation

Material	Instrument	Spectra 1-6	Spectra 1-10	Spectra 1-25	All spectra
Mannitol	A	9.67×10^{-4}	9.51×10^{-4}	1.00×10^{-3}	9.46×10^{-4}
	E	1.01×10^{-3}	1.07×10^{-3}	1.28×10^{-3}	1.65×10^{-3}
	I	9.14×10^{-4}	8.38×10^{-4}	6.10×10^{-4}	4.60×10^{-4}
Polyethylene	A	7.33×10^{-4}	6.84×10^{-4}	7.20×10^{-4}	6.75×10^{-4}
	E	1.17×10^{-2}	1.32×10^{-2}	1.38×10^{-2}	2.14×10^{-2}
	I	8.05×10^{-4}	8.46×10^{-4}	8.75×10^{-4}	8.51×10^{-4}
Stearic	A	5.84×10^{-4}	4.94×10^{-4}	4.66×10^{-4}	4.96×10^{-4}
	E	5.36×10^{-3}	4.68×10^{-3}	6.44×10^{-3}	6.48×10^{-3}
	I	1.95×10^{-3}	4.98×10^{-3}	3.72×10^{-2}	7.46×10^{-2}

2.3.2 桂枝茯苓胶囊软材细粉水分建模结果比较

桂枝茯苓胶囊软材细粉水分定量的内部交叉验证结果如表6示。从预测精度来看，设备A表现出最优的统计性能，其RMSECV为0.192 3，显著低于设备E与I，其 R^2_{cv} 达到了0.946 9，表明其光谱信息与水分参考值之间具有极高的相关性。在模型复杂度方面，设备A最终在使用较少主成分数(LVs=6)的情况下即达到了最优精度，体现了其获得数据的高纯净度。一般来说，波长越大，波段越灵敏和更具体，因此FT-NIR仪器全范围始终能提供更准确的结果^[14]。虽然便携式设备E与I的RMSECV略逊于设备A，但通过CARS筛选后，变量保留率均达到10%左右(见表4)，这证明了光栅式便携设备在处理特定水分特征变量时的灵敏度足以支撑定量分析需求。其性能稍弱，可能归因于这类设备极具经济性的设计理念^[15,18,33]。需要指出的是，作为获得的光谱信息明显少于设备A的设备E，其在建模过程中，所使用的潜变量数却高于设备A。这种复杂性则需要从背景获取效应等方面解释，详情参考Cristina等^[16]的研究，此处不再进行赘述。便携式设备核心特征响应的特性，结合其体积小、成本低廉的优势，使其在制药车间的快速巡检及过程监测中展现出显著性价比优势。

2.3.3 吸光度动态重复性与建模潜力量化评价

利用“1.2.5.3”构建的MCDA，3台设备的各项性能得分及最终综合评分以及3台设备性能特征画像(雷达图)与MCDA综合评分(柱状图)如表7、图8所示。

2.3.3.1 吸光度动态重复性得分(S_{sm})分析 设备A在聚乙烯与硬脂酸的测试中展现出最佳水准。设备I在甘露醇与聚乙烯的测试中获得极高得分，受限于前文所述散热缺陷，在最终分层加权矩阵中，

其硬脂酸的得分产生较为极端的木桶效应,使其该项得分落后于设备 A。设备 E 由于整体基线噪声水平较高,在 3 种物料的测试中处于劣势。

表 6 桂枝茯苓胶囊软材细粉水分定量的内部交叉验证结果

Table 6 Internal cross-validation results for moisture quantification in soft material powder of Guizhi Fuling capsules

Instrument	Model	RMSEC	RMSECV	R^2_c	R^2_{cv}	LV
A	None-PLS	0.223 0	0.276 1	0.926 7	0.890 5	7
	Normalization + CARS + PLS	0.165 9	0.192 3	0.959 4	0.946 9	6
E	None-PLS	0.338 0	0.404 8	0.867 8	0.815 5	9
	Moving Average Smoothing + CARS + PLS	0.332 8	0.384 3	0.871 8	0.833 5	8
I	None-PLS	0.376 7	0.436 6	0.768 8	0.695 1	5
	De-trend + CARS + PLS	0.351 4	0.392 2	0.797 2	0.759 8	4

2.3.3.2 小样本建模潜力得分(S_{Model})分析 在应用层面,设备 A 凭借最优预测精度与最高拟合度,在该项测试中占据绝对优势。由于评分体系采用了相对归一化法,设备 E、I 在与 A 设备的横向对比中,受限于较大的预测偏差,均处于较低水平。这表明硬件底层的精密度优势已直接转化为应用层面的预测可靠性。

2.3.3.3 综合选型决策(S_{Total}) 综合吸光度动态重复性(40%)与小样本建模潜力(60%)两维度加权计算,3 台设备的最终选型得分为:设备 A(92.93) > 设备 I(28.91) > 设备 E(14.83)。结果表明,在本研究针对的桂枝茯苓胶囊软材细粉水分监测场景下,傅里叶变换型光谱仪(设备 A)是性能最优的选型方案;而便携式设备 I 虽然在特定物料上表现优异,但其热稳定性瓶颈限制了其在连续高频监测任务中的可靠性。

表 7 3 台设备的各项性能得分及最终综合评分

Table 7 Performance scores and final composite scores of three instruments

Instrument	$S_{STD}(W_1=0.4)$			$S_{Model}(W_2=0.6)$		S_{Total}
	$S_{mannitol}$	$S_{polyethylene}$	$S_{stearic}$	$f(RMSECV)(w_R=0.7)$	$f(R^2)(w_D=0.3)$	
A	47.00	100	100	100	100	92.93
E	0	0	45.58	3.95	39.39	14.83
I	100	99.01	17.84	0	0	28.91

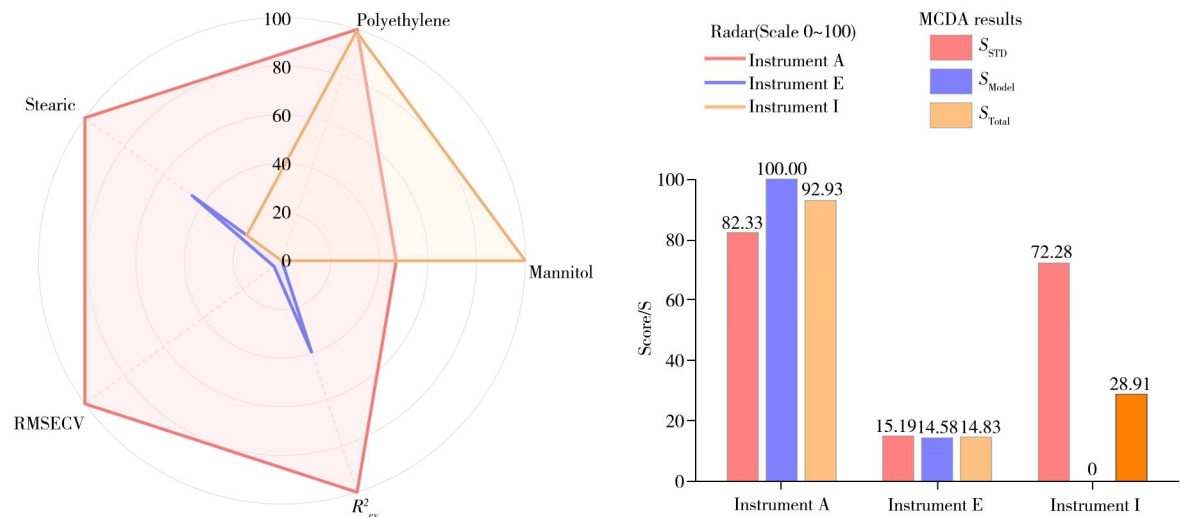


图 8 3 台设备性能特征画像(雷达图)与 MCDA 综合评分(柱状图)对比

Fig. 8 Comparison of performance profiles (radar chart) and multicriteria decision analysis (MCDA) composite scores (bar chart) for three instruments

2.4 补充性评价策略的科学性与创新价值探讨

2.4.1 补充性评价策略的科学逻辑

近红外光谱仪器的吸光度重复性是衡量其性能的关键指标之一,与后续模型构建的准确性密切相关^[34]。本研究认为所使用的评价物料不应只停留在便捷性层面,而应遵循以下通用“4S原则”: (1)理化稳定性(Stability): 在受控环境下性质稳定,确保测试可重复; (2)光谱特征代表性(Spectrum): 具

有宽泛且清晰的近红外吸收特征，其核心官能团特征需覆盖待测成分关键吸收区；(3)光学场景适配性(Scene)：物料的光学特性(如散射、折射)应接近待测样品；(4)动态性能针对性(Sensitivity)：如引入硬脂酸等热敏感物质，可有效监测便携式设备在连续作业下的动态热稳定性，补齐标准品静态检测的短板。

基于上述原则，本研究所选取的非标准物质在光学特性与物化属性上与制药常用物料形成特征映射。甘露醇(高散射粉末模拟)作为经典的结晶型多元醇，其近红外光谱特征主要由O—H键主导，与制药过程分析中所检测的药物水分特征吸收相映射。其粉体状态具有高散射系数，能够模拟检测过程中近红外对晶体以及细粉测量时的情况。聚乙烯(有机化合物模拟)含有大量的C—H二级与一级倍频吸收，这与中药中大量有机基团物质(如多糖类、黄酮类)相匹配。作为致密固体，其漫反射稳定，可作作为标准物质聚苯乙烯的替代品，在 Matthew 等^[35]的研究中，更将其作为设备每日性能验证的低成本物质。硬脂酸(低熔点热敏探针)作为制剂中常见的润滑剂，其熔点 70 °C 的物性特征适合用作动态热敏性探针。标准物质通常对热不敏感，无法识别设备因光源散热缺陷所导致的光谱漂移，硬脂酸物理性质的特点使之能精准识别硬件在高频次采样下的热稳定性边界。

从现有文献、规范来看，SRM2035b、聚苯乙烯薄膜、正庚烷等标准物质是该领域常见的测试载体^[2, 36-37]。但结合采购与溯源信息(表8)分析可知，此类标准物质虽具权威性但存在成本极高、物流周期长、易制毒/易燃安全风险等实操瓶颈。本研究选取的物料，不仅实现了“零采购成本”与高安全性，且其光谱响应特征更契合制药实际场景。

表8 近红外设备吸光度重复性测试所用标准及参考物质的采购与溯源信息

Testing material	Price	Material supplier
聚苯乙烯薄膜	¥900/片 ^[39]	国家标准物质资源库共享平台
SRM2035	¥25 005/个 ^[40]	中亿生物*
正庚烷	¥119/100 mL	国药集团

As of September 2025, NIST SRM 2035b was out of stock, and its price was not available^[38](截至 2025 年 9 月美国国家标准物质公司 SRM2035b 缺少现货，同时也未提供价格^[38])

在设备试用初期(样本量通常 $n<100$)，传统的训练集/验证集划分易导致模型代表性不足。本研究采用留一交叉验证法，最大限度利用了每一个样本的特征信息^[41]，避免了机械划分造成的过拟合风险。该方法在保障计算效率的同时，能快速产出 RMSECV 等量化指标，为企业在缺乏大规模工艺数据支撑的试机阶段提供了一套具备计量学信服力的决策判据。

2.4.2 与传统选型方法的对比及适用边界

设备 I 在硬脂酸测试中暴露的温控隐患，以及设备 A 在小样本建模中展现的性能优势，从硬件画像与应用潜力两个维度，印证了本研究提出的 4S 原则与 MCDA 模型的可靠性。相较于传统方法，本评价策略在用户设备选型试机期间的创新性如表 9 所示。同时本策略引入的多准则综合评价模型具有高度的决策灵活性。公式(6)中的权重 W_1 与 W_2 并非固定参数，用户可根据实际应用场景进行调整。例如，若选型侧重于科研实验室的高精密度分析，可调高 W_1 的比重。这一灵活性确保了评价体系能适配不同的选型优先级。

表9 本策略与传统选型策略对比

Table 9 Comparison between proposed strategy and conventional selection strategy

Dimension	Conventional reference material method	Conventional large-sample modeling method	The proposed supplementary evaluation strategy
评价深度	实验室静态指标	终端应用指标	动态与应用指标相结合
决策成本	高	高	低
风险识别	难以识别散热漂移风险	依赖模型泛化	“热敏探针”量化硬件设计缺陷
适用阶段	年度计量校准、设备出厂检测	生产成熟期模型维护	设备初选、快速选型

3 结 论

本研究针对制药工业近红外设备选型中“试机期短、标准品难求、大样本建模难”的实际痛点，提出并验证了一套“吸光度动态重复性+小样本建模潜力”的双重补充评价策略。通过构建基于“4S

原则”的物料特征映射体系,结合多准则综合评价模型,得出以下结论:

(1)设备性能差异显著:傅里叶变换型设备 A 在物理性能指标与化学建模精度上均表现出绝对优势(RMSECV=0.192 3, R^2_{cv} =0.946 9),最终选型得分(92.93)显著领先,体现了高信噪比硬件对模型稳健性的底层支撑。光栅型设备 E 与 I 虽然在绝对精度上略逊,但通过变量优化筛选,其在水分核心特征区(1 440~1 550 nm)的响应足以满足工业级定量分析需求。

(2)动态评价策略的必要性:吸光度重复性测试成功识别了设备 I 在处理热敏性物质(硬脂酸)时的热稳定性缺陷。这一关键硬件限制是仅依赖供应商提供的静态理论指标或昂贵标准品无法察觉的,充分证明了引入具有物理化学代表性的“非标准物质”进行动态评价的科学性。

(3)小样本策略的可行性:本研究证明,在样本量有限($n=80$)的选型初期,利用留一交叉验证结合预处理与变量筛选,能够最大化表现出设备建模潜力,为快速评估设备的应用适配性提供具备计量学信服力的量化依据。

综上所述,本研究提出的评价策略不仅降低了设备性能验证的经济与时间成本,更通过模拟实际工作负荷识别了硬件隐患。该方法学为制药企业及科研单位在短周期内进行近红外设备选型提供了一套高效、严谨且易于推广的策略参考。

参考文献:

- [1] Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China: Part IV*. 2025 ed. Beijing: China Medical Science Press(国家药典委员会. 中华人民共和国药典: 四部. 2025 年版. 北京: 中国医药科技出版社), 2025.
- [2] T/CIS 17006-2022. General Technical Specification of Fourier Transform Near-infrared Spectrometer. China Instrument and Control Society(傅里叶变换近红外光谱仪通用技术规范. 中国仪器仪表学会).
- [3] Czaja T P, Selga L, Andersson R, Engelsens S B. *Food Res. Int.*, **2025**, 219: 116966.
- [4] Yue J N, Gao L L, Zhong L, Huang R Q, Yang X Y, Tian W L, Cao G Y, Meng Z Q, Nie L, Zang H C. *Microchem. J.*, **2024**, 204: 111130.
- [5] Zhang J, Guo Z, Wang S H, Yue M H, Zhang S S, Peng H H, Yin X, Du J, Ma C Y. *Spectrosc. Spectral Anal.* (张静, 郭榛, 王思花, 岳明慧, 张姗姗, 彭慧慧, 印祥, 杜娟, 马成业. 光谱学与光谱分析), **2023**, 43(7): 2059-2066.
- [6] Wang D, Feng H Z, Li L, Han P. *Spectrosc. Spectral Anal.* (王冬, 冯海智, 李龙, 韩平. 光谱学与光谱分析), **2023**, 43(5): 1351-1357.
- [7] Chen X Y, Siesler H W, Yan H. *Spectrochim. Acta A*, **2021**, 252: 119504.
- [8] Barthès B G, Kouakoua E, Clairotte M, Lallemand J, Chapuis-Lardy L, Rabenarivo M, Roussel S. *Geoderma*, **2019**, 338: 422-429.
- [9] Liu J X, Yin X H, Han S H, Li X, Xu B C, Li P Y, Luo D L. *J. Food Saf. Qual.* (刘建学, 尹晓慧, 韩四海, 李璇, 徐宝成, 李佩艳, 罗登林. 食品安全质量检测学报), **2019**, 10(19): 6432-6437.
- [10] Rubini M, Feuillerat L, Cabaret T, Leroyer L, Leneveu L, Charrier B. *Talanta*, **2021**, 221: 121454.
- [11] Yan H, Siesler H W. *J. Pharm. Biomed. Anal.*, **2018**, 160: 179-186.
- [12] Yifan Z, Ying Z, Fei S, Yue S, Shumei W, Jiang M. *Spectrosc. Lett.*, **2022**, 55(8): 514-526.
- [13] Yang X Q, Luo C, Zhang X X, Wang X L, Yang Z L, Han L J. *Biomass Convers. Biorefin.*, **2026**, 16(3): 108-122.
- [14] Marta P R, Ana D-V, Víctor A, María J A-C. *Agronomy*, **2024**, 14(12): 3061.
- [15] Sun Y, Yuan M, Liu X Y, Su M, Wang L L, Zeng Y Z, Zang H C, Nei L. *Microchem. J.*, **2020**, 159: 105492.
- [16] Cristina M, Emanuel J N M, Sergio T D F, Maria F P, Celio P, Ernestina C. *Talanta*, **2017**, 165: 112-116.
- [17] Ayako T, Elvis A F, Rafael M, Tharanga K, Maggy T S-L. *Insects*, **2025**, 16(11): 1143.
- [18] Tu Z Y, Joseph I, Youngsoo L. *Foods*, **2023**, 12(3): 467-478.
- [19] Yun X, Royston G. *J. Anal. Test.*, **2018**, 2: 249-262.
- [20] Kovács D, Király P, Tóth G. *SAR QSAR Environ. Res.*, **2021**, 32(4): 247-268.
- [21] Andrius V, Emma G, Ellen P, Alexander J C. *Plos One*, **2019**, 14(11): e0224365.
- [22] Chu X L. *Chemometric Methods in Modern Spectral Analysis*. 1st ed. Beijing: Chemical Industry Press(褚小立. 现代光谱分析中的化学计量学方法. 1 版. 北京: 化学工业出版社), **2022**.
- [23] Li S F, Shan Y, Fan W, Yin Y, Zhou Z, Li G Y. *Food Sci.* (李水芳, 单杨, 范伟, 尹永, 周孜, 李高阳. 食品科学), **2011**, 32: 182-185.
- [24] Xu F, Fu D D, Wang Q H, Xiao Z, Wang B. *Food Sci.* (许锋, 付丹丹, 王巧华, 肖壮, 王彬. 食品科学), **2018**, 39: 149-154.
- [25] Liu Z C, Cai W S, Shao X G. *Sci. China B*, **2008**, 51(8): 751-759.

- [26] Diwu P Y, Bian X H, Wang Z F, Liu W. *Spectrosc. Spectral Anal.* (第五鹏瑶, 卞希慧, 王姿方, 刘巍. 光谱学与光谱分析), **2019**, 39(9): 2800–2806.
- [27] Åsmund R, Frans V D B, Søren B E. *TrAC, Trends Anal. Chem.*, **2009**, 28(10): 1201–1222.
- [28] Yun Y H, Li H D, Deng B C, Cao D S. *TrAC, Trends Anal. Chem.*, **2019**, 113: 102–115.
- [29] Åsmund R, Martin A, Carsten R, Søren B E. *J. Chemom.*, **2014**, 28(5): 439–447.
- [30] Yifan C, Mengsheng Z, Chen N, Harse S, Lianbo G. *Spectrochim. Acta A*, **2026**, 351: 127445.
- [31] Deng B C, Yun Y H, Liang Y Z, Cao D S, Xu Q S, Li L Z, Huang X. *Anal. Chim. Acta*, **2015**, 880: 32–41.
- [32] Li Y Q, Hong S J, Huang W, Zhang L G, Ge J, Luan S R, Ni L J. *J. Instrum. Anal.* (李永琪, 洪士军, 黄雯, 张立国, 葛炯, 栾绍嵘, 倪力军. 分析测试学报), **2020**, 39: 1231–1238.
- [33] Sophia M, Krzysztof B B, Justyna G, Elisabeth S, Christian W H. *Talanta*, **2021**, 223: 121809.
- [34] Ke P Y, Liu M X, Wang X Q, Huang S L, Zhang Y G, Fang J X. *J. Infrared Millim. Waves* (柯鹏瑜, 刘梦璇, 王绪泉, 黄松垒, 张永刚, 方家熊. 红外与毫米波学报), **2021**, 40(5): 582–588.
- [35] Matthew E, Michael P, Chyanee C, Mohammed J, Steve S, David J. *Spectrochim. Acta A*, **2022**, 267: 120512.
- [36] Lu W Z, Yuan H F, Chu X L. *Near-infrared Spectroscopic Instruments*. 1st ed. Beijing: Chemical Industry Press (陆婉珍, 袁洪福, 褚小立. 近红外光谱仪器. 第1版. 北京: 化学工业出版社), **2010**: 110–111.
- [37] Lu W Z. *Modern Near-infrared Spectroscopic Analytical Technology*. 2nd ed. Beijing: China Petrochemical Press (陆婉珍. 现代近红外光谱分析技术. 第2版. 北京: 中国石化出版社), **2007**: 134–135.
- [38] National Institute of Standards and Technology (NIST). NIST Standard Reference Material 2035b[EB/OL]. https://shop.nist.gov/ccrz__ProductDetails?sku=2035b&cclcl=en_US.
- [39] National Standard Material Sharing Platform. GBW(E)130414 Polystyrene Infrared Wavelength Reference Material (国家标准物质资源共享平台. GBW(E)130414 聚苯乙烯红外波长标准物质). [2025–10–15]. <https://www.ncrm.org.cn/Web/Ordering/MaterialDetail?autoID=6226>.
- [40] Dongguan Zhongyi Biotechnology Co., Ltd. Standard Reference Material NIST SRM 2035b – Ultraviolet–Visible–Near–infrared Wavelength/Wavenumber Transmission Standard (东莞市中亿生物科技有限公司. 标准参考物质 NIST SRM 2035b – 紫外可见光近红外波长/波数传输标准). [2025–10–15]. <http://www.srmcas.com/cpxq/116379.html>.
- [41] Hu J Y, Xu X, Ji X M, Xu M X, Jiang D F, Wang J. *Chin. J. Eng.* (胡静怡, 徐翔, 季小妹, 徐明贤, 姜岱峰, 王登. 工程科学学报), **2023**, 45(9): 1517–1527.

(责任编辑: 盛文彦)