

近红外光谱的头孢类药品成分分析与模型传递方法

周子堃^{1,2}, 李晨曦^{2*}, 王哲^{1,2}, 刘蓉^{1,2}, 陈文亮^{1,2}, 徐可欣^{1,2}

1. 天津大学精密测试技术及仪器国家重点实验室, 天津 300072
2. 天津大学精密仪器与光电子工程学院, 天津 300072

摘要 药品安全与质量监管迫切需要在线、快速、低成本的成分检测技术。近红外光谱技术在检测成本及速度方面具有显著优势, 基于近红外光谱的药品成分检测方法, 对于提高药品质量监管水平有着十分重要的研究意义和应用价值。在实际应用中, 不同光谱仪器由于性能参数不同, 测量光谱存在一定差异, 很难实现定量校正模型共享。因此, 研究不同光谱仪器之间模型传递对于提高分析效率十分重要。针对头孢类药品成分检测的需要, 研究了头孢类药品中三种组分定量校正模型, 提出了一种基于马尔可夫链(MC)的转换集选择的仪器间定量校正模型传递方法。采用两台不同厂家光谱仪器分别测量56份不同批次的头孢拉定颗粒样品, 针对样品的三种组分: 头孢拉定、头孢苄苄和水分, 使用偏最小二乘法(PLS)建立定量校正模型。通过构建概率矩阵, 选择合适的转换集, 提高模型转换效率及不同仪器得到光谱数据的建模预测精度。实验结果表明, 利用该模型转移算法, 可利用少量转换集样本实现不同光谱仪器间定量校正模型转移, 模型转移前后, 定量校正模型对于三种主成分预测相对误差从9.67%, 52.14%和19.25%, 分别下降到4.37%, 31.12%和11.67%。利用该模型传递方法可以有效修正主成分仪器光谱差异, 实现了不同仪器测量光谱及定量分析模型传递共享。该研究的建模分析与模型传递方法也为药品成分与质量检测提供了技术支撑。

关键词 近红外光谱; 药品成分检测; 定量校正模型; 模型传递; 马尔可夫链

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)11-3562-05

引言

药品安全与质量是关系到民众健康的重要社会问题, 发展在线、快速、低成本的药品成分与质量检测方法是保证其质量与安全的关键^[1]。药品成分检测方法主要有色谱法、物理常数测定法、官能团测定法等^[2-3], 这些方法往往需要较为复杂的预处理步骤, 分析时间较长, 在药品生产以及抽样检查中应用较多, 但是在现场快检与在线监测中应用较少。近红外光谱方法具有无损、快速的优势, 能够提供准确的定性、定量分析结果, 广泛应用于成分检测及过程控制领域^[4]。

近年来, 基于近红外光谱技术的药品成分检测方法成为研究热点, 并在质量与生产现场检测中应用广泛。魏学敏等^[5]详细介绍了光谱技术与化学计量学方法在药物分析方面应用的研究情况, 近红外光谱技术在成本及检测速度等方面

具有显著优势。Boyer等^[6]采用近红外光谱法测定了抗疟疾抗生素中有效成分的含量, 建立了校正模型并进行验证, 结果显示成分预测误差在5%以内。Carvalho等^[7]结合分线性判别分析和遗传算法线性判别分析, 建立了近红外光谱分类模型, 判别精准度达到了94.4%。在实际应用中, 光谱测量仪器种类及型号较多, 由于光谱仪参数及测量条件变化, 所带来的波长偏移及噪声分布特性改变, 均会在一定程度上影响光谱建模分析精度。因此, 研究不同仪器之间预测模型传递方法具有较为重要的意义。Abel等^[8]提出了两种新的模型传递算法, 并在应用中与最大似然主成分分析和分段直接校正法两种成熟的模型传递方法的性能进行比较, 并取得了较好的结果。研究不同仪器之间所建立的光谱定量模型传递方法, 减小或消除不同仪器测得的光谱的差异, 对于提高近红外光谱分析效率, 推广其在各个领域的应用, 具有较高的研究意义和应用价值^[9-11]。

针对头孢类药品成分检测需要, 研究了基于偏最小二乘

收稿日期: 2019-09-19, 修订日期: 2020-01-21

基金项目: 国家自然科学基金(81871396, 81971657, 81671727, 81471698, 81401454), 天津市自然科学基金(19JCYBJC29100), 天津市科技特派员计划项目, 国家重大科学仪器设备开发专项(2014YQ060773), 国家(863)高技术研究发展计划(2012AA022602)资助

作者简介: 周子堃, 1993年生, 天津大学精密测试技术及仪器国家重点实验室硕士研究生 e-mail: 137779278@qq.com

* 通讯联系人 e-mail: lichenxi@tju.edu.cn

法 PLS 与分段直接校正法 PDS 的头孢拉定颗粒组分定量分析与模型传递方法。采用两台傅里叶光谱仪分别采集头孢拉定颗粒样本的光谱数据, 建立基于 PLS 方法的头孢类药品组分的定量分析模型。提出了基于马尔可夫链的转换校正集选择及分段校正模型转移方法, 该方法通过构建概率矩阵, 选择合适的转换集提高模型转换效率及精度。实验结果表明, 利用本文所建立的定量分析模型及传递方法, 可实现头孢拉定颗粒样本中头孢拉定、头孢氨苄和水分三种成分定量精准分析, 减少不同仪器所建立模型传递误差, 与传统的 K-S 算法相比, 其对不同仪器光谱差异的修正效果有着明显的提高。本研究的建模分析与模型传递方法也为药品成分与质量检测提供了技术支撑。

1 实验部分

1.1 样品及光谱测量

实验样品为头孢拉定颗粒, 分别选自于白云山药业、金鸿药业等十余家公司生产的不同批次药物, 共计 56 个样本。与其质量密切相关的 3 种组分: 头孢拉定、头孢氨苄和水分含量的参考值由中国食品药品检定研究院检测给出, 其中头孢拉定组分浓度范围为 $30 \sim 130 \text{ mg} \cdot \text{g}^{-1}$, 头孢氨苄组分浓度范围为 $0.5 \sim 3 \text{ mg} \cdot \text{g}^{-1}$, 水分浓度范围为 $0.1 \sim 2.3 \text{ mg} \cdot \text{g}^{-1}$ 。

光谱测量采用两台傅里叶光谱仪, 其一为珀金埃尔默仪器有限公司(PerkinElmer)的 FrontierTM FT-IR/NIR 分析仪(主机), 使用积分球测量附件, 检测器为 MCT 检测器; 另一台为赛默飞世尔科技(ThermoFisher)的 AntarisTM II FT-NIR 分析仪(从机), 使用积分球测量附件和 InGaAs 检测器。光谱测量范围为 $10\,000 \sim 4\,000 \text{ cm}^{-1}$, 分辨率为 8 cm^{-1} , 测量信噪比可达到 $3\,000:1$ 以上。为了进一步提高光谱信噪比, 光谱扫描平均次数为 32 次。样品放置于入石英平底皿中并压实, 保持每次测量时样品厚度及紧实度一致, 减少散射影响, 并保证样品厚度大于入射光程。经研究对比发现, 两台光谱仪测得光谱的差异主要为基线漂移、少量波长漂移和噪声。

1.2 光谱预处理及建模

光谱预处理主要目的是去除基线漂移, 进一步提高光谱信噪比, 保证光谱建模分析准确度。根据样品特点, 首先采用标准正态变量校正 SNV 消除由于样本颗粒分布不均匀及颗粒大小不同产生的散射对光谱的影响; 然后, 采用 S-G 三次多项式 9 点平滑滤波降低噪声。偏最小二乘法 PLS 是最常用的定量校正模型方法, 它集成了主成分分析、典型相关分析、线性回归分析的优点, 适用于解决实际问题中数据量较少, 且变量之间存在多重相关性的情况^[12]。PLS 不仅分解了光谱矩阵 X, 同样也对浓度矩阵 Y 进行了处理, 同时消除了光谱矩阵和浓度矩阵里包含的噪声, 提升了模型准确性与抗干扰能力。

1.3 基于马尔可夫链的模型传递

多元校正算法容易出现过拟合的现象, 为了避免此类情形, 常在模型传递前选择一部分样品集, 即转换集用于参数

的计算。

马尔可夫链(Markov Chain, MC)是机器学习等领域中一个很重要的方法, 表征状态空间中从一个状态到另一个状态的随机过程^[13]。假设存在随机变量集合, 且随机变量的条件概率满足 $p(X_{t+1} | X_t, \dots, X_1) = p(X_{t+1} | X_t)$, 则称 X 为马尔可夫链。在模型传递过程中, 通常选取部分样本作为转换集, 用于计算模型传递参数。在 MC 思想基础上, 每一个样本都视为一个独立的状态, 模型传递的过程可以近似为状态空间中从一个状态到另一个状态的随机过程。将模型传递的评价参数进行处理后作为状态转移概率 $P_{i,j}$ 。遍历所有的 m 个样本, 可以得到 $m \times m$ 个转移概率值用于构建转移矩阵 P

$$P = \begin{bmatrix} P^*(1, 1) & P^*(1, 2) & \cdots & P^*(1, m) \\ P^*(2, 1) & P^*(2, 2) & \cdots & P^*(2, m) \\ \vdots & \vdots & \ddots & \vdots \\ P^*(m, 1) & P^*(m, 2) & \cdots & P^*(m, m) \end{bmatrix} \quad (1)$$

对于不可约的马尔可夫链, 其平稳分布就是它的极限分布, 即对于任意初始分布 v_0 , 在转移矩阵 P 不变的情况下有: $v_0 \lim_{T \rightarrow \infty} P^T = \pi$, 其中 T 为演变步数。在确定样品转换集时, 设置初始概率分布 v_0 服从均匀分布, 即 $v_0 \sim U(0, 1)$ 。通过计算马尔可夫链的平稳分布, 以当前分布与上一步分布的均方误差 MSE 来表示分布的收敛情况, 计算公式如式(2)

$$\text{MSE} = \frac{1}{m} \| \pi_k - \pi_{k-1} \|^2 \quad (2)$$

式中, π_k 表示第 k 步时的概率分布, m 为校正集数目。根据其平稳分布, 可以优化选择转换集样品, 得到以转移概率 $P_{i,j}$ 作为约束条件下的最优解, 从而提高模型转换效率及建模准确度。

1.4 模型评价

光谱定量校正模型中, 一般采用校正均方根误差(RMSEC)、预测均方根误差(RMSEP)、决定系数(R^2)和相对预测误差(RPE)等指标评价定量模型预测精度。为了客观的评定模型传递效果, 本文中定义光谱平均差异(ARMS)和光谱校正率($\text{Pr}_{\text{corrected}}$)两个参数评价模型传递效果^[14], 其计算公式如式(3)和式(4)

$$\text{ARMS} = \frac{1}{n_c} \sum_{i=1}^{n_c} \sqrt{\frac{1}{p} \sum_{\lambda=1}^p (S_{2\lambda}^i - S_{1\lambda}^i)^2} \quad (3)$$

$$\text{Pr}_{\text{corrected}}(\%) = \frac{\text{ARMS}_{\text{un}}^2 - \text{ARMS}^2}{\text{ARMS}_{\text{un}}^2} \quad (4)$$

其中, $S_{1\lambda}^i$ 和 $S_{2\lambda}^i$ 表示样本 i 中的波长点 λ 在主从仪器上的光谱数据; ARMS_{un} 表示传递前主从仪器光谱的平均差异, ARMS 表示传递后主从仪器光谱的平均差异, ARMS 越小、Pr 越大表示传递后主从仪器光谱越接近。

2 结果与讨论

2.1 单台仪器光谱建模预测结果

分别利用两台光谱仪测量得到光谱, 建立头孢拉定中三种组分定量校正模型, 使用 RANK 法将 56 个样本分为校正集和验证集, 并采用蒙特卡洛交互验证法去除校正集和验证

集中的异常样本,模型预测结果如表 1 所示。实验结果可知,样品中主要成分头孢拉定含量的预测精度最高,验证集样本的相对预测误差均在 5% 以内,而组分头孢氨苅和水分的含量相对较低,部分样品中的组分含量已经接近检测限,所以这两种组分建立模型的相对预测误差较大。综合对比,三种成分在两台仪器上建立的校正模型的预测效果差距不大,其中 Frontier 建立模型的预测结果略优于 Antaris 的建模结果。

表 1 两台光谱仪建模参数和预测结果

Table 1 Modeling parameters and prediction results of two spectrometers

成分 仪器	头孢拉定		头孢氨苅		水分	
	Frontier	Antaris	Frontier	Antaris	Frontier	Antaris
RMSEC/(mg · g ⁻¹)	1.83	2.09	0.089	0.188	0.121	0.147
RMSEP/(mg · g ⁻¹)	2.92	3.63	0.304	0.316	0.147	0.189
R ²	0.988	0.982	0.623	0.611	0.874	0.767

2.2 基于 MC 的模型传递分析

光谱校正率 P_r 表征了模型传递效率,即从仪器光谱传递后越是贴近主仪器光谱,则模型传递效果越好。 P_r^* 定义为对样本 i 作转换样本时对应的 m 个光谱校正率作归一化的结果

$$P_r^*(i, j) = \frac{P_r(i, j)}{\sum_{k=1}^m P_r(i, k)} \quad (5)$$

将光谱校正率 P_r 作为转移概率构建马尔可夫链状态转移矩阵,设置初始最大步数为 20,演变步数和 MSE 的关系如图 1 所示。

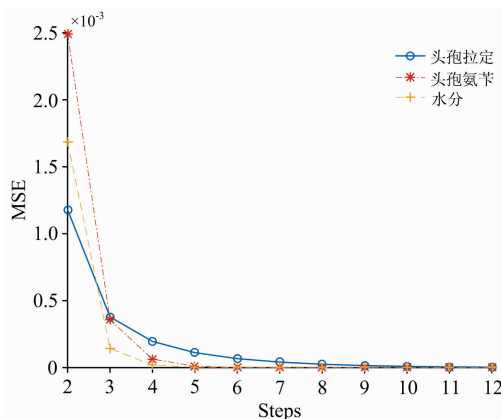


图 1 演变步数与均方误差的关系

Fig. 1 Relation between evolution step and mean square error

结果表明,在步数大于 10 步后三个马尔可夫链的分布均已接近收敛。根据马尔可夫链的平稳分布特性以 RMSEP 和 $Pr_{correct}$ 为标准,依次选择若干个样本作为转换集。三种组分的主从机模型传递中转换集选择情况分别如图 2(a), (b) 和 (c) 所示。当转换集数目为 6 时,光谱校正率的变化已经趋于平稳,当转换集数目大于 10 个后,随着样本数的继续增

加, RMSEP 减小并不明显。故在预测均方根误差的变化平稳后,基于增强模型稳健性的原则,选择头孢拉定组分的转换集数目为 10 个,头孢氨苅组分的转换集数目为 13 个,水分的转换集数目为 11 个。

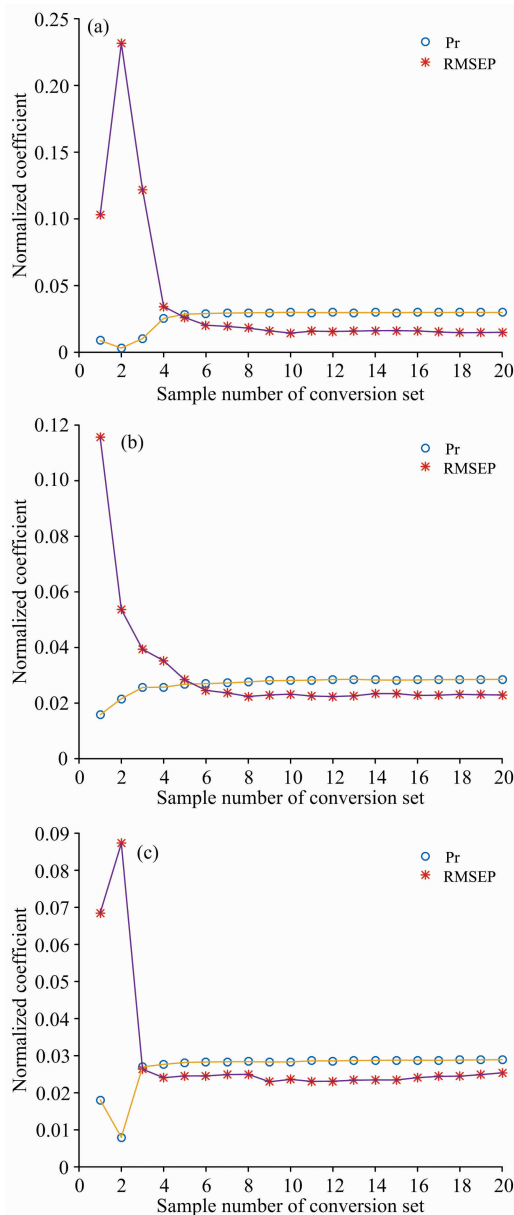


图 2 RMSEP 和 Pr 的归一化参数与转换集数目关系

(a): 头孢拉定; (b): 头孢氨苅; (c): 水分

Fig. 2 The relation between the normalized parameters (RMSEP and PR) and the number of conversion sets

(a): Cefradine; (b): Cephalosporin; (c): Moisture content

2.3 模型传递后建模预测结果分析

由于主从仪器上测得的光谱具有较大差异,在模型传递前,直接使用主仪器模型对从仪器上测得的光谱数据进行预测,得到的预测结果和参考值的误差较大。在 MC 方法选择转换集基础上,使用 PDS 算法进行模型传递,并使用 DS 算法对窗口数据进行传递,三种组分的模型传递参数及结果如

表 2 所列。根据实验结果,传递后可以有效减少光谱差异,以主机建立校正模型,利用从机光谱数据传递后进行预测为例,预测结果 RMSEP 和 RPE 在模型传递后得到了极大的改善,模型预测精度与主仪器光谱的预测精度基本相当。

表 2 模型传递参数及传递前后光谱差异

Table 2 Calibration transfer parameters and spectral differences

成分	头孢拉定	头孢氨苄	水分
窗口宽度	9	11	9
传递后 ARMS	0.075	0.076	0.083
Pr/%	91.1	94.7	94.2
主仪器 RMSEP/(mg·g ⁻¹)	2.264	0.210	0.131
从仪器传递前 RMSEP/(mg·g ⁻¹)	8.82	0.481	0.362
从仪器传递后 RMSEP/(mg·g ⁻¹)	3.50	0.314	0.172
从仪器传递前 RPE/%	9.67	52.14	19.25
从仪器传递后 RPE/%	4.37	31.12	11.67

将 MC 算法与使用 K-S 算法选择转换集^[15]时的相关结果进行对比,表 3 列出了两种方法使用后的模型参数及评价参数。对比光谱校正率 Pr, MC 的结果明显优于 K-S 方法。对于质量较好的模型,如头孢拉定定量模型,两种方法建立传递模型后的 RMSEP 相差较小;对于头孢氨苄定量模型,结合表 1 中的主仪器模型相关数据,使用 MC 选择的转换集建立的传递模型使从仪器光谱更好的拟合了主仪器光谱。这些数据充分说明了使用马尔可夫链选择转换集是可行的,并在一些情况下会比使用 K-S 方法取得更好的结果。

References

- [1] Roggo Y, Chalou P, Maurer L, et al. *Journal of Pharmaceutical and Biomedical Analysis*, 2007, 44(3): 683.
- [2] Srikar A, Swapna D, Swathi G, et al. *International Journal of Pharmacy and Technology*, 2010, 2(1): 16.
- [3] MA Bo-kai, GOU Xin-lei, ZHAO Xin-ying(马博凯, 勾新磊, 赵新颖). *Food Safety and Quality Detection Technology(食品安全质量检测学报)*, 2016, 7(11): 4295.
- [4] Măntele W. *Journal of Biological Physics*, 2003, 9(2): 87.
- [5] WEI Xue-min, WU Qian, LIU Qiang, et al(魏学敏, 吴倩, 刘强, 等). *Chinese Journal of Pharmaceutical Analysis(药物分析杂志)*, 2013, 33(8): 1447.
- [6] Boyer C, Gaudin K, Kauss T, et al. *Journal of Pharmaceutical and Biomedical Analysis*, 2012, 67: 10.
- [7] Carvalho L C, Morais C L M, Lima K M G, et al. *Food Analytical Methods*, 2018, 11(7): 1857.
- [8] Abel F F, Raffaele V, Onno E de Noord, et al. *Journal of Chemometrics*, 2017, 31(3): e2874.
- [9] Filzmoser P, Todorov V. *Analytica Chimica Acta*, 2011, 705(1-2): 2.
- [10] Workman J J. *Applied Spectroscopy*, 2018, 72(3): 340.
- [11] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). *Chinese Journal of Analytical Chemistry(分析化学)*, 2002, 30(1): 114.
- [12] Cortés V, Talens P, Barat J M, et al. *Postharvest Biology and Technology*, 2019, 148: 236.
- [13] Chotard A, Auger A, Hansen N. Markov Chain Analysis of Evolution Strategies on a Linear Constraint Optimization Problem. 2014 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2014. 159.
- [14] TIAN Gao-you, CHU Xiao-li, YUAN Hong-fu, et al(田高友, 褚小立, 袁洪福, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2006, (7): 927.
- [15] Harshvardhan S, Ajaya K P, Hare K M. *Measurement*, 2019, 134: 698.

表 3 两种模型传递方法建模预测结果比较

Table 3 Calibration transfer and evaluation parameters of two algorithms

成分 方法	头孢拉定		头孢氨苄		水分	
	MC	K-S	MC	K-S	MC	K-S
ARMS _{un}	0.251	0.244	0.332	0.317	0.345	0.328
ARMS	0.075	0.081	0.076	0.083	0.083	0.091
Pr/%	91.1	89.0	94.7	93.1	94.2	92.3
RMSEP/(mg·g ⁻¹)	3.50	3.59	0.314	0.288	0.172	0.179

3 结论

近红外光谱方法具有无损、快速的优势,在药品成分检测与质量控制方面应用广泛。针对药品质量与成分检测的需要,研究了基于近红外光谱的头孢类药品的定量分析及模型传递方法。采用两台傅里叶光谱仪(主机: Frontier, 从机: Antris)分别采集 56 个头孢拉定颗粒样本的近红外光谱,并建立针对头孢拉定、头孢氨苄和水分三种成分的定量分析预测模型。在此基础上,提出了基于马尔可夫链的转换校正集选择方法,通过构建概率矩阵,选择合适的转换集提高模型转换效率及不同仪器得到光谱数据的建模预测精度。实验结果表明,与 K-S 算法相比,本文模型传递方法对不同仪器光谱差异的修正效果有着明显的提高。模型传递前后,不同仪器使用同一模型预测三种成分相对误差从 9.67%, 52.14% 和 19.25%, 分别下降到 4.37%, 31.12% 和 11.67%。利用本模型传递方法可以有效修正主从仪器光谱差异,实现了不同仪器测量光谱及定量分析模型传递共享,也为药品成分与质量检测提供了技术支撑。

Study on Cefradine Granules Component Analysis and Calibration Transfer Method Based on Near-Infrared Spectroscopy

ZHOU Zi-kun^{1,2}, LI Chen-xi^{2*}, WANG Zhe^{1,2}, LIU Rong^{1,2}, CHEN Wen-liang^{1,2}, XU Ke-xin^{1,2}

1. State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
2. School of Precision Instrument and Optic Electronic Engineering, Tianjin University, Tianjin 300072, China

Abstract Near-infrared spectroscopy (NIRS) technology has distinct advantages in component detection for its characteristics of high-speed and low-cost, which is essential for the supervision of drug quality and safety. Studying the method of drug component detection based on NIRS technology is significant for improving the level of drug quality supervision. In fact, owing to differences in performance parameters of different spectroscopic instruments, spectra measured are discrepancy, which brings hardship to the realization for quantitative correction models sharing. Therefore, in order to improve analysis efficiency, the calibration transfer method is discussed. In this paper, the establishment of cephalosporins component correction model and calibration transfer method are studied, and a transformation set selection method based on Markov chain (MC) is proposed. Fifty-six samples of cefradine granules in different batches were used. Spectral data were measured by two Fourier spectrometers. For three components of the sample: cefradine, Cefalexin and water, partial least squares (PLS) method was used to establish a quantitative correction model. MC algorithm is used to construct the probability matrix and select the conversion set, which improves the efficiency of model transformation and the prediction accuracy of spectral data. The experimental results show that the quantitative calibration model transfer between different spectroscopic instruments can be realized by using a small number of sample sets. After the model transfer, the relative error of the quantitative calibration model for the three principal components prediction decreases from 9.67%, 52.14%, 19.25% to 4.37%, 31.12%, 11.67%, respectively. The spectral differences between master and slave instruments can be corrected effectively, and the transfer and sharing of measurement spectra and quantitative analysis models of different instruments can be realized. The modeling analysis and model transfer methods studied in this paper also provide technical support for drug composition and quality detection.

Keywords Near-infrared spectroscopy; Drug composition detection; Quantitative calibration model; Calibration transfer; Markov chain

(Received Sep. 19, 2019; accepted Jan. 21, 2020)

* Corresponding author