

模型转移技术在小儿消积止咳口服液提取过程中的应用研究

许秀华¹, 聂磊¹, 马晓博¹, 庄晓琪¹, 张晋¹, 董海玲¹, 梁文燕¹, 杜昊忱²,
袁晓梅², 关永霞², 李连¹, 张惠¹, 郭学平^{1,3*}, 臧恒昌^{1*}

(1. 山东大学药学院, 国家药品监督管理局药物制剂技术与评价重点实验室, 山东 济南 250012;
2. 鲁南制药股份有限公司, 中药制药共性技术国家重点实验室, 山东 临沂 276006; 3. 华熙生物科技股份
有限公司, 山东 济南 250012)

摘要: 中药产业的现代化发展对中药产品质量提出了更高的要求, 提取过程作为中药生产的关键基础环节, 直接关系到产品的最终质量。然而, 目前提取工艺的质量保证主要依靠简单的湿化学分析, 无法达到提取过程的精准控制, 而且存在费时、耗力等诸多弊端。因此, 将近红外光谱技术应用于中药生产环节的过程监控对于中药产品质量控制具有重要意义。本研究以小儿消积止咳口服液的提取过程为研究对象, 利用傅里叶变换近红外光谱仪和便携式近红外光谱仪收集近红外光谱, 监测提取过程中药典规定的质控指标成分辛弗林的浓度变化, 实现提取过程的快速、准确检测。同时创新模型转移方法, 实现两种不同类别(分析级与便携式)的近红外光谱仪间的模型转移, 解决了便携式仪器分辨率低、性能差、预测精度不足的问题, 以实现小儿消积止咳口服液进行现场快速筛查和质量分析, 对中药生产过程质量监控具有指导意义。

关键词: 近红外光谱; 模型转移技术; 小儿消积止咳口服液; 提取过程; 辛弗林

中图分类号: R917 文献标识码: A 文章编号: 0513-4870(2023)10-2900-09

Study on the application of model transfer technology in the extraction process of Xiao'er Xiaoji Zhike oral liquid

XU Xiu-hua¹, NIE Lei¹, MA Xiao-bo¹, ZHUANG Xiao-qi¹, ZHANG Jin¹, DONG Hai-ling¹,
LIANG Wen-yan¹, DU Hao-chen², YUAN Xiao-mei², GUAN Yong-xia², LI Lian¹, ZHANG Hui¹,
GUO Xue-ping^{1,3*}, ZANG Heng-chang^{1*}

(1. NMPA Key Laboratory for Technology Research and Evaluation of Drug Products, School of Pharmaceutical Sciences, Shandong University, Ji'nan 250012, China; 2. State Key Laboratory of Generic Manufacture Technology of Chinese Traditional Medicine, Lunan Pharmaceutical Group Co. Ltd., Linyi 276006, China; 3. Bloomage Biotechnology Co. Ltd., Ji'nan 250012, China)

Abstract: The modernization and development of traditional Chinese medicine has led to higher standards for the quality of traditional Chinese medicine products. The extraction process is a crucial component of traditional Chinese medicine production, and it directly impacts the final quality of the product. However, the currently relied upon methods for quality assurance of the extraction process, such as simple wet chemical analysis, have several limitations, including time consumption and labor intensity, and do not offer precise control of the extraction process. As a result, there is significant value in incorporating near-infrared spectroscopy (NIRS) in the production

收稿日期: 2023-03-01; 修回日期: 2023-04-26.

基金项目: 山东省重点研发计划项目 (2021CXGC010507); 山东省重点研发计划资助 (2022CXGC020515); 国家重点研发计划 (2019YFC1711200); 山东省重大关键技术攻关 (重大科技创新工程) 项目 (2021CXGC010511).

*通讯作者 Tel: 86-531-88380268, E-mail: guoxp@bloomagefreda.com; zanghcw@126.com

DOI: 10.16438/j.0513-4870.2023-0239

process of traditional Chinese medicine to improve the quality control of the final products. In this study, we focused on the extraction process of Xiao'er Xiaoji Zhike oral liquid (XXZOL), using near-infrared spectra collected by both a Fourier transform near-infrared spectrometer and a portable near-infrared spectrometer. We used the concentration of synephrine, a quality control index component specified by the pharmacopoeia, to achieve rapid and accurate detection in the extraction process. Moreover, we developed a model transfer method to facilitate the transfer of models between the two types of near-infrared spectrometers (analytical grade and portable), thus resolving the low resolution, poor performance, and insufficient prediction accuracy issues of portable instruments. Our findings enable the rapid screening and quality analysis of XXZOL onsite, which is significant for quality monitoring during the traditional Chinese medicine production process.

Key words: near-infrared spectroscopy; model transfer technique; Xiao'er Xiaoji Zhike oral liquid; extraction process; synephrine

提取过程是中药制备过程中必不可少的一步,是中药生产的关键基础环节,直接关系到产品的最终质量。然而,目前提取工艺的质量保证主要依靠提取结束时的湿化学分析,甚至只依靠工人的经验知识进行判断,客观依据和分析不足。已有研究报道采用高效液相色谱法^[1]、紫外分光光度法^[2]和薄层色谱法^[3]等分析方法用于提取物的质量控制,然而这些方法存在样本前处理复杂、环境不友好和费时费力等缺点^[4]。在中药制药工业中,提取过程无法达到精准控制,且离线分析存在诸多弊端:分析时间长、样本氧化变质等,这些问题都可能导致提取过程控制依据不充分,控制精度不准,进而使产品质量不稳定。

过程分析技术 (process analysis technology, PAT) 是在 2004 年提出的新技术,旨在帮助人们了解药物制备过程中的变化。光谱技术是过程分析领域强大的分析工具之一,其中,近红外 (near-infrared, NIR) 光谱是 PAT 的常用工具^[5],在中药生产过程监测中得到了广泛应用^[6],如原药材、中间体和成品的定性鉴定和定量测定,并且也应用于提取工艺、浓缩工艺、醇沉淀工艺和共混工艺的监测^[7]。中药提取过程是一个复杂的动态系统,具有许多变量和干扰。因此,过程监控对于中药产品质量控制具有重要意义^[8]。近红外光谱技术的应用依赖于光谱仪器。大型光谱仪器精度高、分析准确,但受限于体积和环境要求,较难在生产过程中应用,虽然当前也有能集成于提取罐上的近红外检测装备,对提取过程进行实时监测,但所集成的高精度近红外光谱仪的应用成本较高。便携式小型光谱仪体积小、重量轻、环境适应性好,适合现场在线监测和分析,但其光谱存在分辨率低、精度不足等问题,导致模型预测结果准确度不高^[9]。为解决上述问题,模型转移技术应运而生。

模型转移技术通过数学方法建立源仪器和目标仪器所测光谱间的转移函数,通过转移函数对光谱进行

校正来提高模型适用性^[4,10]。如果能够在保持近红外光谱模型预测能力不显著降低的前提下,提高模型预测能力的同时达到缩小仪器体积及降低仪器成本的目的,将对近红外技术在中药生产过程中的推广和应用产生重要意义。

本研究以小儿消积止咳口服液在中试规模中的提取过程为研究对象,利用傅里叶变换近红外光谱仪 (高精度) 和便携式近红外光谱仪 (低精度) 收集 NIR 光谱,监测该药提取过程中药典规定的质量指标成分辛弗林的浓度变化,实现提取过程的快速、准确检测,同时创新模型转移方法,对比传统直接标准化 (direct standardization, DS) 和改进的主成分分析 (improved principal component analysis, IPCA) 方法^[11],实现两种不同类别 (分析级与便携式) 的近红外光谱仪间的模型转移,不仅能改善便携式近红外光谱仪的模型预测能力,而且也能对小儿消积止咳口服液进行现场快速监测和质量分析,在一定程度上降低近红外光谱仪的应用成本,同时对中药实际生产中在线实时监测模型的质量提升具有指导意义。

材料与amp;方法

药材与试剂 炒山楂、槟榔 (块)、枳实 (粉)、蜜枇杷叶、瓜蒌 (块)、炒莱菔子、桔梗 (段)、连翘 (净)、蝉蜕 (净)、炒葶苈子共 10 味药材由中国山东鲁南厚普制药有限公司提供,符合 2020 年版《中华人民共和国药典》一部相关项下要求。

色谱纯乙腈、甲醇 (德国 Merck 公司); 对照品辛弗林 (批号: M27GB149635, 含量 $\geq 98\%$, 上海源叶生物科技有限公司); 分析纯高氯酸 (旭化成化学公司); 超纯水 (自制)。

仪器 Acquity Arc 型高效液相色谱仪 (PDA 型紫外检测器, Empower 色谱管理软件, 美国 Waters 公司); XS105DU 型电子分析天平 (Mettler Toledo 公司);

JY-15 超声波清洗机 [鼎泰 (湖北) 生化科技设备制造有限公司]; 津腾 GM-0.5B 型隔膜真空泵; 离心机 (Thermo Fisher Scientific 公司); Milli-Q 超纯水仪 (Merck Millipore 公司); MATLAB 2021b (美国 MathWorks 公司)。

参考值的测定

样本制备 按照 2020 版《中华人民共和国药典》一部中小儿消积止咳口服液制剂处方 20 倍量分别称取所需 10 味药材投料至 400 L 的提取罐, 模拟小儿消积止咳口服液实际生产过程提取液的制备方法, 10 味药材, 加水煎煮两次。待药液沸腾后, 调小蒸汽, 使料液保持微沸状态并开始计时, 第一次提取 2.5 h, 第二次提取 2 h。提取过程利用自制循环设备间隔时间取样, 自蒸汽阀开启时开始取样, 第一次提取沸腾前间隔 2 min, 沸腾后间隔 5 min; 第二次提取沸腾前间隔 2 min, 沸腾后间隔 10 min, 每次取样 15 mL。当日不能测完的样本, 放置 -20°C 冰箱冷冻保存。共平行进行了 4 个批次的中试提取实验。

供试品溶液制备 精密吸取上述提取液, 于 $12\,000\text{ r}\cdot\text{min}^{-1}$ 离心 5 min, 取上清液, 用 $0.22\ \mu\text{m}$ 微孔滤膜过滤, 即得供试品溶液。

对照品溶液制备 精密称取对照品辛弗林约 5.00 mg, 置于 10 mL 量瓶中, 加入 50% 甲醇, 超声溶解, 得到浓度为 $0.50\text{ mg}\cdot\text{mL}^{-1}$ 的对照品储备液。精密移取对照品储备液 1 mL 置 10 mL 量瓶中, 加超纯水稀释至刻度, 摇匀, 滤过, 即得对照品溶液。

色谱条件 色谱柱: Waters Xselect[®] HSS T3 色谱柱 ($4.6\text{ mm}\times 150\text{ mm}$, $2.5\ \mu\text{m}$); 流动相: 0.1% 高氯酸水溶液 (A) 和乙腈: 甲醇 (1:1) (B); 柱温: 30°C ; 进样量: $5\ \mu\text{L}$; 流速: $0.6\text{ mL}\cdot\text{min}^{-1}$; 检测波长: 225 nm。洗脱梯度程序如下: 0~25 min, 95%~80% A; 25~45 min, 80%~70% A; 45~55 min, 70%~50% A; 55~60 min, 50%~40% A; 60~61 min, 40%~95% A; 61~70 min, 95% A。

光谱采集

傅里叶变换近红外 (FT-NIR) 光谱 采用傅里叶变换近红外光谱仪 (MPA II, Bruker, 德国) 获取 NIR 光谱。将样本装入路径长度为 1.0 mm 的石英比色皿池中, 分辨率为 8 cm^{-1} , 扫描次数为 32 次。光谱范围为 $12\,000\sim 4\,000\text{ cm}^{-1}$, 每张光谱 2 801 个数据点, 采集每个样本光谱前以空气为参考采集背景光谱。每个样本采集 3 次, 平均光谱用于建模。FT-NIR 光谱作为源仪器光谱, 仪器外形见图 1A。

Micro-NIR 光谱 采用装有 InGaAs 探测器的 Micro-NIR 1700 光谱仪 (Viavi solutions, 中国), 收集漫

反射光谱, 环境条件与 FT-NIR 光谱一致。每批次采集光谱之前, 都设置自动积分, 波长范围 $908\sim 1\,676\text{ nm}$, 每张光谱 125 个数据点, 每个样本采集 3 次, 平均光谱用于建模。Micro-NIR 光谱作为目标仪器光谱, 仪器外形见图 1B。

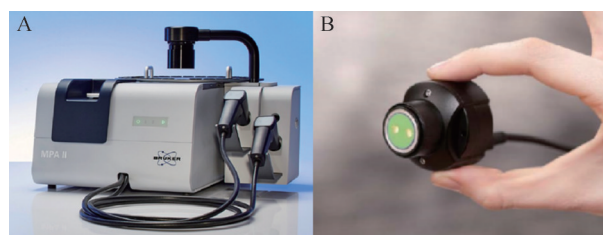


Figure 1 A: Fourier transform near-infrared (FT-NIR) spectrometer. B: Micro-NIR 1700

建模方法与模型转移方法 首先采用主成分与马氏距离 (PCA-DA) 结合的方法剔除两台光谱仪中共有的异常光谱对应的样本, 选择两台光谱仪光谱波长的交集部分即波长范围在 $908\sim 1\,676\text{ nm}$ 的光谱数据, 然后使用 Kennard and Stone (KS) 法进行校正集和验证集的划分, 目标仪器校正集及验证集样本与源仪器相同, 以公平比较不同模型转移方法的模型转移效果。随后采用多元散射校正 (multiple scatter correction, MSC), 标准正态变量变换 (standard normal variate transformation, SNV) 和 Savitzky-Golay (S-G) 卷积求导及其组合的预处理方法进行光谱预处理, 以消除非目标因素对模型转移结果的影响。在确定预处理方法相关参数时, 对源仪器光谱和目标仪器光谱数据分别建立偏最小二乘回归 (partial least squares regression, PLSR) 模型, 根据各自模型质量确定最终预处理方法。并采用“留一法”, 以交叉验证均方根误差 (root mean squares error of cross validation, RMSECV) 确定校正模型的潜在因子数。

然后进行两台仪器光谱间的模型转移, 本文对比了传统 DS 和 IPCA 方法^[12]。

DS 法本质是通过求 Moore-Penrose 伪逆运算构建转移矩阵 F , 如公式 (1) 表示:

$$X_{t(\text{trans})} = X_{m(\text{trans})} F \quad (1)$$

式中, trans 表示转移集样本, X_t 和 X_m 分别表示目标仪器和源仪器光谱矩阵。 F 矩阵通过求 Moore-Penrose 伪逆得到, 如公式 (2) 所示:

$$F = X_{m(\text{trans})}^+ X_{t(\text{trans})} \quad (2)$$

式中, 上标“+”表示求 Moore-Penrose 伪逆运算。通过构建 F 矩阵后, 可将源仪器的校正集样本光谱 $X_{m(\text{Cal})}$ 代入公式 (1), 计算转移后适合目标仪器的校正集光谱, 进而建立转移后光谱 $X_{t(\text{Cal})}$ 的目标仪器校正

集, 如公式 (3) 所示:

$$X_{t(\text{Cal})} = X_{m(\text{Cal})} F \quad (3)$$

将目标仪器测定的验证集光谱 $X_{t(\text{Val})}$ 代入此校正模型进行预测。

IPCA 方法是将待转移的目标仪器光谱矩阵进行主成分分解, 得到主成分得分矩阵 (T) 和负载矩阵 (P), 然后用源仪器的光谱与目标仪器的得分矩阵建立转换矩阵, 如公式 (4) 和 (5) 所示:

$$X_{t(\text{trans})} = T_{t(\text{trans})} P_{t(\text{trans})}^T \quad (4)$$

$$T_{t(\text{trans})} = X_{m(\text{trans})} \beta \quad (5)$$

其中, $T_{t(\text{trans})}$ 和 $P_{t(\text{trans})}$ 分别表示目标仪器转移集光谱主成分分解后的得分矩阵和负载矩阵, 上标 T 表示矩阵转置。 $T_{t(\text{trans})}$ 和 $P_{t(\text{trans})}$ 可通过优化主成分数来确定。通过公式 (5) 将源仪器的转移集光谱与目标仪器转移集光谱的得分矩阵关联, 系数矩阵 β 可通过 PLS 求解。将公式 (5) 代入公式 (4) 可得到公式 (6):

$$X_{t(\text{trans})} = X_{m(\text{trans})} \beta P_{t(\text{trans})}^T \quad (6)$$

对比公式 (1) 和 (6), 可知 F 矩阵可表示为 $\beta P_{t(\text{trans})}^T$ 。在 IPCA 方法中, 可选择不同的主成分数来优化模型转移效果。当选择较大的主成分数时, DS 方法中的 F 矩阵就与 $\beta P_{t(\text{trans})}^T$ 结果趋于一致, 因此 DS 法与 IPCA 法的模型转移结果近乎一致。将 $X_{m(\text{Cal})}$ 代入

公式 (6) 计算得到适合目标仪器校正集样本的光谱, 见公式 (7):

$$X_{t(\text{Cal})} = X_{m(\text{Cal})} \beta P_{t(\text{trans})}^T \quad (7)$$

本研究进行了两种模型转移方法由源仪器向目标仪器转移和由目标仪器向源仪器转移的双向转移研究, 转移过程如图 2 所示。在模型转移的过程中, 以 R_p 和 RMSEP 值作为模型转移效果的评价指标。

结果与讨论

1 参考值测定结果

通过参考方法测定的样本参考值结果如表 1 所示。

Table 1 Results of determination of synephrine in four batches of extracts

Marker component	Batch	Min value/ mg·mL ⁻¹	Max value/ mg·mL ⁻¹	Average value/ mg·mL ⁻¹	Standard deviation
Synephrine	1	0.014 8	0.100 5	0.052 5	0.033 0
	2	0.011 3	0.103 0	0.053 9	0.033 1
	3	0.012 1	0.104 3	0.052 7	0.033 6
	4	0.012 7	0.108 8	0.060 6	0.036 0

2 原始光谱分析

图 3A、B 分别为两台仪器测定的 4 个批次 (共 182 个样本) 的原始 NIR 吸收光谱图。源仪器和目标

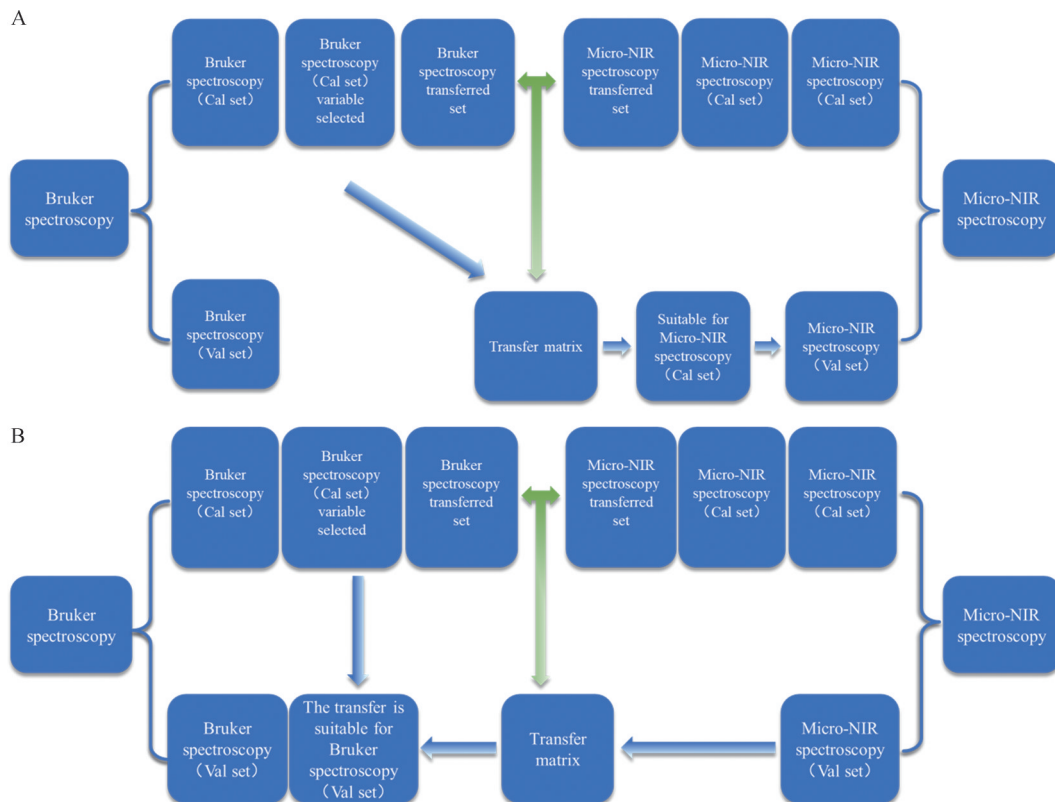


Figure 2 Model transfer process. A: Transfer from source instrument to target instrument; B: Transfer from target instrument to source instrument

仪器测定的光谱波数或波长范围和光谱的强度不同。在图3中,观察到 $12\ 000\sim 9\ 000\ \text{cm}^{-1}$ 没有明显吸收, $5\ 260\sim 5\ 060\ \text{cm}^{-1}$ 的光谱区域过饱和,因此这两部分光谱被认为是本研究的无效信息。吸收带主要集中在 $6\ 900\ \text{cm}^{-1}$ (O-H的第一泛频和C-H的组合频)和 $5\ 200\ \text{cm}^{-1}$ (O-H的组合频)左右,这是水溶液的NIR光谱的典型特征。此外, $8\ 700$ 和 $5\ 600\ \text{cm}^{-1}$ 的峰表示C-H基团的吸收,前者表示C-H基团第二泛频和其组合频,后者表示第一泛频。由于使用水作为提取溶剂,样本中其他化学组分的光谱信号几乎被覆盖,因此需进一步使用化学计量学方法来提取挖掘有效的光谱信息。

3 异常光谱的检测与样本集划分

受到样本本身和测定环境的影响,样本的近红外光谱数据可能存在异常。异常值的存在会一定程度影响模型的训练,进一步影响模型质量,因此需对异常光谱进行检测、剔除。使用主成分分析(PCA)与马氏距离相结合,分别对两个仪器采集到的182个样本的光谱数据进行分析,基于Hotelling T^2 检验在 $\alpha = 0.05$ 置信水平上计算关键值,异常光谱检测结果如图4所示,源仪器的4个样本(图4A)光谱位于椭圆所示的95%置信限外,目标仪器的21个样本(图4B)光谱位于椭圆所示的95%置信限外,其中图3B中所示吸收强度较低的光谱样本包括在异常样本中(即46、47号样本),因此共剔除25个异常样本,剩余157个样本用于建模研究。采用K-S方法,按校正集与验证集比例为3:1进行划分,得到校正集样本118个,验证集样本39个。

4 预处理方法选择

本研究中,源仪器和目标仪器的光谱测定范围及变量单位不同,源仪器光谱变量单位为波数,采集样本的光谱范围为 $4\ 000\sim 12\ 000\ \text{cm}^{-1}$,而目标仪器的光谱变量单位是波长,采集的样本光谱范围为 $908\sim 1\ 676\ \text{nm}$ 。为了便于比较,将光谱变量的单位统一转换为波长,即源仪器光谱变量范围为波长 $833\sim 2\ 500\ \text{nm}$ 。在进行预处理前,首先选取共同波长,即 $908\sim 1\ 676\ \text{nm}$,图5为相同波长范围内的光谱对比图(采用统一坐标表示)。从图中可看出,尽管两台仪器所采集光谱在波长范围、波长(即光谱变量)点数、吸光度值存在较大差异,但同一波长范围内光谱的轮廓基本相似,波长点数仍不同。

样本的近红外光谱中除了样本自身含氢基团振动吸收外,还包含有噪声等干扰信息。因此,在建立校正模型前需对光谱数据进行预处理,以获得稳定、可靠的光谱校正模型。针对共有波长光谱采用MSC、SNV和S-G卷积求导及其组合方法对原始光谱进行预处理,使用预处理后的光谱数据建立辛弗林浓度的定量分析模型,以较高的 R_p 值和较低的RMSEP值为依据分别选择源仪器和目标仪器的最佳预处理方法,模型结果见表2。

由表2所示,选用SNV和一阶导数,25点平滑作为源仪器光谱数据定量模型的预处理方法,选用MSC和一阶导数,25点平滑作为目标仪器光谱数据定量模型的预处理方法,所建模型的效果最佳。对于共有波长范围光谱,源仪器所建模型的效果更好(较高的 R_p

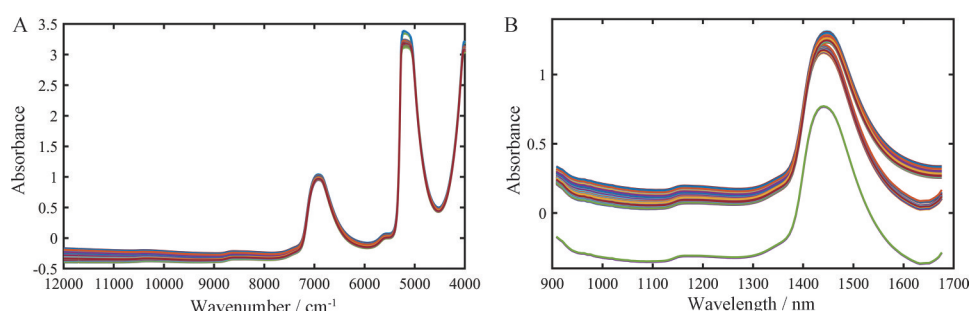


Figure 3 A: Source instrument original spectra; B: Target instrument original spectra

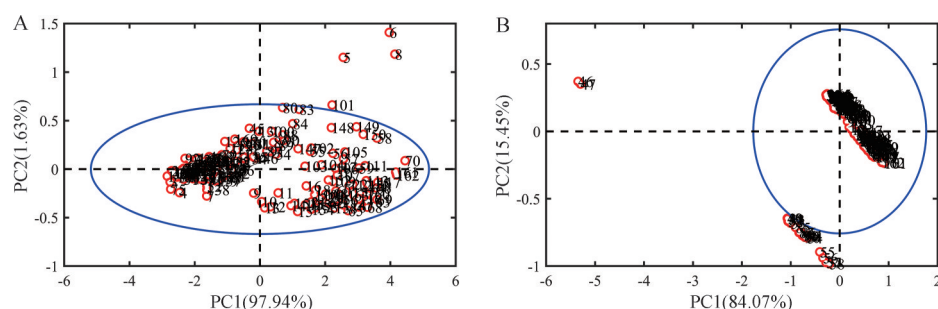


Figure 4 Score chart of spectral principal component analysis of sample. A: Source instrument data; B: Target instrument data

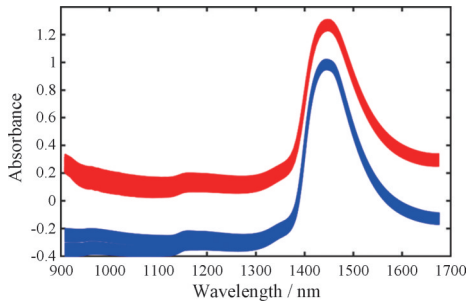


Figure 5 Near infrared spectra of the same wavelength range of the two instruments. Blue: Source instrument common band data; Red: Target instrument data

值, 较低的RMSEP值)。由于高精度的源仪器不便于在中药生产环境中直接使用, 而便携式光谱仪体积小, 环境适应性好, 适合于在中药生产过程中使用, 因此可将高精度光谱仪模型转移到便携式光谱仪上, 提升其模型性能和质量, 无须更多硬件投入, 不仅节省生产成本, 也便于模型的共享和推广应用。

5 模型转移结果与比较

5.1 模型由源仪器向目标仪器方向转移 采用KS法

从校正集样本中选择一定数量的样本作为转移集样本优化模型转移效果。以 R_p 和RMSEP值为评价指标, 当转移集样本数为78时, 模型转移效果最好。基于转移集样本采用DS或IPCA方法建立转换矩阵, 通过转换矩阵将源仪器校正集样本的光谱校正为适合目标仪器校正集样本的光谱。基于适合目标仪器校正集样本的光谱重新建立校正模型, 将目标仪器测定的验证集光谱代入此校正模型进行预测, 以实现模型由源仪器向目标仪器方向转移。模型转移结果总结于表3和图6。

由图6可知, 光谱预测值与真实值均匀分布在直线及直线的两侧, 表明提取液中辛弗林的含量参考值与经近红外光谱校正模型计算的预测值大致相同, 线性关系较强。经表3和图6结果对比可知, 主仪器的建模结果最优, 相较于目标仪器的原始模型, 经DS与IPCA方法将性能较好的源仪器模型转移为适合目标仪器的模型后, 目标仪器在保持住原有模型的预测能力的基础上, RMSEP值有所降低(从0.0067到0.0063), 说明有一定的提升。这可能是因为源仪器的光谱分辨

Table 2 Comparison of the results of different pretreatment methods in the quantitative analysis model of synephrine concentration. 1st: The first derivative; 2nd: The second derivative. LVs: Number of latent variables; R_c : Decision coefficient of calibration; RMSEC: Root mean squares error of calibration; R_p : Determination coefficient of predication; RMSEP: Root mean squares error of validation; MSC: Multiplicative scatter correction; SNV: Standard normal variate transformation; S-G: Savitzky-Golay

Spectrometer	Pretreatment	LVs	R_c	RMSEC	R_p	RMSEP
Source instrument	MSC	10	0.994 0	0.003 8	0.992 3	0.003 5
	SNV	10	0.991 8	0.004 5	0.991 6	0.004 2
	MSC+S-G-1 st -(25,2)	9	0.995 0	0.003 5	0.992 4	0.003 4
	MSC+S-G-2 nd -(25,2)	10	0.997 8	0.002 3	0.984 6	0.005 2
	MSC+S-G-1 st -(35,2)	10	0.996 1	0.003 1	0.991 8	0.003 6
	MSC+S-G-2 nd -(35,2)	8	0.994 9	0.003 5	0.990 7	0.004 1
	SNV+S-G-1 st -(25,2)	8	0.994 3	0.003 7	0.994 0	0.003 1
	SNV+S-G-2 nd -(25,2)	7	0.993 6	0.004 0	0.983 6	0.005 2
	SNV+S-G-1 st -(35,2)	10	0.995 1	0.003 5	0.991 8	0.003 6
Target instrument	SNV+S-G-2 nd -(35,2)	8	0.994 9	0.003 5	0.990 6	0.004 1
	MSC	8	0.981 6	0.006 7	0.972 2	0.007 1
	SNV	8	0.982 0	0.006 6	0.968 5	0.007 5
	MSC+S-G-1 st -(25,2)	8	0.973 1	0.008 1	0.976 8	0.006 7
	MSC+S-G-2 nd -(25,2)	7	0.972 0	0.008 3	0.973 6	0.006 9
	MSC+S-G-1 st -(35,2)	6	0.960 6	0.009 8	0.973 7	0.006 7
	MSC+S-G-2 nd -(35,2)	8	0.973 1	0.008 1	0.976 4	0.006 7
	SNV+S-G-1 st -(25,2)	9	0.976 1	0.007 6	0.976 6	0.006 7
	SNV+S-G-2 nd -(25,2)	5	0.968 1	0.008 8	0.974 0	0.006 9
SNV+S-G-1 st -(35,2)	9	0.973 1	0.008 1	0.968 0	0.007 5	
SNV+S-G-2 nd -(35,2)	8	0.973 2	0.008 1	0.976 4	0.006 7	

Table 3 Results of model transfer from source instrument to target instrument based on the same wavelength range using two methods. DS: Direct standardization; IPCA: Improved principal component analysis

Instrument	Number of transfer set samples	LVs/variable	R_c	RMSEC	R_p	RMSEP
Source instrument	-	8	0.994 3	0.003 7	0.994 0	0.003 1
Target instrument	-	8	0.973 1	0.008 1	0.976 8	0.006 7
Target instrument after DS transfer	78	7	0.971 0	0.008 4	0.980 2	0.006 3
Target instrument after IPCA transfer	78	7	0.971 0	0.008 4	0.980 2	0.006 3

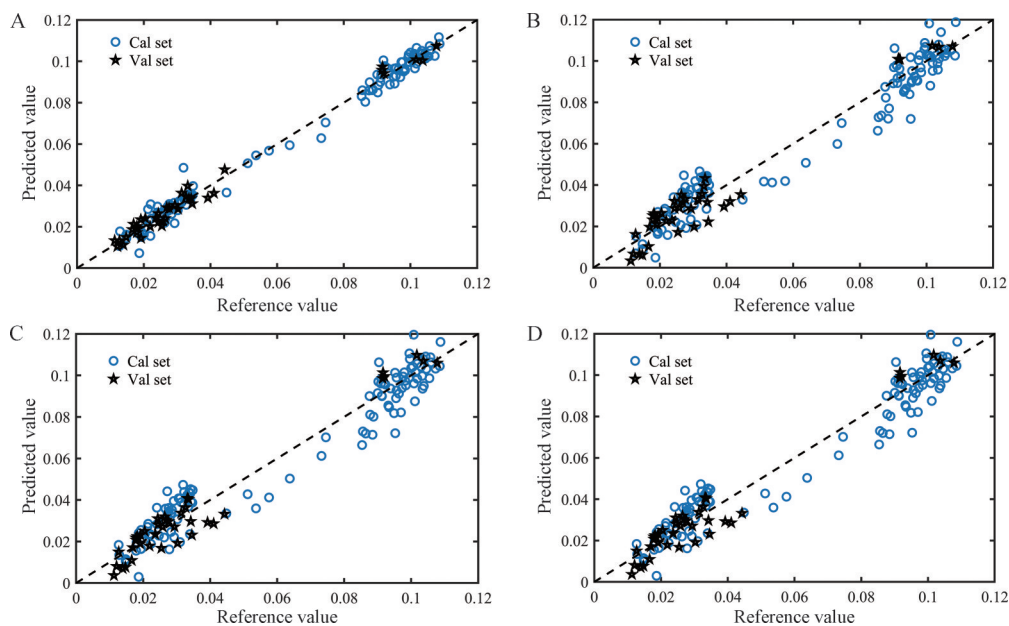


Figure 6 DS method and IPCA method before and after the transfer from source instrument to target instrument model prediction results. A: Source instrument modelling results; B: Target instrument modeling results; C: Modeling results after DS method model transfer; D: Modeling result after IPCA model transfer

率较高, 908~1 676 nm 包含 1 767 个数据点, 光谱信息较为丰富。而目标仪器在此范围内的光谱只包括 125 个数据点, 分辨率相对较低。通过模型转移, 源仪器的光谱信息得以传递到目标仪器中, 提高了模型预测性能。

同时由表 3 和图 6 可见, DS 与 IPCA 方法的结果近乎一致, 这可从两种方法的原理角度来说明。对比模型转移方法介绍中 DS 法和 IPCA 法的基本原理 [公式 (3) 和 (7)] 可知, IPCA 方法可通过调节主成分数及转

移集样本数优化转移效果, 而 DS 方法只能通过调节转移集样本的数量来优化转移效果, 可调节的参数较少, 灵活性较低。图 7 对比了 DS 法与 IPCA 法对校正集光谱的转移结果。可见, 两种转移方法对目标仪器校正集光谱的转移效果近乎一致, 与原始测定的光谱差异很小 (差异量级在 10^{-4})。

5.2 模型由目标仪器向源仪器方向转移 将目标仪器校正集样本的光谱校正为适合源仪器校正集样本的光谱, 以实现模型由目标仪器向源仪器方向转移。在

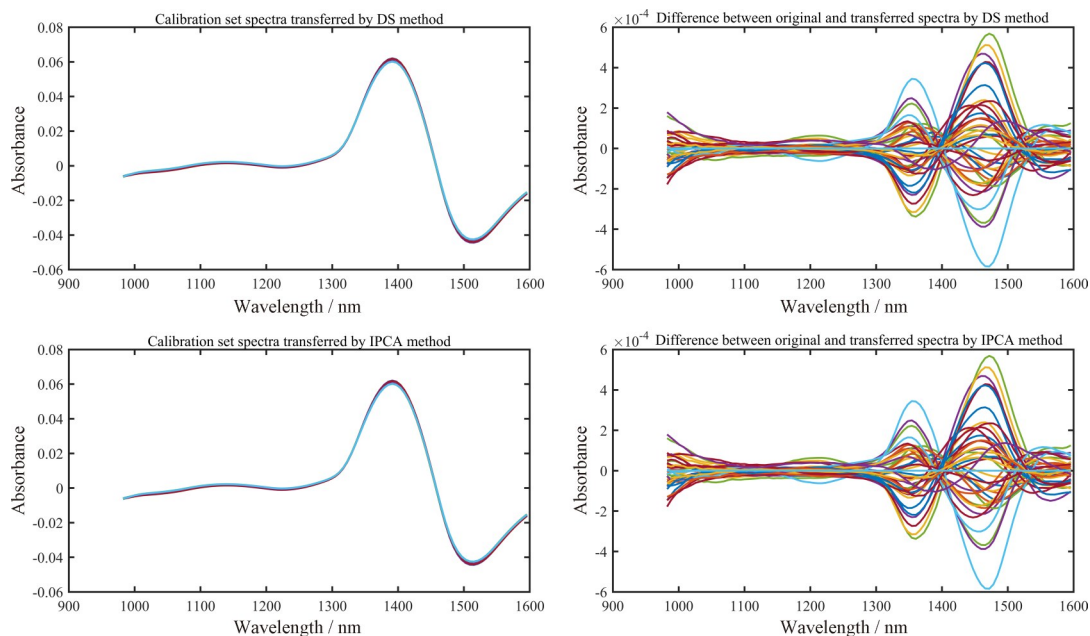


Figure 7 Comparison of transfer spectra between DS method and IPCA method

此方向转移中, 由较低分辨率的目标仪器光谱转移为较高分辨率的源仪器光谱, 按照 DS 与 IPCA 方法原理, 转换矩阵运算是可以实现的。为了预测验证集样本的辛弗林含量, 需要将验证集光谱转换为适合源仪器的光谱, 然后直接代入源仪器的校正模型进行预测即可。

图 8 展示了两种模型转移方法对目标仪器验证集光谱转换为适合源仪器的光谱图。由图可见, 相较于

DS 法, IPCA 法转换后的图谱与源仪器测定的验证集图谱的差异波动相对较小。DS 法与 IPCA 法的模型转移结果见表 4 和图 9。

因转移后的验证集光谱直接使用源仪器的原始校正模型, 因此潜在因子数与源仪器的校正模型是一致的。DS 法是通过 F 矩阵将低分辨率光谱 (目标仪器光谱数据点少) 向高分辨光谱 (源仪器光谱数据点多) 转换, 计算

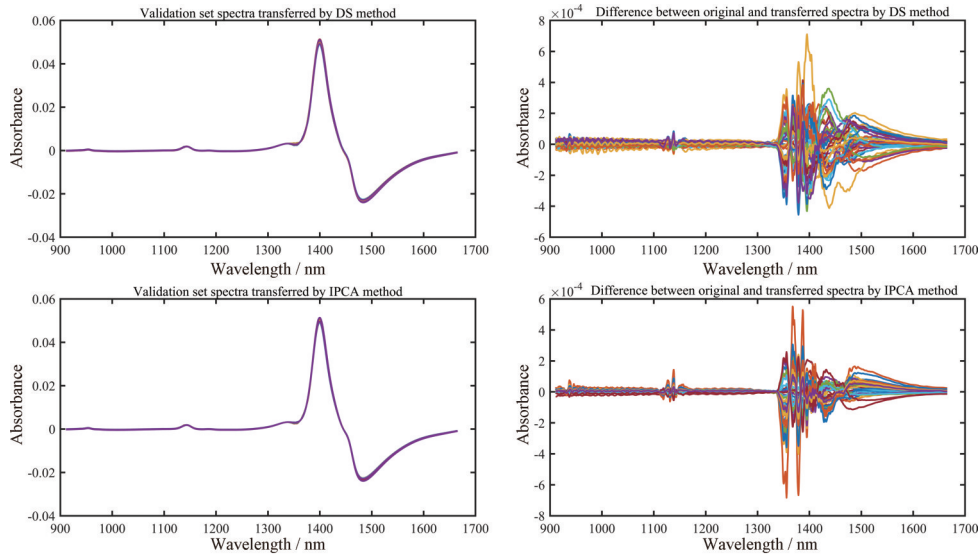


Figure 8 Comparison of the spectra transferred from the target to the source instrument by the DS method and the IPCA method

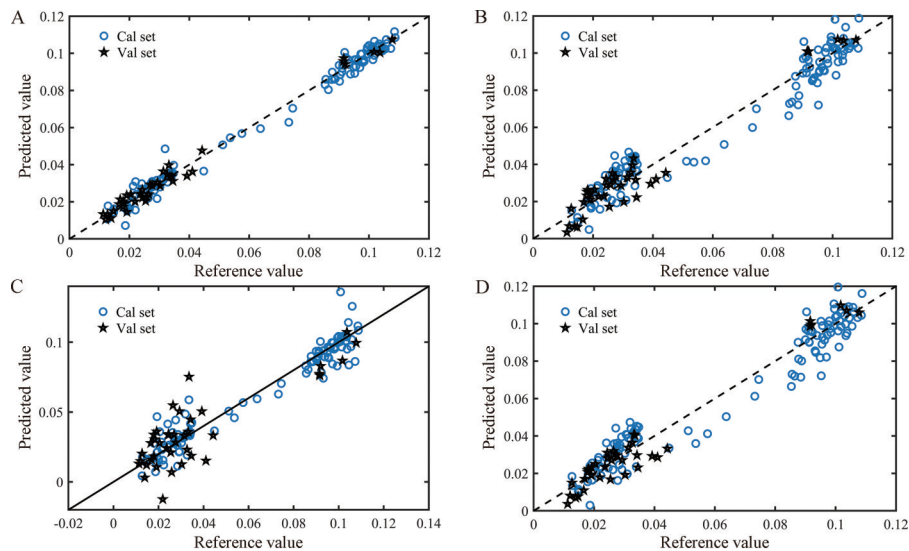


Figure 9 DS method and IPCA method before and after the transfer from target instrument to source instrument model prediction results. A: Modeling results of source instruments; B: Target instrument modeling results; C: Modeling results after DS method model transfer; D: Modeling results after IPCA model transfer

Table 4 Results of model transfer from target instrument to source instrument based on the same wavelength range using two methods

Instrument	Number of transfer set samples	LVs/variable	R_c	RMSEC	R_p	RMSEP
Source instrument	-	8	0.994 3	0.003 7	0.994 0	0.003 1
Target instrument	-	8	0.973 1	0.008 1	0.976 8	0.006 7
Target instrument after DS transfer	78	8	0.970 5	0.006 9	0.857 3	0.014 7
Target instrument after IPCA transfer	78	8	0.968 1	0.008 8	0.980 9	0.005 9

结果存在一定的不稳定因素,而IPCA法可通过调整主成分因子,先将源仪器转移集光谱进行主成分分解得到得分和负载矩阵,然后目标仪器转移集光谱与得分矩阵关联建模,由于较少数量的主成分得分对源仪器转移集光谱有较好的综合解释能力,这样关联建模模型稳定性较好。而源仪器光谱的负载矩阵保留了高分辨光谱的特征信息,再通过主成分得分与负载矩阵重构源仪器的光谱,计算结果比较稳定。由目标仪器向源仪器方向的转移,IPCA法表现的转移效果相较DS法更优。

6 模型的外部验证

为进一步考察模型的稳健性及应用性,本研究利用未参与模型建立的一批提取液中37个样本作为独立验证集,对模型的性能进行了验证。外部验证的结果如表5,由表可知,DS法的由源仪器向目标仪器转移与IPCA法的两个方向的转移都表现出较好的效果,且IPCA方法转移效果相对DS方法更优,进一步证明了本研究所提出的方法可应用于实际生产。

Table 5 Results of independent verification set. R_1 : Determination coefficient of external validation; RMSET: Root mean squares error of external validation

Direction of transfer	Method	R_1	RMSET
Transferred from the source to the target instrument	DS	0.986 1	0.009 2
	IPCA	0.986 2	0.006 5
Transferred from the target to the source instrument	DS	0.606 3	1.040 7
	IPCA	0.986 5	0.006 6

结论

模型转移实际上是将源仪器光谱和目标仪器光谱建立关联,模型转移后的模型性能不仅与源仪器的性能有关,也与目标仪器的性能有关。本文以小儿消积止咳口服液的提取过程为研究对象,对比DS和IPCA两种模型转移方法,实现两种不同类别(分析级与便携式)的近红外光谱仪间模型转移。研究表明,采用IPCA法将模型由目标仪器向源仪器方向转移效果优于模型由源仪器向目标仪器转移效果,而DS法则是由源仪器向目标仪器转移的效果较好。本研究解决了便携式仪器分辨率低、性能差、预测精度不足的问题,以实现对小兒消积止咳口服液进行现场快速筛查和质量分析,对中药生产过程质量监控具有指导意义。

作者贡献: 许秀华是本文的主要完成者; 聂磊提供数据分析指导; 马晓博、庄晓琪、张晋进行数据分析和图表绘制; 董海玲、梁文燕协助查询相关文献并进行数据整理; 杜昊忱、袁晓梅、关永霞提供仪器样本和实验方法指导; 张惠、李连、郭学平、臧恒昌对文章撰写进行指导并提出修改意见。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Zhao CF, Li QJ, Wang LP, et al. High performance liquid chromatography fingerprint and composition analysis of *Senecio cannabifolius* Less [J]. Chin J Anal Chem (分析化学), 2013, 41: 133-136.
- [2] Shen X, Liu HN. Effects of the extracts of traditional Chinese medicines on the extracellular polysaccharide production of *Grifola frondosa* fermentation system [J]. J Fung Res (菌物研究), 2021, 19: 115-121.
- [3] Huang Q, Chen YM, Chou GX, et al. Study on thin-layer chromatographic fingerprint of *Rhizoma Atractylodis* [J]. Shanghai J Tradit Chin Med (上海中医药杂志), 2013, 47: 107-109.
- [4] Su M. Application of Near Infrared Spectroscopy in Quality Evaluation of Radix Astragali Extract (近红外光谱技术在黄芪提取物质量评价中的应用研究) [D]. Ji'nan: Shandong University, 2020.
- [5] Li LS. Study on Near-infrared Spectroscopy Detection Technology and Calibration Transfer of Fruit Quality (水果品质近红外光谱检测技术及模型传递方法研究) [D]. Nanchang: East China Jiaotong University, 2022.
- [6] Hu LP, Huang SQ, Tian SH, et al. Model building and transfer between spectrometers in the application of near infrared spectroscopy to online quality control of Chinese herbal liquid tonic [J]. Chem Ind Eng Prog (化工进展), 2020, 39: 3263-3272.
- [7] Liu XS, Chen JS, Chen GQ, et al. On-line detection of concentration process of Ganmaoling granules by near infrared spectroscopy combined with automatic control system [J]. Acta Pharm Sin (药学报), 2017, 52: 462-467.
- [8] Yuan M. Application of Chemometrics in Quality Analysis and Evaluation of Xinkeshu Tablet (化学计量学在心可舒片质量分析及评价中的应用研究) [D]. Ji'nan: Shandong University, 2022.
- [9] Xu YY. Quantitative Analysis of Methanol in Methanol Gasoline Based on NIR Spectroscopy Combined with Calibration Transfer Technolog (基于NIR光谱结合模型转移技术的甲醇汽油中甲醇定量分析方法研究) [D]. Xi'an: Northwestern University, 2022.
- [10] Chen Q. Model Transfer Strategy of NIR Analysis for Pulpwood Lignin Based on DS and PDS (基于DS与PDS的纸浆材木质素近红外分析模型传递策略研究) [D]. Nanjing: Nanjing Forestry University, 2022.
- [11] Zhang H, Tan H, Lin B, et al. Improved principal component analysis (IPCA): a novel method for quantitative calibration transfer between different near-infrared spectrometers [J]. Molecules, 2023, 28: 406.
- [12] Zhang H. Chemometrics Methods Development and Analyzer Design for Raw Plasma with Near-infrared Spectroscopy Technology (原料血浆近红外光谱分析的化学计量学方法开发及专用分析仪设计) [D]. Ji'nan: Shandong University, 2021.