

## 不同分光原理近红外光谱仪光谱标准化方法 在小麦粉品质检测中的应用

田 静<sup>1</sup>, 陈 斌<sup>1\*</sup>, 陆道礼<sup>1</sup>, 盛龙禹<sup>2</sup>, 蔡贵民<sup>3</sup>

<sup>1</sup> 江苏大学食品与生物工程学院 江苏镇江 212013

<sup>2</sup> 上海中科航谱光电技术有限公司 上海 200086

<sup>3</sup> 上海棱光技术有限公司 上海 200023)

**摘要** 为探究不同分光原理的近红外光谱仪的光谱标准化方法,降低光谱差异性,实现模型共享,以小麦粉的粗蛋白为检测指标,以 3 台不同原理的近红外光谱仪采集的小麦粉样品近红外光谱为研究对象,分别采用直接标准化(DS)、分段直接标准化(PDS)和一元线性回归直接标准化(SLRDS)算法进行近红外光谱标准化研究,通过欧氏距离( $D$ )、光谱标准化误差率(SSER)等指标分析主、从机光谱间的光谱差异性,以预测相关系数( $R_p$ )、标准偏差(RMSEP)和相对标准偏差(RPD)等参数,评估模型转移后的预测效果;结果表明:1)经 3 种算法的光谱标准化,从机与主机之间的光谱差异性显著降低,转移后模型的从机预测效果大幅度提升,其中从机  $S_2$  的预测效果优于从机  $S_1$ ;2)3 种算法中,光谱经 DS 算法标准化后 2 台从机的预测误差率均最小,其中从机  $S_1$  的  $SSER_{ave}$  和  $SSER_{max}$  分别为 0.9057 和 3.3667,  $R_p$ 、RMSEP、RPD 分别为 0.8949, 0.7052, 2.2408, 从机  $S_2$  的  $SSER_{ave}$  和  $SSER_{max}$  分别为 0.6595 和 4.3691,  $R_p$ 、RMSEP、RPD 分别达到 0.9687, 0.4105, 4.0284。3 种算法的光谱标准化方法均可应用于不同分光原理的近红外光谱仪,其中 DS 算法效果较好,实现了小麦粉粗蛋白含量近红外校正模型的共享。

**关键词** 近红外光谱; 小麦粉; 光谱标准化; 模型传递

**文章编号** 1009-7848(2022)10-0286-09 **DOI:** 10.16429/j.1009-7848.2022.10.031

近红外光谱技术(Near infrared spectroscopy, NIRS)经过多年的发展已相当成熟,尤其是在食品、农产品、医药、材料等领域得到广泛的应用<sup>[1-3]</sup>。NIRSA 是基于样品组分中基团在近红外光谱区有特征吸收,通过构建吸收光谱信息和待测指标之间的定量关系,实现对样品品质的快速检测技术。其优势体现在快速高效、绿色无损、可多组分同时检测<sup>[4-6]</sup>。近红外光谱模型在使用过程中,由于检测条件、检测环境或仪器设备发生变化,近红外光谱的吸光度会出现差异,导致建成的校正模型失效,而重新建立模型需耗费大量的人力、物力<sup>[7-8]</sup>,因此,实现仪器间的模型共享是 NIRS 实际应用中的瓶颈问题之一,是该技术走向生产实践必须解决的问题<sup>[9-10]</sup>。

近年来,国内外大量研究人员针对近红外光谱分析的模型传递问题进行了较为深入的研究,

并取得许多成果<sup>[11-13]</sup>。目前近红外模型传递的研究多数局限于相同分光原理的仪器之间,而对不同分光原理的近红外光谱仪间的模型传递鲜有报道。如能实现不同分光原理的仪器间的光谱标准化、模型传递,可有效降低不同分光原理仪器需独立建立校正模型的成本,对 NIRSA 技术在各行业的应用与推广具有重要的意义。

以小麦粉为试验样品,以光栅扫描型、法布里-珀罗干涉仪型、声光可调滤光器型 3 种不同分光原理的 3 台近红外光谱仪的小麦粉近红外光谱为研究对象,分别采用直接标准化(Direct standardization, DS)、分段直接标准化(Piecewise direct standardization, PDS)、一元线性回归直接标准化(Simple linear regression direct standardization, SLRDS)等 3 种算法对 3 种分光原理近红外光谱仪采集的同一样品光谱进行标准化,使用欧氏距离、光谱标准化误差率 2 个指标定量描述同一样品的从机光谱标准化前、后与主机光谱间的差异,距离和误差率越小,表示样本光谱间的差异越小,从而使近红外光谱校正模型在不同原理仪

收稿日期: 2021-10-11

基金项目: 国家重点研发计划项目(2018YFE0196600)

作者简介: 田静(1997—),女,硕士生

通信作者: 陈斌 E-mail: ncp@ujs.edu.cn

器间的共享的预测误差最小。

## 1 模型传递原理

模型传递主要是通过建立主机和从机光谱、模型参数或预测值之间的函数关系,校正因仪器差异、检测条件或检测环境因素变化而导致的预测误差。模型传递方法依照校正对象可分为 3 类:对模型参数进行校正,如两步偏最小二乘方法等;对预测结果进行校正,如斜率/偏差算法<sup>[14]</sup>等;对光谱进行校正,即光谱标准化后再带入模型计算,如分段直接标准化等。

对模型参数进行校正,是通过添加一系列新环境、新仪器等条件下测定的样品光谱来扩充原模型,并通过光谱预处理方法、稳健回归算法等增强模型的预测效果,使模型能适用于新测试条件下采集的样品光谱。这种方法在使用过程中,需尽可能多地选择不同的样本、仪器、测样环境等,然后将这些光谱都应用到模型的建立过程,以扩充其变化范围<sup>[15]</sup>。同时,当仪器或环境发生改变时,

需对建模样本进行不断扩充,这是一个相当复杂的过程。另外,随着仪器的不断老化、仪器部件的更换,这种方法很难满足模型传递所需达到的目的。

对预测结果进行校正的前提是假设主机和从机的预测结果呈线性关系。当主机、从机光谱差异不大时,该方法能获得较好的校正效果。然而,绝大部分情况下由于假设并不成立,因此其对模型传递的效果较差,适用范围较窄。

光谱标准化的思路与前面两种不同(如图 1 所示),光谱标准化首先是对模型的输入进行校正,即对同一样品在不同仪器所采集的光谱数据进行校正,减少各仪器间的光谱背景与强度的影响,使得同一样品在不同仪器得到的光谱尽可能一致(此时标准化后的光谱是样品的虚拟光谱),降低光谱间的差异性,有效解决不同仪器间光谱校正模型的传递,光谱标准化方法可以实现不同分光原理的仪器间的模型共享。

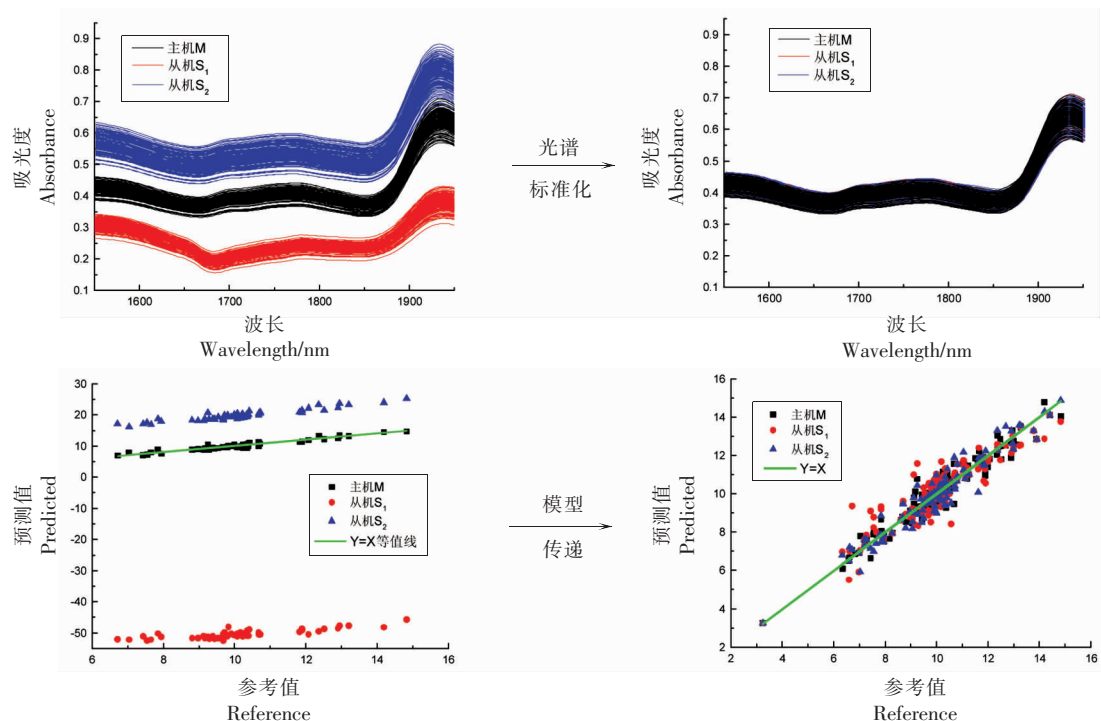


图 1 光谱标准化与模型传递的差异

Fig.1 The differences between spectral standardization and model transfer

## 2 试验部分

### 2.1 样品与仪器

试验所用样品购自国内各地区线下超市和线

上的小麦粉,共计 154 个样本,其中包括:高筋小麦粉、中筋小麦粉、低筋小麦粉、自发粉和全麦粉等品种,收集到的样品冷藏条件下保存,备用。

取出小麦粉样品,置于室温下平衡 24 h 后采集光谱。

光谱采集仪器:1 台光栅扫描 (Dispersive scanning) 型 S450 近红外光谱仪,上海棱光科技有限公司 (简称 S450), 工作波长范围 900~2 500 nm, 波长间隔 1 nm; 1 台法布里-珀罗干涉仪 (Fabry-Perot interferometer, FPI) 型 N500 近红外光谱仪, 济南海能仪器股份有限公司 (简称 N500), 扫描波长范围 1 550~1 950 nm, 波长间隔 2 nm; 1 台声光可调滤光器 (Acousto optical tunable filter, AOTF) 型近红外光谱仪, 上海中科航谱光电技术有限公司研发 (简称 AOTF), 扫描波长范围 1 100~2 300 nm, 波长间隔 1 nm。以 S450 作为主机, 记为 M, 以 N500 及 AOTF 作为从机, 分别记为 S<sub>1</sub>、S<sub>2</sub>, 采集样品光谱。每个样品重复扫描 3 次, 结果取平均光谱。

理化指标测定仪器:采用 D200 杜马斯定氮仪 (济南海能仪器股份有限公司) 测定样品的粗蛋白含量。由于光谱标准化的前提是主、从机采集的光谱具有相同波段范围和同样的数据间隔, 而 3 台不同类型的近红外光谱仪器的工作波长和光谱数

据的波长间隔不同, 其中 N500 的光谱数据的波长范围最短 (1 550~1 950 nm), 波长间隔最大 (2 nm), 因此在光谱标准化前需将 3 台仪器的波长范围和间隔统一。N500 仪器每 2 nm 取 1 个波长点, 共 201 个波长点, S450、AOTF 仪器每隔 1 nm 取 1 个波长点, 共 401 个波长点。使用 Excel 2016 软件对 S450、AOTF 仪器光谱数据的波长范围进行剪切, 在截取 1 550~1 950 nm 波长范围后, 抽取与 N500 相同的波长点, 以得到 201 个波长点后的数据作为该样品的光谱数据, 并在该波长区间进行光谱标准化。采用 SPXY 算法 (Sample set partitioning based on joint x-y distance), 按 6:4 的比例将样本划分为校正集 (92 个) 与预测集 (62 个)。

## 2.2 化学值测定

小麦粉样品的粗蛋白含量检测按照 GB/T 31578-2015《粮油检验 粮食及制品中粗蛋白测定 杜马斯燃烧法》方法, 采用 D200 杜马斯定氮仪, 每个样品平行检测 3 次 (相对极差不超过 4%), 以 3 次检测结果的均值为最终值。小麦粉样品的粗蛋白含量测定结果见表 1。

表 1 小麦粉样品的粗蛋白含量分布表

Table 1 Crude protein content of wheat flour samples

| 样品  | 数量/个 | 范围/%       | 均值/%  | 标准差  | 变异系数 |
|-----|------|------------|-------|------|------|
| 小麦粉 | 154  | 6.71~14.83 | 10.16 | 1.78 | 0.18 |

## 2.3 光谱差异分析

采用欧氏距离和光谱标准化误差率 2 个指标定量评价光谱间的差异性, 对同一样品的从机光谱标准化前、后与主机光谱之间的差异进行评价。

2.3.1 欧氏距离 采用光谱间的欧式距离 ( $D$ ) 定量评价仪器间的光谱差异。 $D$  越大, 仪器间光谱的差异越明显。

$$D = \sqrt{\sum_{k=1}^n (A_{ik} - \bar{A}_k)^2}, (k=1 \cdots n) \quad (1)$$

式中,  $k$ ——波长点;  $A_{ik}$ ——从机上采集的光谱在第  $k$  个波长点处、第  $i$  条光谱的吸光度值;  $\bar{A}_k$ ——标准光谱 (主机) 在第  $k$  个波长点的吸光度值。

2.3.2 光谱标准化误差率 使用光谱标准化误差率 (Spectral standardization error rate, SSER) 评价

指标表征不同仪器间的光谱标准化后的准确度, 定量描述同一样品的从机光谱经标准化后与主机光谱之间的差异。光谱标准化误差率越小, 光谱与主机光谱的差异越小。

对第  $i$  个样本的光谱标准化误差率定义为:

$$SSER_i = \frac{\frac{1}{K} \sum_{j=1}^K |M_{ij} - Z_{ij}|}{\frac{1}{K} \sum_{j=1}^K |Z_{ij}|} \times 100\% \quad (2)$$

$$Z_{ij} = \frac{M_{ij} + T_{ij}}{2}, i=1, \dots, N, j=1, \dots, K \quad (3)$$

式中,  $\{M_{ij}, i=1, \dots, N, j=1, \dots, K\}$ ——主机采集的样本光谱矩阵;  $\{T_{ij}, i=1, \dots, N, j=1, \dots, K\}$ ——从机光谱经光谱标准化后的样本光谱矩阵;  $N$ ——样本数;  $K$ ——光谱的波长点数。

对于全部样品集合,其平均误差率  $SSER_{ave}$  和最大误差率  $SSER_{max}$  定义为:

$$SSER_{ave} = \frac{1}{N} \sum_{i=1}^N SSER_i \quad (4)$$

$$SSER_{max} = \max(SSER_i) \quad (5)$$

## 2.4 模型评价标准

模型建立过程中采用校正相关系数 ( $R_c$ ) 和校正标准偏差 (RMSEC) 对模型的性能进行评价,建立最优的校正模型。模型建立完成后,通常采用预测相关系数 ( $R_p$ )、预测标准偏差 (RMSEP) 和相对标准偏差 (RPD) 等指标来综合评价模型的预测性能。RMSEC、RMSEP 越小,  $R_c$  和  $R_p$  越接近 1, 所建模型的稳定性与预测性能越好。RPD 用来评价模型精度的指标,当  $RPD < 1.75$  时,预测精度过低,模型不可用;当  $RPD > 3$  时,表明模型的预测精度较高。

## 2.5 数据处理与分析

采用江苏大学近红外工作室自主研发的 NIRSA 5.8.8 系统 (计算机软件著作权登记号为 2007SR06801)、IBM SPSS Statistics 25 和 Excel 2016 等软件平台进行数据分析。

## 2.6 光谱标准化方法

采用 DS、PDS、SLRDS 算法对光谱进行标准化,以实现不同分光原理的仪器间模型传递。

DS 算法主要是通过主机标样集的光谱与从机标样集的光谱之间的数学关系建立光谱标准化

转移矩阵,借助转移矩阵对从机上采集的光谱进行校正,从而减少不同仪器间所测同一样本光谱数据的差异。PDS 算法与 DS 算法原理相似,不同点是 PDS 从采集到的光谱上分出连续的波段,在每个波段窗口分别计算变换系数,根据每个波段窗口的变换系数建立光谱标准化转移矩阵,利用转移矩阵对从机上采集的光谱进行校正,从而实现主、从机光谱最大程度的相似,试验采用左右波段窗口区域宽度  $\omega=3$ 。SLRDS 算法是假设不同波长点间的吸光度是相互独立的,利用一元线性回归对从机上的光谱进行校正。

## 3 结果与分析

### 3.1 主机模型的建立

以主机 M 上采集的 92 个校正集样品的近红外光谱及其粗蛋白含量数据为研究对象,通过偏最小二乘法 (Partial least square, PLS) 建立分析模型,用于模型传递。为了消除光谱数据中无关信息和噪音的干扰,使用多元散射校正 (Multiplicative scatter correction, MSC)、标准正态变量变换 (Standard normal variate transformation, SNV)、归一化 (Normalization) 和中心化 (Centralized) 等常用的近红外光谱预处理方法对样品光谱进行预处理,通过建立的 PLS 校正模型评价预处理方法的优、劣,选择最优预处理方法。不同预处理方法的校正模型评价结果见表 2。

表 2 不同预处理方法的样品蛋白 PLS 校正模型评价

Table 2 Evaluation of PLS correction model for protein samples under different pretreatment methods

| 预处理方法   | 主成分数 | 校正集    |        | 预测集    |        | RPD    |
|---------|------|--------|--------|--------|--------|--------|
|         |      | $R_c$  | RMSEC  | $R_p$  | RMSEP  |        |
| 无       | 7    | 0.9552 | 0.5002 | 0.9312 | 0.6072 | 3.3788 |
| SNV     | 8    | 0.9697 | 0.4760 | 0.9449 | 0.5120 | 4.0933 |
| 归一化     | 6    | 0.9678 | 0.4872 | 0.9329 | 0.5655 | 3.9727 |
| MSC     | 9    | 0.9678 | 0.4811 | 0.9330 | 0.5655 | 3.9727 |
| SNV+MSC | 7    | 0.9690 | 0.4830 | 0.9443 | 0.5146 | 4.0476 |
| 归一化+MSC | 7    | 0.9715 | 0.4552 | 0.9620 | 0.4267 | 4.2187 |
| 一阶微分    | 5    | 0.9606 | 0.5353 | 0.9408 | 0.5427 | 3.5898 |
| 二阶微分    | 5    | 0.7465 | 1.3499 | 0.8215 | 0.9863 | 1.5029 |

由表 2 可知,对比不同预处理方法建模效果,其中主成分数为 7 时,归一化+MSC 预处理后的校正模型效果较好,此时 PLS 模型的  $R_c$  值为 0.971,

RMSEC 值为 0.4552,  $R_p$  值为 0.9620, RMSEP 为 0.4267, RPD 为 4.2187, 模型具有较高的预测精度。MSC 可以消除样品杯中因小麦粉样品分布不

均匀等因素导致的散射系数差异对其漫反射光谱的影响。经归一化+MSC 预处理后的光谱建模效果最佳。

### 3.2 不同仪器采集光谱的差异

在3台近红外光谱仪上采集全部小麦粉样品的光谱图如图2所示。可清楚看到,两台从机与主机所采集的光谱有较明显的差异。从机所采样品光谱的吸光度在纵轴上存在一定的平移。主机与从机采集的光谱在1650~1750 nm及1850~1950 nm波段范围差异较大,这可能是因光学元器件和分光效率引起的仪器间的系统误差所致。

计算主机所有样品的平均光谱记为标准光

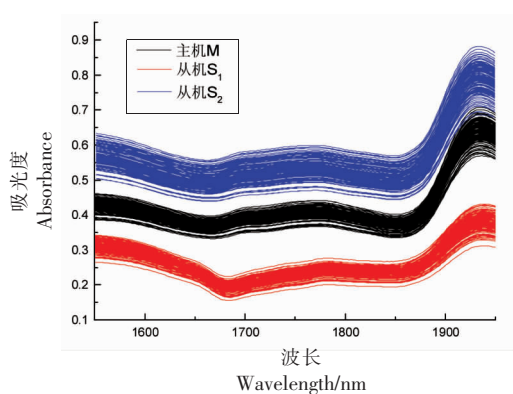


图2 3台仪器采集的近红外光谱平均图

Fig.2 Mean near-infrared spectrogram obtained by three instruments

### 3.3 从机光谱标准化前的预测结果与分析

利用主机模型分别对3台仪器测量的62个预测集样品进行预测分析,3台仪器的预测结果及预测散点图如表3和图4所示。

由表3可知,将主机建立的小麦粉粗蛋白校正模型用于主机验证集样品预测时,模型预测精度较高, $R_p$ 达0.9620, RMSEP和RPD分别为0.4267和4.2187。当该模型对2台从机的验证集

表3 模型传递前主机模型对3台仪器预测集的预测结果

Table 3 The prediction results of the host model to 3 instrument verification sets before model transfer

| 仪器       | $R_p$  | RMSEP   | RPD    |
|----------|--------|---------|--------|
| 主机 M     | 0.9620 | 0.4267  | 4.2187 |
| 从机 $S_1$ | 0.8169 | 60.5540 | 1.0028 |
| 从机 $S_2$ | 0.9389 | 9.6848  | 1.7333 |

谱,分别计算各仪器上测得样品光谱与标准光谱的欧式距离,记录各距离值的频数分布,其分布情况如图3所示。

由图3可看出,主机采集的样品光谱与标准光谱的距离大部分在0.5以下,而从机采集的样品光谱与标准光谱的欧式距离较大。其中,从机 $S_1$ 采集的样品光谱与标准光谱的距离大多在2.0~3.0范围,从机 $S_2$ 采集的样品光谱与标准光谱的距离大多在1.0~2.5范围。这可能是因选用的3台近红外光谱仪的分光原理不同,故所采样品光谱信息差异较为显著。

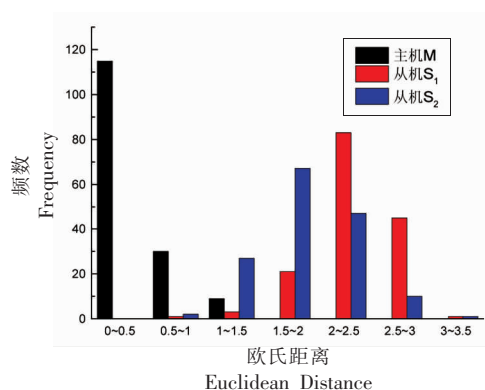


图3 不同仪器小麦粉样品光谱的欧式距离频数分布

Fig.3 Spectrum Euclidian distance frequency distribution of wheat flour samples with different instruments

样品预测时, $R_p$ 也较高,分别为0.8169和0.9389,而RMSEP和RPD两项评价指标较差,说明3台光谱仪器间存在较大的系统误差,主机的校正模型不能与2台从机共享。单一使用 $R_p$ 指标不能全

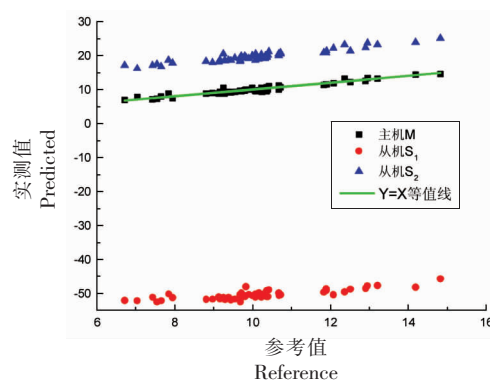


图4 校正前主机模型对3台仪器的预测散点图

Fig.4 Forecast scatter diagram of 3 instruments by host model before calibration

面评价模型的性能,需与 RMSEP 和 RPD 等指标同时使用,才能全面评价模型的预测效果。

### 3.4 光谱标准化方法

DS、PDS、SLRDS 算法均为有监督的算法,需选择标样集,且标样集样品数的选取对光谱标准化效果有重要的影响,样品数过少会导致信息不

充分,而过多会增加数据处理的难度,出现过拟合的假象。采用 Kennard-Stone(K-S)算法从主、从机的校正集中分别取 10,20,30,40,50,60 个样品作为光谱标准化的标样集。3 种算法的标样集样品数与 RMSEP 大小的关系如图 5 所示。

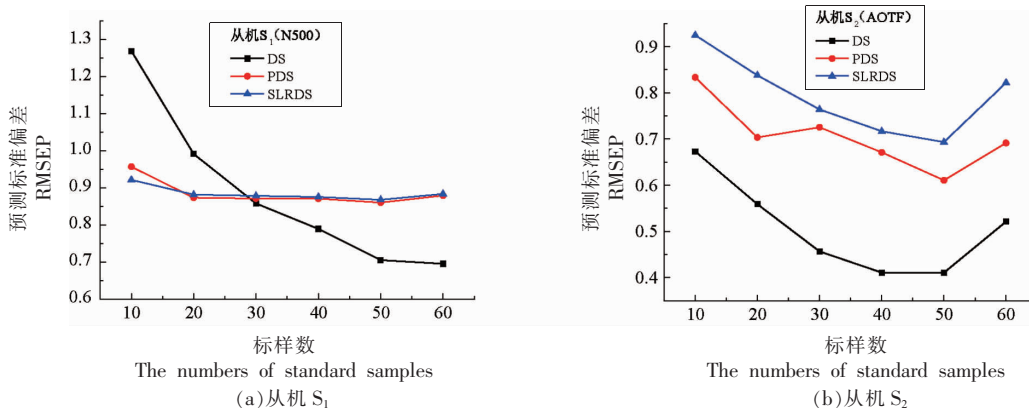


图 5 采用 DS、PDS 和 SLRDS3 种算法的 2 台从机 RMSEP 随标样集样品数的变化

Fig.5 RMSEP of two slave machines varies with the number of samples in the standard sample set

图 5 显示,随着样品数的增加,标样集中所含有用信息增多,RMSEP 的数值降低。由图 5a 可知,针对从机 S<sub>1</sub>,在 DS 算法中,RMSEP 数值随标样集样品数的增加而显著降低,当样品数大于 50 时,RMSEP 变化不大。在 PDS 和 SLRDS 算法中,随着样品数的增加,RMSEP 数值逐渐降低,变化趋势较为平稳。由图 5b 可知,针对从机 S<sub>2</sub>,3 种算法的 RMSEP 数值整体呈现随标样集样品数的增

加而减少的趋势,当样品数大于 50 时,RMSEP 出现增大的趋势。最终选择 50 个样品作为标样集。

### 3.5 标准化后的光谱差异及模型传递效果

选取 50 个样品为标样集,分别采用 DS、PDS、SLRDS 算法对 2 台从机采集的样品光谱进行光谱标准化,计算标准化后的样品光谱与标准光谱的欧式距离,记录各距离值的频数分布,如图 6 所示。

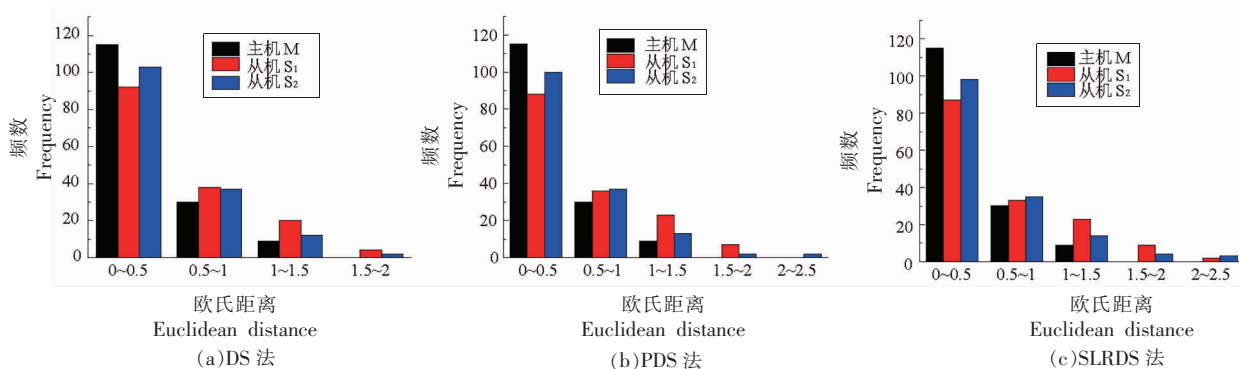


图 6 光谱标准化后不同仪器小麦粉样品光谱的欧式距离频数分布

Fig.6 Euclidean distance frequency distribution of wheat flour samples with different instruments after spectral standardization

图 6 显示,经 3 种方法光谱标准化后,2 台从机光谱与标准光谱的欧式距离均降至 2.0 以下,

距离大多数分布在 0~1.0 范围。这表明光谱标准化方法大幅度提高了主、从机采集光谱的一致性,

而不能单从光谱间的欧氏距离评价3种光谱标准化方法的优、劣,还需使用光谱标准化误差率和模型预测效果进行评价。

将经DS、PDS、SLRDS算法的从机预测集标

准化光谱输入主机模型中进行预测。分别计算3种算法应用于2台从机后的光谱标准化误差率,比较光谱标准化平均误差率、最大误差率及模型传递后预测效果,结果如表4所示。

表4 标准化后的光谱差异及模型传递效果

Table 4 Spectral differences after standardization and model transfer effect

| 仪器                | 算法    | 光谱标准化               |                     | 模型共享           |         |        |
|-------------------|-------|---------------------|---------------------|----------------|---------|--------|
|                   |       | SSER <sub>ave</sub> | SSER <sub>max</sub> | R <sub>p</sub> | RMSEP   | RPD    |
| 从机 S <sub>1</sub> | 无     | 48.7666             | 57.2004             | 0.8169         | 60.5540 | 1.0028 |
|                   | DS    | 0.9057              | 3.3667              | 0.8949         | 0.7052  | 2.2408 |
|                   | PDS   | 2.9271              | 8.2078              | 0.8376         | 0.8704  | 1.8305 |
|                   | SLRDS | 2.9265              | 8.2062              | 0.8370         | 0.8779  | 1.8274 |
| 从机 S <sub>2</sub> | 无     | 27.3701             | 38.5168             | 0.9389         | 9.6848  | 1.7333 |
|                   | DS    | 0.6595              | 2.8471              | 0.9687         | 0.4105  | 4.0284 |
|                   | PDS   | 0.7558              | 3.0622              | 0.8700         | 0.8436  | 2.0282 |
|                   | SLRDS | 0.7572              | 3.0624              | 0.8703         | 0.8376  | 2.0304 |

由表4可知,2台从机采集的样品光谱经3种算法标准化后,与主机光谱的平均误差率、最大误差率大幅度降低,这表明标准化后从机光谱与主机光谱的差异性明显降低。同时,主机模型对其的预测效果较标准化前也有不同程度的提升。结果表明,3种算法对从机S<sub>2</sub>的标准化效果较好,优于从机S<sub>1</sub>;主机模型对从机S<sub>2</sub>的预测效果也优于从机S<sub>1</sub>。2台从机光谱均经DS算法标准化后效果最好,其对从机S<sub>1</sub>的SSER<sub>ave</sub>和SSER<sub>max</sub>分别为0.9057和3.3667,R<sub>p</sub>、RMSEP、RPD分别达0.8949,0.7052,2.2408;对从机S<sub>2</sub>的SSER<sub>ave</sub>和SSER<sub>max</sub>分别为0.6595和2.8471,R<sub>p</sub>、RMSEP、RPD分别达0.9687,0.4105,4.0284,模型预测精度较光谱标准化前有大幅度的提升。

#### 4 结论

针对小麦粉中粗蛋白含量的近红外光谱校正模型,在3台不同分光原理的近红外光谱仪间共享问题,采用DS、PDS、SLRDS算法开展光谱标准化及模型传递研究。结果表明:

1) DS、PDS、SLRDS算法均属于有监督的光谱标准化算法,随着样品数的增加,标样集中所含有用信息增多,3种算法的RMSEP数值整体呈现随转换集样品数的增加而减小的总体趋势;也要避免因样品数增加过多而出现拟合的现象。经

比较,小麦粉样品的光谱标准化时,选择50个样品参与光谱标准化,小麦粉粗蛋白的近红外校正模型传递的误差最小。

2) 经3种算法对光谱标准化后,2台从机采集的光谱与主机光谱差异性显著降低,主机模型对其的预测效果大幅提升,其中对从机S<sub>2</sub>的预测效果的提升优于从机S<sub>1</sub>。

3) 3种算法中,2台从机光谱均经DS算法标准化后误差率最小,主机模型预测效果最好,模型预测精度较光谱标准化前有大幅提升。

#### 参 考 文 献

- [1] WILLIAMS P C. Application of chemometrics to prediction of some wheat quality factors by near-infrared spectroscopy[J]. Cereal Chemistry, 2020, 97(5): 958-966.
- [2] 赵昕,张任,王伟,等.基于近红外光谱快速定量检测面粉中曲酸的方法建立[J].食品科学,2018,39(8):249-255.  
ZHAO X, ZHANG R, WANG W, et al. Using near-Infrared (NIR) spectroscopy for rapid, quantitative detection of kojic acid in wheat flour[J]. Food Science, 2018, 39(8): 249-255.
- [3] 赵丽华,巩元勇,张洁,等.近红外光谱法快速测定藜麦籽粒粗蛋白含量[J].食品工业科技,

- 2020, 41(15): 233–236, 243.
- ZHAO L H, GONG Y Y, ZHANG J, et al. Rapid determination of quinoa seeds crude protein content using near infrared spectroscopy[J]. *Science and Technology of Food Industry*, 2020, 41(15): 233–236, 243.
- [4] YE D D, SUN L J, ZOU B R, et al. Non-destructive prediction of protein content in wheat using NIRS[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2018, 189: 463–472.
- [5] TIAN W F, CHEN G J, ZHANG G R, et al. Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics[J]. *Food Chemistry*, 2021, 344: 128633.
- [6] JIANG H, LIU T, CHEN Q S. Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system[J]. *Infrared Physics and Technology*, 2020, 109: 103423.
- [7] 吴彦华, 范永涛. 近红外光谱分析中温度影响的修正[J]. *光谱学与光谱分析*, 2020, 40(6): 1690–1695.
- SUN Y H, FAN Y T. Correction of temperature influence in near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(6): 1690–1695.
- [8] 刘燕德, 饶宇, 孙旭东, 等. 尺寸差异对脐橙糖度可见近红外光谱检测模型影响[J]. *光谱学与光谱分析*, 2020, 40(10): 3241–3246.
- LIU Y D, RAO Y, JIANG X D, et al. Size effect on the near-infrared spectroscopy detection model of navel orange[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(10): 3241–3246.
- [9] HUANG Z X, SANAEIFAR A, TIAN Y, et al. Improved generalization of spectral models associated with Vis-NIR spectroscopy for determining the moisture content of different tea leaves [J]. *Journal of Food Engineering*, 2021, 293: 110374.
- [10] CHEN J, ZHU S P, ZHAO G H. Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR[J]. *Food Chemistry*, 2017, 221: 1939–1946.
- [11] WANG A D, YANG P, CHEN J, et al. A new calibration model transferring strategy maintaining the predictive abilities of NIR multivariate calibration model applied in different batches process of extraction[J]. *Infrared Physics and Technology*, 2019, 103: 103046.
- [12] MCVEY C, MCGRATH T F, HAUGHEY S A, et al. A rapid food chain approach for authenticity screening: The development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices[J]. *Talanta*, 2021, 222: 121533.
- [13] ZHANG L G, LI Y Q, HUANG W, et al. The method of calibration model transfer by optimizing wavelength combinations based on consistent and stable spectral signals[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, 227: 117647.
- [14] 张进, 蔡文生, 邵学广. 近红外光谱模型转移新算法[J]. *化学进展*, 2017, 29(8): 902–910.
- ZHANG J, CAI W S, SHAO X G. New algorithms for calibration transfer in near infrared spectroscopy [J]. *Progress in Chemistry*, 2017, 29(8): 902.
- [15] 刘耀瑶, 杨浩, 熊智新, 等. 制浆材木质素含量近红外分析模型传递研究[J]. *中国造纸学报*, 2019, 34(3): 43–49.
- LIU Y Y, YANG H, XIONG Z X, et al. Study on near-infrared calibration model transfer for lignin content in pulpwood[J]. *Transactions of China Pulp and Paper*, 2019, 34(3): 43–49.

### Application of Spectral Standardization of Different Spectral Types of Near-infrared Analyzers in the Quality Detection of Wheat Flour

Tian Jing<sup>1</sup>, Chen Bin<sup>1\*</sup>, Lu Daoli<sup>1</sup>, Sheng Longyu<sup>2</sup>, Cai Guimin<sup>3</sup>

<sup>1</sup>School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu

<sup>2</sup>Shanghai Zhongke Aerospectrum Optoelectronic Technology Co., Ltd., Shanghai 200086

<sup>3</sup>Shanghai Lengguang Technology Co., Ltd., Shanghai 200023)

**Abstract** To explore the spectral standardization method of NIR spectrometers with different spectral principles, the



crude protein of wheat flour was taken as the specific detection standard, and the spectra of wheat flour samples collected by three NIR spectrometers with different principles were taken as the research object. The standardization method of near infrared spectrum is studied by using direct standardization(DS) and piecewise direct standardization (PDS) and simple linear regression direct standardization (SLRDS) algorithms respectively, by Euclidean distance ( $D$ ), spectral standardized error rate (SSER) spectral differences between the spectral analysis of the main, from the machine, The prediction effect was evaluated by predicting correlation coefficient ( $R_p$ ), standard deviation of prediction (RMSEP) and relative standard deviation (RPD). The results show that: 1) After the spectral standardization of the three algorithms, the difference between the master and the slave spectrum is significantly reduced, and the prediction effect of the transferred model is greatly improved. The prediction effect of the slave  $S_2$  is better than that of the slave  $S_1$ . 2) Among the three algorithms, the prediction error rates of the two slave machines are the smallest after the DS algorithm is standardized.  $SSER_{ave}$  and  $SSER_{max}$  of the slave machine  $S_1$  are 0.9057 and 3.3667,  $R_p$ , RMSEP and RPD are 0.8949, 0.7052 and 2.2408,  $SSER_{ave}$  and  $SSER_{max}$  of the slave machine  $S_2$  are 0.6595 and 4.3691,  $R_p$ , RMSEP and RPD are 0.9687, 0.4105 and 4.0284, respectively. The spectral standardization of the three algorithms can be applied to different types of NIR analyzers, and the NIR correction model of crude protein content in wheat flour can be shared.

**Keywords** near infrared spectroscopy; wheat flour; spectral standardization; model transfer