

文章编号 1004-924X(2007)12-1952-07

小型近红外玉米蛋白质成分分析 仪器设计的波段选择

曹 璞¹, 潘 涛¹, 陈星旦^{1,2}

(1. 暨南大学 理工学院 应用光谱实验室, 广东 广州 510632;

2. 中国科学院 长春光学精密机械与物理研究所 应用光学国家重点实验室, 吉林 长春 130033)

摘要:采用傅里叶变换近红外漫反射光谱技术和偏最小二乘法(PLS)建立了玉米蛋白质含量的定标模型。按照预测效果优选光谱波段,为设计小型近红外玉米蛋白质成分分析仪器提供依据。采用多元散射校正方法对光谱进行预处理,然后利用 Savitzky-Golay 平滑法对原谱、一阶导数谱和二阶导数谱进行平滑处理。选取全谱、合频、一倍频、二倍频和蛋白质基团等 5 个波段,每个波段分别采用原光谱、一阶导数谱、二阶导数谱,共建立 15 个定标模型。同时调整 Savitzky-Golay 平滑点数和 PLS 因子数,通过多次 PLS 数值实验比较,按照预测效果确定每个模型的最优平滑点数、因子数,再从中择优。结果表明,采用一阶导数谱的一倍频波段($7\ 000\sim 5\ 500\ \text{cm}^{-1}$)的定标效果最好,模型的预测相关系数、预测均方根偏差、相对预测均方根偏差分别为 0.945, 0.357, 3.340%。一倍频波段可以代替全谱波段并得到更好的定标效果。

关键词:玉米;蛋白质;近红外光谱分析;偏最小二乘法;波段选择

中图分类号:TH744.1;Q-337 **文献标识码:**A

Choice of wave band in design of minitype near-infrared corn protein content analyzer

CAO Pu¹, PAN Tao¹, CHEN Xing-dan^{1,2}

(1. *Laboratory of Applied Spectroscopy, College of Science and Engineering, Jinan University, Guangzhou 510632, China*; 2. *State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*)

Abstract: The calibration models for protein concentration in corn samples were established by Fourier transform near-infrared diffuse reflection spectroscopy and Partial Least Square (PLS) regression. According to the prediction effect the best wave band was chosen to provide a basis for designing a minitype near-infrared corn protein content analyzer. The spectra were processed by Multiplicative Scatter Correction(MSC) method firstly, then the original spectra, the first derivative spectra and the second derivative spectra were processed by Savitzky-Golay smoothing method. The following 5 wave

收稿日期:2007-09-24;修订日期:2007-10-09

基金项目:国家自然科学基金资助项目(No. 10771087);教育部科学技术研究重点项目(教技司[2006]52号 No. 106154);教育部留学回国人员科研启动基金资助项目(教外司留[2005]383号);暨南大学引进优秀人才科研启动基金资助项目

bands, the whole region, combination region, the first overtone region, the second overtone region and the protein functional group bands were selected for establishing 15 calibration models adopting the original spectra, the first derivative spectra and the second derivative spectra respectively in each band. By adjusting the number of Savitzky-Golay smoothing points and number of PLS factors simultaneously and comparing with PLS computational experiments several times, the optimal number of smoothing points and number of factors for each models were obtained based on the prediction effect, and then the best one was selected. The research results show that the prediction effect using the first derivative spectra in the first overtone region is best one, and the correlation coefficient, the Root Mean Square Error of Prediction(RMSER) and the Relative Root Mean Square Error of Prediction (RRMSEP) for the corresponding model are 0.945, 0.357, and 3.340%, respectively. It shows that the first overtone band ($7\ 000\sim 5\ 500\text{ cm}^{-1}$) can replace the whole band, and get better calibration effect.

Key words: corn; protein; near infrared spectroscopic analysis; partial least square regression; wave band choice

1 引 言

随着光电技术、计算机信息技术和化学计量学方法的发展,现代近红外光谱(NIR)技术以其分析速度快、精度适中、成本低、非破坏性和易于实现无损、实时分析等优点在农业、食品、医药、烟草、石油化工等领域中得到广泛应用^[1-6]。玉米的蛋白质含量是玉米品质的一个重要理化指标,采用传统的化学分析方法需要做复杂的预处理,且需要化学试剂进行化学反应,不利于环保,而且测样时间长,过程繁琐,对分析人员的技能要求很高,无法适合在农田实地进行快速简便的定量测试分析。近年来,国内外已在实验室开展了用近红外光谱技术分析玉米成分的工作^[3,5,7-8],其中如何通过优化定标模型指导光电系统设计、研究小型专用光谱仪等具有重要意义。

本文采用近红外漫反射光谱技术,用傅里叶变换近红外光谱仪对 177 个玉米样品采集光谱。利用偏最小二乘法(Partial Least Square,PLS)建立玉米蛋白质含量的定标模型。首先采用多元散射校正方法对光谱进行预处理,再进行原谱、一阶导数谱和二阶导数谱的平滑化处理。选取波段:全谱、合频、一倍频、二倍频和蛋白质基团等 5 个波段,每个波段都分别采用原光谱、一阶导数谱、二阶导数谱,建立了 15 个定标模型。为了提高模型精度,本文通过同时调整平滑点数和因子数,对每个模型都分别做了多次 PLS 数值实验比较,得

到每个模型的最优平滑点数、因子数和和预测均方根偏差(Root Mean Squar Error of Predication, RMSEP),按照 RMSEP 值再从 15 个模型中选优,确定定标效果最好的波段,为设计小型便携式玉米蛋白质测定专用近红外光谱仪提供指导和依据。

2 实验部分

2.1 实验材料

成都某研究所提供了东北产玉米样品。经过外观观察,剔除颜色异常、质量异常的玉米,筛选得到 177 份样品。样品用带有 1.0 mm 孔径筛网的粉碎机磨碎。玉米样品蛋白质含量的参考化学值的测定采用凯氏定氮法(参见国标 GB/T5511-85)获得。

选取其中 139 个为定标样品,其余 38 个为验证样品。表 1 给出玉米样品蛋白质含量化学值的数据统计。

表 1 玉米样品蛋白质含量化学值的数据统计
Tab.1 Chemical datum statistics
of protein concentration of corn samples

	蛋白质含量(%)			
	最小值	最大值	平均值	标准差
定标集	7.40	16.60	10.60	1.31
验证集	8.90	13.40	10.69	1.04

2.2 仪器和测量方法

选用了 Nicolet 公司的 Nexus™ 870 傅里叶变换近红外光谱仪和漫反射附件。设置分辨率为 16 cm^{-1} , 光谱波段为 $10\,000 \sim 4\,000 \text{ cm}^{-1}$ ($1\,000 \sim 2\,500 \text{ nm}$)。

转动样品池可有效减小由于样品不均匀带来的误差, 反复装样可有效减小装样环节引入的测定误差, 多次扫描可有效减小背景噪声的影响。本实验将充分混合均匀的样品装入漫反射附件, 在样品池转动的情况下扫描光谱, 每个样品经过反复装样 5 次连续扫描, 每次扫描 32 遍, 然后计算平均吸收光谱, 实验室温度 (23 ± 1) °C。

2.3 光谱数据处理

采用目前使用最广泛的 PLS^[1,4,5,9,10] 建立了定标模型。模型评价指标包括定标相关系数 (R_c)、预测相关系数 (R_p)、定标均方根偏差 (Root Mean Square Error of Calibration, RMSEC)、预测均方根偏差 (Root Mean Square Error of Prediction, RMSEP)、相对定标均方根偏差 (Relative Root Mean Square Error of Calibration, RRMSEC %) 和相对预测均方根偏差 (Relative Root Mean Square Error of Prediction, RRMSEP %):

$$R_c^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_{ic})^2}{\sum_{i=1}^n (y_i - y_{mc})^2}; R_p^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n (y_i - y_{mp})^2}$$

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{ic})^2}{n - 1}};$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_i - y_{ip})^2}{m - 1}};$$

$$\text{RRMSEC} = \frac{\text{RMSEC}}{y_{mc}} \times 100(\%);$$

$$\text{RRMSEP} = \frac{\text{RMSEP}}{y_{mp}} \times 100(\%).$$

式中: y_i 为第 i 样品的常规方法测定得到的参考化学值; y_{ic} 为定标集中第 i 样品的预测值; y_{ip} 为验证集中第 i 样品的预测值; y_{mc} 为定标集中样品化学值的平均值; y_{mp} 为验证集中样品化学值的平均值; n 为定标集的样品数; m 为验证集的样品数。

多元散射校正 (Multiplieative Scatter Cor-

rection, MSC) 的作用是校正吸收基线并降低样品散射作用对光谱的影响, 降低样品的不均匀性带来的光谱差异性^[9-11]。因为本实验的样品为固体, 并采用漫反射附件测定, 颗粒度不是很均匀, 因此, 实验中的所有光谱都先用多元散射校正来做预处理。

合理的光谱平滑化和光谱求导 (一阶、二阶导数光谱) 处理可以消除基线漂移和平缓背景干扰形成的高频随机噪声的影响, 从而提供比原光谱更高的分辨率和更清晰的光谱轮廓变化, 提高光谱信噪比^[12-16]。本文利用 Savitzky-Golay 平滑化方法^[17] 来计算原谱、一阶导数谱和二阶导数谱。它与移动平均平滑法的基本思想类似, 是利用多项式来对移动窗口内的数据进行多项式最小二乘拟合。平滑点数是重要的参数指标, 平滑点数过少容易产生新的计算误差 (噪音) 从而造成模型精度下降, 平滑点数过多则会使得包含样品信息的光谱数据磨光丢失, 同样也会造成模型精度下降。因此, 合理使用平滑点数非常重要, 但必须通过多次数值实验的反复比较, 才能选取确定最优平滑点数。下面以平滑点数 15 为例, 给出对应的一阶、二阶导数光谱的计算公式:

$$\begin{aligned} dA(\nu) = & \frac{1}{334152\Delta\nu^2} \\ & \{12922[A(\nu+7\Delta\nu) + A(\nu-7\Delta\nu)] - \\ & 4121[A(\nu+6\Delta\nu) + A(\nu-\Delta\nu)] - \\ & 14150[A(\nu+5\Delta\nu) + A(\nu-5\Delta\nu)] - \\ & 18334[A(\nu+4\Delta\nu) + A(\nu-4\Delta\nu)] - \\ & 17842[A(\nu+3\Delta\nu) + A(\nu-3\Delta\nu)] - \\ & 13843[A(\nu+2\Delta\nu) + A(\nu-2\Delta\nu)] - \\ & 7506[A(\nu+\Delta\nu) + A(\nu-\Delta\nu)]\} \end{aligned}$$

$$\begin{aligned} d^2A(\nu) = & \frac{1}{6188\Delta\nu^2} \\ & \{91[A(\nu+7\Delta\nu) + A(\nu-7\Delta\nu)] + \\ & 52[A(\nu+6\Delta\nu) + A(\nu-\Delta\nu)] + \\ & 19[A(\nu+5\Delta\nu) + A(\nu-5\Delta\nu)] - \\ & 8[A(\nu+4\Delta\nu) + A(\nu-4\Delta\nu)] - \\ & 29[A(\nu+3\Delta\nu) + A(\nu-3\Delta\nu)] - \\ & 48[A(\nu+2\Delta\nu) + A(\nu-2\Delta\nu)] - \\ & 53[A(\nu+\Delta\nu) + A(\nu-\Delta\nu)] - \\ & 56A(\nu)\} \end{aligned}$$

其中, ν 为波数 (cm^{-1}), $A(\nu)$ 、 $dA(\nu)$ 、 $d^2A(\nu)$ 分别

为原光谱、一阶导数谱、二阶导数谱的吸光度值, $\Delta\nu$ 为光谱数据的波数间隔, 在本实验中, 分辨率 16 cm^{-1} 对应 $\Delta\nu=7.713(\text{cm}^{-1})$ 。

另一方面, 偏最小二乘法是融合主成分分析和多元线性回归的新型化学计量学方法, 其本质是先消除光谱共线性再做回归, 其中的因子数是最重要的参数, 它对应了代表样品信息的光谱数据的加权组合的组数。如果使用的因子数过少, 就不能充分反映样品信息, 模型预测精度会降低, 称为不充分拟合。如果使用的因子数过多, 就会引入一些代表噪声干扰的成分数据, 模型的预测能力也会下降, 称为过拟合。因此, 合理确定偏最小二乘法的因子数, 对于充分利用光谱信息和消除噪声非常重要, 但必须通过多次数值实验的反复比较, 才能选取确定。

3 结果与讨论

图 1 给出了 177 个玉米样品的在仪器测定范围($10\ 000\sim4\ 000\text{ cm}^{-1}$)内的吸收光谱。由图可见, 光谱重叠严重, 吸收较弱。其中 $5\ 500\sim4\ 000\text{ cm}^{-1}$ 、 $7\ 000\sim5\ 500\text{ cm}^{-1}$ 、 $10\ 000\sim7\ 000\text{ cm}^{-1}$ 分别对应分子振动的合频、一倍频、二倍频(含高频)波段。

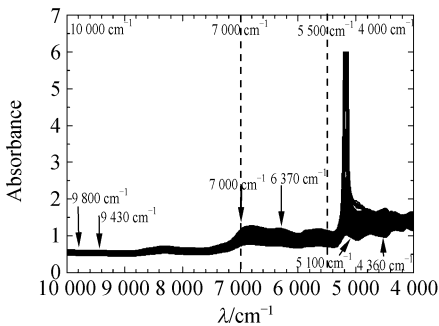


图 1 177 个玉米样品的吸收光谱
Fig. 1 Absorption spectra of 177 corn samples

图 2 给出了玉米样品的吸光度与蛋白质含量化学值的相关系数谱。由图可见, 在每一个波长点上, 177 个玉米样品的吸收度与蛋白质含量化学值的相关系数都很低(绝对值低于 0.22)。因此, 玉米样品的蛋白质含量与任何单独波长点的吸光度没有显著的相关性, 直接从图谱中某一

个波长点来确定其含量是不可行的。由于样品光谱有多个平缓的吸收峰, 因此必须考虑用多个谱区内的光谱数据来建立定标模型。

根据蛋白质的标志基团 N-H 基团在近红外谱区的吸收波长分布, 本文对应筛选了下列三个波段 $9\ 800\sim9\ 430\text{ cm}^{-1}$ 、 $7\ 000\sim6\ 370\text{ cm}^{-1}$ 、 $5\ 100\sim4\ 360\text{ cm}^{-1}$ 来组合(参见图 1), 简称为“基团”, 因为这个波段组合对应了样品的分子结构特征, 尝试采用它的吸光度来定标建模, 也有可行性。

本文分别对 5 个波段进行建模并比较预测效果从中选优。参见图 1, 5 个波段分别是: 全谱($10\ 000\sim4\ 000\text{ cm}^{-1}$)、合频波段($5\ 500\sim4\ 000\text{ cm}^{-1}$)、一倍频波段($7\ 000\sim5\ 500\text{ cm}^{-1}$)、二倍频(含高频)波段($10\ 000\sim7\ 000\text{ cm}^{-1}$)和蛋白质基团波段($9\ 800\sim9\ 430\text{ cm}^{-1}+7\ 000\sim6\ 370\text{ cm}^{-1}+5\ 100\sim4\ 360\text{ cm}^{-1}$)。5 个波段都分别采用原光谱、一阶导数谱、二阶导数谱, 共建立 15 个定标模型来比较优选。

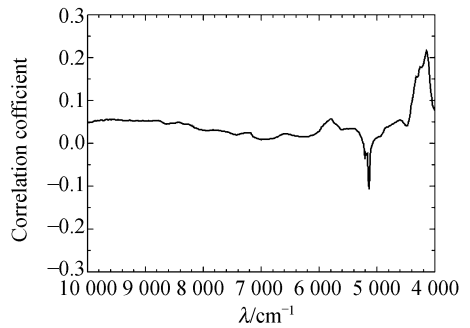


图 2 玉米样品的吸光度与蛋白质含量化学值的相关系数谱
Fig. 2 Correlation coefficient spectrum for absorbance and chemical value of protein concentration of corn samples

采用偏最小二乘法建立了定标模型, 其中 RMSEP 是最实质性的描述模型精度的评价指标, 本文将它作为建模的优化目标, 光谱预处理的平滑点数和偏最小二乘法的因子数是建模的预设参数, 因此, 这是一个双参数的目标优化问题。每个定标模型都设定 3、4、5、6、7 共 5 个因子数, 并设定 15、17、19、21、23、25、27、29、31 共 9 种平滑点数, 从而有 45 个参数组合, 对每一个模型, 每一个参数组合都分别做一次 PLS 数值实验, 共做了

15 × 5 × 9 = 675 次 PLS 数值实验来比较选优。利用 Nicolet 公司提供的 TQ Analyst 定量软件完成每次 PLS 数值实验。记录每次数值实验的参数(平滑点数、因子数等)和结果(每个验证样品的蛋白质含量预测值、RMSEP 等)进行比较。

表 2 采用全谱段一阶导数谱的 PLS 的 RMSEP (%)

Tab. 2 RMSEP of PLS using the first derivative spectra of whole region (%)

	因子数				
	3	4	5	6	7
15	0.507	0.442	0.459	0.615	0.519
17	0.512	0.395	0.494	0.669	0.585
19	0.550	0.398	0.456	0.572	0.553
21	0.573	0.394	0.444	0.608	0.563
23	0.579	0.392	0.393	0.512	0.498
25	0.596	0.418	0.411	0.538	0.551
27	0.607	0.427	0.433	0.546	0.579
29	0.613	0.419	0.446	0.484	0.577
31	0.620	0.412	0.472	0.411	0.525

以全谱段的一阶导数谱为例,表 2 给出了每一个的参数组合(平滑点数、因子数)对应的数值实验结果(RMSEP)。其中,方块内的每一个 RMSEP 值的所在行最左端的数和所在列最上端的数分别是所采用的平滑点数和因子数。表内小方框内的 RMSEP 值,表示它是同一列中的最小值,通过它标明了同一因子数对应的最优平滑点数;表内用黑体和有下横线的 RMSEP 值,表示它是同一行中的最小值,通过它标明了同一平滑点数对应的最优因子数。从表 2 可以看出,每个因子数固定后对应的最优平滑点数并不相同,而每个平滑点数固定后对应的最优因子数也不完全相同。最优参数组合的因子数和平滑点数分别为 4 和 23,最优 RMSEP 值是 0.392(%)。如果先随意固定了某个因子数或平滑点数,就不一定会得到这样的全局最优结果。

由于因子数和平滑点数同时变化寻求最优定标模型的运算量很大,以往的研究很少能够做到这一步,但这是改善提高近红外光谱分析预测能力的一个重要措施,应该给予很好的关注。

表 3 给出了通过全部的数值实验比较得到的 15 个模型对应的最优的平滑点数、因子数和 RM-

SEP 值。可以看出,所有 15 个模型的最优参数组合(平滑点数,因子数)都不相同,5 个波段的最好的模型是:全谱(10 000~4 000 cm^{-1})为采用二阶导数谱、最优因子数、平滑点数和 RMSEP 值为(3,23,0.370);合频波段(5 500~4 000 cm^{-1})采用一阶导数、最优因子数、平滑点数和 RMSEP 值为(7,25,0.564);一倍频波段(7 000~5 500 cm^{-1})采用一阶导数、最优因子数、平滑点数和 RMSEP 值为(5,19,0.357);二倍频(含高频)波段(10 000~7 000 cm^{-1})采用一阶导数、最优因子数、平滑点数和 RMSEP 值为(4,25,0.561);蛋白质基团波段(9 800~9 430 cm^{-1} +7 000~6 370 cm^{-1} +5 100~4 360 cm^{-1})采用原光谱、最优因子数、平滑点数和 RMSEP 值为(7,23,0.490)。

表 3 15 个定标模型的最优 Savitzky-Golay 平滑点数、PLS 因子数和 RMSEP 的比较

Tab. 3 Comparison of number of optimal Savitzky-Golay smoothing points, number of PLS factors and RMSEP in 15 calibration models

		最优因子数	最优平滑点数	最优 RMSEP
全谱	原光谱	7	15	0.550
	一阶导数谱	4	23	0.392
	二阶导数谱	3	23	0.370
合频	原光谱	5	15	0.856
	一阶导数谱	7	25	0.564
	二阶导数谱	6	21	0.705
一倍频	原光谱	6	15	0.608
	一阶导数谱	5	19	0.357
	二阶导数谱	3	23	0.465
二倍频	原光谱	5	15	0.831
	一阶导数谱	4	25	0.561
	二阶导数谱	4	23	0.742
基团	原光谱	7	23	0.490
	一阶导数谱	6	29	0.530
	二阶导数谱	5	29	0.522

最好的定标模型是采用一阶导数预处理后的一倍频波段(7 000~5 500 cm^{-1}),与这个模型的预测效果非常接近的是采用二阶导数预处理后的全谱波段(10 000~4 000 cm^{-1})。这两个模型对应得到的玉米蛋白质含量的预测值和化学值的比较分别在图 3、4 中给出。由图可见,这两个模型

的预测相关系数(R_p)、RMSEP、RRMSEP 分别为 0.945, 0.357, 3.340% 和 0.950, 0.444, 3.461%。因此,模型的预测精度比较高,预测效果很好。特别是采用一倍频波段($7\ 000\sim 5\ 500\text{ cm}^{-1}$)的定标效果完全可以代替全谱波段,这为设计小型便携式的玉米蛋白质测定专用的近红外光谱仪提供了依据。

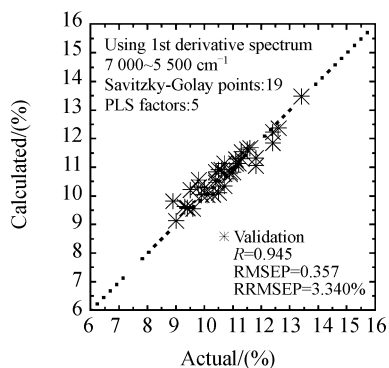


图 3 采用一倍频波段($7\ 000\sim 5\ 500\text{ cm}^{-1}$)一阶导数谱的最优定标模型的玉米蛋白质含量的预测值和化学值的比较

Fig. 3 Comparison of the calculated values with the actual ones for protein concentration of corn samples in optimal calibration model by the first derivative spectra in the first overtone region ($7\ 000\sim 5\ 500\text{ cm}^{-1}$)

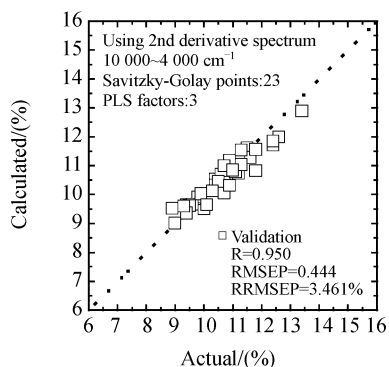


图 4 采用全谱段二阶导数谱的最优定标模型的玉米蛋白质含量的预测值和化学值的比较

Fig. 4 Comparison of the calculated values with the actual ones for protein concentration of corn samples in optimal calibration model by the second derivative spectra for whole region

同时还看到,在原谱平滑化、一阶导数谱和二阶导数谱的预处理方式中,哪一个定标效果好,不能一概而论,而采用多少平滑点数和因子数等,也都要经过多次的数值实验对比来确定。

4 结 论

采用近红外漫反射光谱技术,用傅里叶变换型近红外光谱仪对 177 个玉米样品采集光谱。利用 PLS 建立玉米蛋白质含量的定标模型。首先采用多元散射校正方法对光谱进行预处理,再利用 Savitzky-Golay 平滑化方法来进行原谱、一阶导数谱和二阶导数谱的平滑化处理。选取下列 5 个波段:全谱($10\ 000\text{ cm}^{-1}\sim 4\ 000\text{ cm}^{-1}$)、合频($5\ 500\sim 4\ 000\text{ cm}^{-1}$)、一倍频($7\ 000\sim 5\ 500\text{ cm}^{-1}$)、二倍频(含高频)($10\ 000\sim 7\ 000\text{ cm}^{-1}$)和蛋白质基团($9\ 800\sim 9\ 430\text{ cm}^{-1}+7\ 000\sim 6\ 370\text{ cm}^{-1}+5\ 100\sim 4\ 360\text{ cm}^{-1}$),每个波段都分别采用原光谱、一阶导数谱、二阶导数谱,共建立了 15 个定标模型。(1)通过同时调整 Savitzky-Golay 平滑点数和 PLS 因子数,对每个模型都分别做了 45 次 PLS 数值实验,从中比较选优,得到每个模型的最优平滑点数、因子数和 RMSEP 值,按照 RMSEP 值最小的原则,再从 15 个模型中选优,结果表明,采用一阶导数预处理后的一倍频波段($7\ 000\sim 5\ 500\text{ cm}^{-1}$)的定标效果最好。(2)采用一阶导数预处理后的一倍频波段($7\ 000\sim 5\ 500\text{ cm}^{-1}$)定标建模,相应的预测相关系数、预测均方根偏差、相对预测均方根偏差分别为 0.945, 0.357, 3.340%,模型的预测效果很好。(3)一倍频波段($7\ 000\sim 5\ 500\text{ cm}^{-1}$)完全可以代替全谱波段,并得到更好的定标效果,这为设计小型便携式的玉米蛋白质测定专用的近红外光谱仪提供了指导依据。(4)要获得高精度的定标模型必须考虑波段的不同选取、不同光谱预处理方式的选取、不同平滑点数和不同因子数的选取等多种条件和参数变化下来比较选优,虽然运算量很大,但这是改善提高近红外光谱分析预测能力的一个重要措施,应该给予很好的关注。

参考文献:

[1] BURNS D A, CIURCZAK E W. *Handbook of Near-Infrared Analysis, Second Edition, Revised and Expanded*

- [M]. New York: Marcel Dekker, Inc., 2001.
- [2] SIESLER H W, OZAKI Y, KAWATA S. *Near-infrared Spectroscopy: Principle Instruments and Applications* [M]. Weinheim (Germany): Wiley-VCH, 2002.
- [3] WILLIAMS P, NORRIS K. *Near-infrared Technology in the Agricultural and Food Industries (Second Edition)* [M]. Minnesota (USA): the American Association of Cereal Chemists, Inc. St. Paul, 2001.
- [4] 陆婉珍. 现代近红外光分析技术(第二版)[M]. 北京: 中国石化出版社, 2007.
LU W ZH. *Modern Near Infrared Spectroscopy Analytical Technology (Second Edition)* [M]. Beijing: Chinese Petrochemistry Press, 2007. (in Chinese)
- [5] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
YAN Y L. *Near Infrared Spectroscopy Analysis Fundament and Application* [M]. Beijing: Chinese Light Industry Press, 2005. (in Chinese)
- [6] 陈华才, 吕进, 陈星旦, 等. 基于径向基函数网络的茶多酚总儿茶素近红外光谱检测模型的研究[J]. 光学精密工程, 2006, 14(1): 58-62.
CHEN H C, LU J, CHEN X D, *et al.*. Near infrared spectroscopic model for determinating total catechins in tea polyphenol powder based on radical basis function network[J]. *Opt. Precision Eng.*, 2006, 14(1): 58-62. (in Chinese)
- [7] 张军, 郑咏梅, 王芳荣, 等. 谷物近红外光谱分析中常用数据处理方法讨论[J]. 吉林大学学报(信息科学版), 2003, 21(1): 4-9.
ZHANG J, ZHENG Y M, WANG F Y, *et al.*. Discussion on some regular methods for cereal near infrared spectra analysis[J]. *Journal of Jilin University (Information Science Edition)*, 2003, 21(1): 4-9. (in Chinese)
- [8] 吴军, 白琪林, 苏胜宝, 等. 近红外反射光谱法分析玉米秸秆纤维素含量的研究[J]. 分析化学, 2005, 33(10): 44-49.
WU J, BAI Q L, SU SH B, *et al.*. Near-infrared reflectance spectroscopy analysis of cellulose content in corn stalk [J]. *Chinese Journal of Analytical Chemistry*, 2005, 33(10): 44-49. (in Chinese)
- [9] 梁逸曾, 俞汝勤. 分析化学手册(10)-化学计量学[M]. 北京: 化工出版社, 2001.
LIANG Y Z, YU R Q. *Analytical Chemistry Manual (10)-Chemometrics* [M]. Beijing: Chemical Industry Press, 2001. (in Chinese)
- [10] 赵强, 张工力, 陈星旦. 多元散射校正对近红外光谱分析定标模型的影响[J]. 光学精密工程, 2005, 13(1): 53-58.
ZHAO Q, ZHANG G L, CHEN X D. Effects of multiplicative scatter correction on a calibration model of near infrared spectral analysis[J]. *Opt. Precision Eng.*, 2005, 13(1): 53-58. (in Chinese)
- [11] 高荣强, 范世福, 严衍禄, 等. 近红外光谱的数据预处理研究[J]. 光谱学与光谱分析, 2004, 12(24): 1563-1565.
GAO R Q, FAN SH F, YAN Y L, *et al.*. Preprocessing of near infrared spectroscopic data[J]. *Spectroscopy and Spectral Analysis*, 2004, 12(24): 1563-1565. (in Chinese)
- [12] MICHAEL R WHITBECK. Second derivative infrared spectroscopy[J]. *Applied Spectroscopy*, 1981, 35(1): 93-95.
- [13] NAKANISHI K, HASHIMOTO A, PAN T, *et al.*. Mid-infrared spectroscopic measurement of ionic dissociative materials in metabolic pathway[J]. *Applied Spectroscopy*, 2003, 57(12): 1510-1516.
- [14] PAN T, HASHIMOTO A, KANOU M, *et al.*. Development of a quantification system of ionic dissociative metabolites using an FT-IR/ATR method[J]. *Bioprocess and Biosystems Engineering*, 2003, 26(2): 133-139.
- [15] PAN T, HASHIMOTO A, KANOU M, *et al.* Mid-infrared spectroscopic quantification of ionic dissociative metabolites based on three spectral extraction methods[J]. *Japan Journal of Food Engineering*, 2004, 5(1): 22-31.
- [16] 陈洁梅, 潘涛, 陈星旦. 二阶导数光谱预处理在用 FTIR/ATR 方法定量测定葡萄糖-6-磷酸和果糖-6-磷酸中的应用[J]. 光学精密工程, 2006, 14(1): 1-7.
CHEN J M, PAN T, CHEN X D. Application of second derivative spectrum prepares in quantification measuring glucose-6-phosphate and fructose-6-phosphate using a FTIR/ATR method[J]. *Opt. Precision Eng.*, 2006, 14(1): 1-7. (in Chinese)
- [17] SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by Simplified Least Squares procedures[J]. *Analytical Chemistry*, 1964, 36: 1627-1637.

作者简介: 曹 璞 (1982—), 男, 暨南大学理工学院光电工程系研究生, 从事近红外光谱方面的研究。E-mail: cphnly@126.com

通讯作者: 潘 涛 (1964—), 男, 日本理学博士, 暨南大学理工学院光电工程研究所教授(特聘)、博士生导师。从事应用光谱、生物信息、应用数学等方面的研究。E-mail: tpan@jnu.edu.cn