

基于 Origin 8.0 的马铃薯光谱图像处理

李 鑫,陈争光

(黑龙江八一农垦大学 信息技术学院,黑龙江 大庆 163319)

摘要:在进行近红外光谱分析中,最初得到的光谱图像会因为环境的影响而受到干扰,光谱处理的是否得当,会直接影响到后期建立模型的品质。为消除干扰因素,探索提取有效光谱信息的分析手段,经平均、平滑处理的光谱图像,有效地滤除原始光谱中的毛刺,降低了噪声。对预处理后的光谱进行寻峰、拟合操作,所得结果能够直观地显示峰值、区分吸收峰,拟合后能够保留图像的原始特征,在分析过程中能更好地提取成分含量相关的信息。

关键词:Origin 8.0;近红外光谱;光谱分析;数据处理

中图分类号:O657.33;S532 **文献标识码:**A **文章编号:**1002-2767(2015)02-0132-04 DOI:10.11942/j.issn1002-2767.2015.02.0132

光谱分析技术发展至今,在理论和技术上已经趋于成熟。由于其分析速度快、非破坏性分析、远距测定和实时分析等特点而被广泛应用^[1-2]。近红外光谱在测量过程中由于复杂背景的变化、测量环境的影响,往往最初得到的光谱会有漂移的现象,并且参杂着一定的噪声,对建立高品质的检测模型会有一些的影响,因此需要一个有效的分析手段,消除干扰因素,提取有效光谱信息。

Origin 是美国 OriginLab 公司开发的专业光谱数据分析软件^[3],由于它操作灵活、功能强大,因此被广泛应用于各科研领域。尤其是其数据分析功能,几乎涵盖了光谱分析领域的所有常规分析操作,是科研工作者常用的软件。此外该软件能与各类编程软件无缝链接使其功能扩展到很多应用领域。本研究即是利用 Origin 8.0 软件对马铃薯光谱进行分析处理。

1 近红外光谱数据处理

1.1 平均光谱

光谱图像是由德国 BRUKER 公司生产的 MPA 近红外光谱仪于 16 cm⁻¹波数下对马铃薯样品进行测定获得的。为克服样品粒度差异和样品颗粒分布不均所导致的光谱漂移,近红外光谱仪对每一次装样进行 32 次扫描取平均值作为扫描结果。每个样品装样 3 次,得到 3 组平行扫描结

果(见图 1)。Origin 提供了对重复实验数据求平均值的功能,在“Analysis”菜单下,打开“Mathematics”中的“Average Multiple Curves”。选取要求平均的 3 组光谱数据,在“Method”中选择简单平均法“Average, SD/SE, N”将同一 X 值上的 Y 值做算术平均,得到新的 Y 值,结果见图 2。并且还可以在“Additional Output”中选择输出标准差、标准误差等参数值。

波数	吸光度1	吸光度2	吸光度3
12473.92708	0.07065	0.07088	0.07293
12466.21284	0.07122	0.07138	0.07417
12458.4986	0.07193	0.07185	0.07482
12450.78436	0.07188	0.07218	0.07451
12443.07012	0.07142	0.072	0.07396
12435.35587	0.07122	0.07185	0.07389
12427.64163	0.07099	0.07219	0.07437

3664.26429	1.67532	1.65619	1.6661
3656.55005	1.83177	1.80204	1.82081
3648.83581	1.97899	1.94493	1.94773
3641.12157	2.09894	2.08999	2.06359
3633.40733	2.14287	2.17764	2.15799
3625.69309	2.09857	2.15812	2.1619
3617.97885	2.04559	2.11356	2.11929
3610.26461	1.97039	2.04202	2.04859
3602.55037	1.9179	2.01699	2.03174
3594.83613	1.85122	1.96978	1.9988

图 1 相同波数下三组光谱参数

Fig. 1 Three groups of spectrum parameter under the same wave

1.2 光谱平滑处理

光谱图像主要反映物质的化学组成和各组分的含量,同时也受环境中各种各样的因素(粒度、粘度、密度和表面纹理等物理因素)干扰,因此最初获得光谱总是伴有部分噪声,即光谱图像会有

收稿日期:2014-10-30

基金项目:黑龙江省教育厅科研资助项目(12521370)

第一作者简介:李鑫(1990-),男,河南省淇县人,硕士,从事农产品品质分析研究。E-mail:407676696@qq.com。

通讯作者:陈争光(1973-),男,湖北省黄陂县人,博士,副教授,硕士生导师,从事农产品检测方面研究。E-mail:ruzee@sina.com。

大量的毛刺。因此,需要借助数学方法将毛刺滤除掉,得到相对平滑的光谱图像。通过光谱平滑处理可以有效消除背景噪声和物理因素的干扰,提高光谱与化学参数的相关性,有助于后期模型的建立,提高模型的稳定性和可靠性。具体操作过程是在“Analysis”菜单下“Signal Processing”信号处理中选择“Smoothing”平滑处理。平滑处理是消除光谱噪声的常用方法,Origin 软件提供给 4 种方法进行平滑。

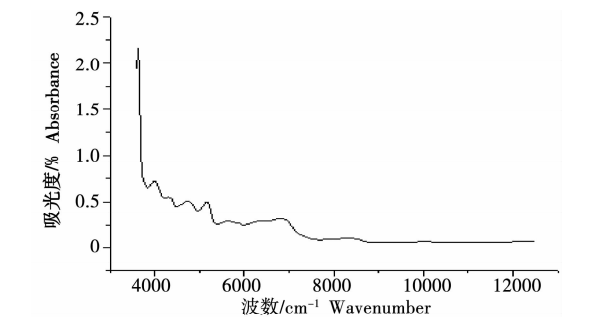


图 2 平均光谱图像
Fig. 2 Average spectral image

- (1)“Adjacent-Averaging”相邻平均法,对局部数据计算相邻几个数据的算术平均值来实现曲线的平滑;
- (2)“Savitzky-Golay”卷积平滑法,对局部数据进行多项式回归来实现曲线的平滑;
- (3)“Percentile Filter”分位数滤波法,对局部数据计算一个指定的分位值,然后将原始数据替换为这个分位值来实现曲线的平滑;
- (4)“FFT Filter”FFT 滤波,是基于快速傅里叶变换的低通滤波算法,通过滤除高频信号来实现曲线平滑。

为了能够有效地保留数据原始特征,首选卷积平滑法。首先在“Input”中选定需要平滑的数据,在“Method”中选用“Savitzky-Golay”(卷积平滑)。同时在选择平滑点数时要注意,点数选择过

少,平滑效果不明显,平滑点数过多会导致失真,应勾选“Auto Preview”选项,实时观测处理效果,经比较选用 9 点平滑的光谱预处理效果较好^[4-5],其原理是从图像左边开始,对每 9 个数据点重复拟合,每进行一次就剔除第一个数据点然后加上右边一个数据点,直到加到最后一个数据点为止。Origin 通过 Savitzky-Golay 法的卷积系数表^[6](见表 1)可以计算出多项式系数的大小。

例如,用一个三次多项式 $f_i = b_{n0} + b_{n1}i + b_{n2}i^2 + b_{n3}i^3$ 来拟合 9 个点,当 X 轴上每一个点代入曲线方程后, f_i 值与该点纵坐标之差的平方和最小时拟合度最高。根据卷积系数表(见表 1)求 b_{n0} ,找到 9 个点对应的一系列系数。

表 1 Savitzky-Golay 卷积系数表

Table 1 Convolution coefficient of Savitzky-Golay

POINTS	13	11	9	7	5
-6	-11				
-5	0	-36			
-4	9	9	-21		
-3	16	44	14	-2	
-2	21	69	39	3	-3
-1	24	84	54	6	12
0	25	89	59	7	17
1	24	84	54	6	12
2	21	69	39	3	-3
3	16	44	14	-2	
4	9	9	-21		
5	0	-36			
6	-11				
NORM	143	429	231	21	35

设多项式拟合的 9 个数据值大小分别为 $y_{-4}, y_{-3}, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, y_4$

则有

$$a_{30} = \frac{-21y_{-4} + 14y_{-3} + 39y_{-2} + 54y_{-1} + 59y_0 + 54y_1 + 39y_2 + 14y_3 - 21y_4}{231} \quad b_{30} = \frac{a_{30}}{0!} \tag{1}$$

根据以上方法参考卷积系数表可以分别求出 a_{31}, a_{32}, a_{33} 的大小。

又

$$b_{n1} = \frac{a_{n1}}{1!}, b_{n2} = \frac{a_{n2}}{2!}, b_{n3} = \frac{a_{n3}}{3!} \tag{2}$$

则可求得 b_{n1}, b_{n2}, b_{n3} 的值。

1.3 光谱的峰值标注

物质受到近红外光的照射,分子中的各基团就会发生震动,光谱图中的峰值就反映了基团的震动形式,基团都有自己特定的红外吸收区域,分子其它部分对其吸收位置影响较小。通常把这种能够代表基团存在并有较高强度的吸收谱带称为

基团频率,其所在的位置一般又称为特征吸收峰。因此,图谱中波峰的分析处理,是对物质进行定量或定性分析的关键。Origin 针对波峰处理是在“Analysis”菜单下的“Spectroscopy”光谱分析选项,在二级子菜单中选择“Baseline and Peaks”选项,首先在“Input”中指定需要分析的数据,然后根据操作向导完成后面一系列的操作。包括创建基线、扣除基线、寻峰和峰值积分四部分功能。由于光谱测量时的背景已经被扣除了,所以在“Method”基线的设置方式中选择“None”。点击“Next”直接进入寻峰功能的页面,这里提供的三种寻峰方法极大值法、窗口搜索法和一阶导数法。每种寻峰方法都有各自的特点,根据实际情况作相应的选择,这里选择一阶导数法,“Smooth Points”平滑点数量设置为 9(同平滑处理时设置的点数),“Direction”峰向为“Both”,再分别设置峰值参数和显示信息,点击下方的“Apply”按钮,可以实时浏览处理结果。然后进入到峰值积分的操作页面,方法包含全面积积分和峰值面积积分两种,其中,全面积积分可以得到累加的峰面积,而峰值面积积分可以得到每个峰各自的峰面积。选择峰值面积积分,点击“Finish”完成设定,得到积分后的图谱。积分结果在原工作表中以新表单的方式给出。为了更加直观、清晰的浏览图谱,可以进一步设置布局。在图谱中峰值的数值部位单击右键,选择“Plot Details”,将“Rotate”的值调整成 90,“Y Offset”为-100,“Size”为 18,“Justify”为 Right,点击“OK”,完成设置,结果见图 3。

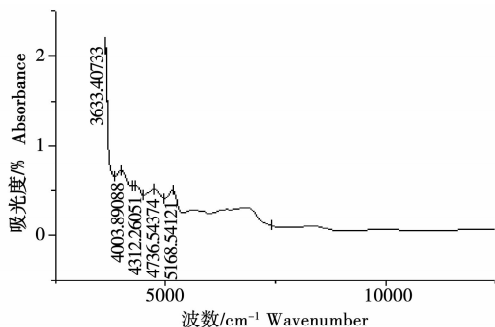


图 3 光谱峰值标注

Fig. 3 Spectral peak annotation

1.4 光谱曲线拟合

Origin 的曲线拟合功能分为线性拟合与非线性拟合^[7-11],提供的拟合方法很多,基本满足了一般性数据拟合中的绝大多数情形。在光谱分析

中,对不同形状的分析曲线要有相应的数学表达式描述,这就需要一个合适的拟合方法进行处理。在“Analysis”菜单下的“Fitting”中选择“Nonlinear Curve Fit”来实现拟合。打开对话框,首先在“Data Selection”菜单中的“InputData”指定需要拟合的数据,“Weights”选项根据实际情况进行选择,这里选用“No weights”。下方的显示窗口可以看见数据的散点图,并且在权函数中选择不使用权函数。在“Function Selection”这一栏中进行函数类型和方法的设置,在“Category”函数类型中以形式和领域划分,一共有二百多种的函数类型供选择,这里选择“Spectroscopy”光谱学领域,在“Function”方法选用“Lorentz”洛伦兹函数,拟合函数如式(3)所示。此时下方的显示窗口会根据初始参数值绘制一条红色的函数图像。在下方显示窗口中点击“Formula”会显示当前调用函数的公式。打开“Parameters”参数页面,此页面列出了当前拟合函数的所有参数及属性,当前页面会给出参数的初始值,如果参数的初始值给定的不合理,拟合的结果就会显示“Fit did not converge”拟合不收敛的提示,此时需要对初始值重新调整,并固定 y_0 和 w 两个参数,这样可以提高运算速率,加速迭代的收敛,这时拟合的结果提示收敛,结果见图 4。

$$y = y_0 + \frac{2A}{\pi} \frac{W}{4(x - x_c)^2 + W^2} \quad (3)$$

式中: y_0 为偏移量,取值为 0; x_c 为波峰所对应的 x ,取值为 5; W 为宽度,取值为 2; A 为峰面积,取值为 1。

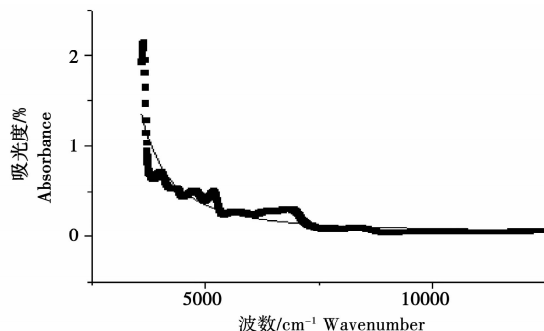


图 4 光谱图像拟合后效果图

Fig. 4 Spectral image rendering after fitting

1.5 光谱分析模板的建立

近红外光谱分析研究需要处理大量的光谱图像,相同物质的光谱研究,所用处理方法也是一致

的。为了避免每条光谱都进行单独的操作,应建立光谱的分析模板。分析模板是完成若干项数据操作的一个流程,建立分析模板实际上是保留这个操作流程,这样在进行大量的重复性分析时,可直接调用分析模板,节省时间,提高分析效率。首先对光谱进行拟合,这里需要注意的是拟合过程中,将“Recalculate”重算模式,改为“Auto”。此时,如果将新的数据拖入表单,会自动触发拟合过程,这样就建立了一个完整的数据分析流程。在建立分析模板时,首先清空表单中现有的数据,然后点击“File”菜单,选择“Save Template As”,弹出的对话框中“Category”代表保存的模板类,包括内置模板和自定义模板,“Template Name”为模板名称,在“Option”中可以设置存放路径和模板说明。点击“OK”完成整个操作。当要使用该模板时,在“File”菜单下的“Recent Books”可以进行调用。充分利用这项功能,可以减少数据处理的工作量,并且可以保证数据处理条件的一致性。

2 结论

应用 Origin 8.0 对马铃薯光谱数据进行处理,能够有效地滤除环境因素干扰,可以更加准确、直观地反映马铃薯光谱数据的特征属性。它的分析模板功能能够批处理大量的光谱数据,省时省力,处理结果以表格、图像的形式呈现,更具科学性。出色的绘图功能有助于提高撰写论文的规范性,提高学术水平,为今后从事科研工作奠定了良好的基础。

参考文献:

- [1] 李宗朋,王健,宋全厚,等.近红外光谱技术在食品检测与质量控制中的应用[J].食品与发酵工业,2012,38(8):125-131.
- [2] Ainara López, Silvia Arazuri, Carmen Jarén, et al. Crude Protein content determination of potatoes by NIRS technology[J]. Procedia Technology, 2013, 8(8): 488-492.
- [3] 李宝惠,平梅. Origin 7.0 在科技绘图、数据处理方面的应用[J]. 临沂师范学院学报, 2005, 27(3): 124-128.
- [4] 尼珍,胡昌勤,冯芳近. 红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志, 2008(5): 824-829.
- [5] Jiang B, Li W, Huang Y D. Influence of pretreatment methods of the spectra on the calibration model of the precuring-degree[J]. Journal of Applied Polymer Science, 2012, 124(2): 1529-1533.
- [6] Savitzky A, Golay M J. Smoothing and differentiation of data by simplified least squares procedures[J]. Analytical Chemistry, 1964, 36(7): 1627-1639.
- [7] 邓晓敏,张军朋,吴先球. 利用 Origin 确定实验中非线性函数的曲线关系[J]. 大学物理实验, 2011, 24(1): 73-76.
- [8] 陈旭红. 用 Origin 软件的线性拟合和非线性曲线拟合功能处理实验数据[J]. 江苏技术师范学院学报, 2006, 12(6): 85-89.
- [9] 易均辉,莫惠媚,龚福忠,等. 巧用 Origin 软件处理物理化学实验数据[J]. 实验室研究与探索, 2009, 18(8): 56-59.
- [10] 武新,张永胜. Origin 在曲线拟合中的应用[J]. 计算机工程与应用, 2005, 41(17): 206-208.
- [11] 王宇,胡赞,曹仲义,等. Origin 软件在化工原理实验数据拟合中的应用[J]. 试验技术与管理, 2010, 27(1): 86-88, 93.

Potato Spectral Image Processing Based on Origin 8.0

LI Xin, CHEN Zheng-guang

(College of Information Technology, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319)

Abstract: In near infrared spectral analysis, spectral image originally obtained would be interference because of the environmental impact, proper spectral processing, can directly affect the quality of the late model building. In order to explore the effective analysis method of spectrum information extraction for eliminating interference factors, the spectral image averaging and smoothing could effectively filter out the burr in the original spectra and reduced the noises. For peak, fitting operation on the spectrum after pretreatment, the results showed the peak, distinguish the absorption peak of intuitive, original feature fitting could preserve the image information content, extract components could be better in the process of related analysis.

Keywords: Origin 8.0; near infrared spectroscopy; spectral analysis; data processing