第 28卷第 5期
2009年 10月

红 外 与 毫 米 波 学 报
J. Infrared Millim. Waves

Vol. 28, No. 5
October, 2009

文章编号: 1001 - 9014(2009)05 - 0371 - 05

基于小波变换的苹果汁多光程近红外光谱信息提取研究

朱大洲[1,2], 籍保平[1], 史波林[3], 潘立刚[2], 屠振华[1], 孟超英[4], 庆兆珅[1]

(1. 中国农业大学 食品科学与营养工程学院,北京 100083;
2. 北京农产品质量检测与农田环境监测技术研究中心,北京 100097;
3. 上海大学 通信与信息工程学院,上海 200072;
4. 中国农业大学 信息与电气工程学院,北京 100083)

摘要:利用浸入式光纤采集鲜榨苹果汁分别在 5mm、10mm、15mm和 20mm光程下的透/反射近红外光谱,实现对苹果汁中糖度(可溶性固形物, SSC)和酸度(pH值)的定量预测. 结果表明, SSC和 pH具有不同的最佳光程长,分别为 5mm和 20mm. 为了兼顾各待测量对象的浓度范围和各组分的最佳光程长,从而提高模型的性能,采用多光程光谱混合建模,研究了多光程光谱信息的提取方法. 采用原始光谱直接展开所建的模型虽然能有效利用多光程光谱的信息,但增加了模型复杂度,致使建模时间增长. 因此,提出了两种基于小波变换的信息提取方法,它们在高效提取多光程信息的同时,能显著缩短建模时间并简化模型. 其中基于展开光谱的小波近似系数建立的模型性能最优, SSC和 pH值模型的 SECV值分别达到 0. 4761°Brix和 0. 0779.

关 键 词:近红外;光程;信息提取;小波变换;苹果汁
中图分类号: O657. 33, TS255. 44 **文献标识码**:

INFORMATION EXTRACTION OF MULTI-OPTICAL-PATH NIR SPECTRA FOR APPLE JUICE BASED ON WAVELET TRANSFORMATION

ZHU Da-Zhou[1,2], JI Bao-Ping[1], SHI Bo-Lin[3], PAN Li-Gang[2], TU Zhen-Hua[1], MENG Chao-Ying[4], QING Zhao-Shen[1]

(1. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083;
2. Beijing Research Center for Agri-food Testing and Farmland Monitoring, Beijing 100097, China;
3. School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;
4. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: The transmittance/reflectance NIR spectra of fresh apple juice were collected by immersed fiber at four optical path of 5mm, 10mm, 15mm, and 20mm. The NIR spectra were used to predict quantitatively the soluble solid contents (SSC) and pH value of apple juice. The results show that the optimal optical path lengths for SSC and pH are different, which are 5mm and 20mm, respectively. In order to consider the concentration ranges of different components and their optimal path lengths, multi-optical path spectra were used to construct the model to improve the model quality. The method for extracting information from multi-optical-path spectra was studied. The method based on directly unfolded spectra could effectively use the spectral information, but the model became more complex and the calculation time was very long. Two methods based on wavelet transformation were proposed to extract the information. They can extract the multi-optical path spectra information effectively, and at the same time the calculation time decreases significantly and the model is simplified. One of the methods that uses the approximate wavelet coefficient of the unfolded spectra to construct model has the best performance: SECV for SSC and pH value are 0. 4761 °Brix and 0. 0779, respectively.

Key words: NIR; optical path; information extraction; wavelet transform; apple juice

收稿日期: 2008 - 10 - 27,**修回日期**: 2009 - 06 - 10 **Received date:** 2008 - 10 - 27, **revised date:** 2009 - 06 - 10
基金项目:国家"十一五"科技支撑计划资助项目(2006BAD05A06 - Z1);北京市科技计划项目(Z09090001040000)
作者简介:朱大洲(1981-),男,湖北宜昌人,博士,主要从事食品无损检测研究.

引言

近红外光谱进行液体定量检测的基础是比尔定律[1],即:$A=\varepsilon * L * C$,其中 A 为吸光度,ε 为待测组分的吸光系数,L 为光程长,C 为待测组分的浓度.比尔定律是针对单色光真溶液检测而言的.对于待测对象,各波长对应的 ε 不同,光程 L 对 A 的影响很大.吸光度与光程成正比,因此增大光程可提高光谱对低浓度成分的检测灵敏度,但增大光程的同时也增强了噪声.于海燕等人的研究表明光程对黄酒金属元素的近红外检测精度具有很大影响[2].特别是对于在线液体测量,光程成为影响其测量误差的重要因素.因此,光程的选择至关重要[3].

根据分光误差理论,单波长测量中存在最佳吸光度,如 Willard等人[4]提出了忽略参考光谱噪声下的最佳吸光度为 0.434,Howard等人[5]提出了考虑参考光谱噪声下的最佳吸光度为 0.4816.一般认为,吸光度的较佳范围为 0.3～0.7,吸光度太小,灵敏度较低,吸光度太大,噪声信号也较大.根据比尔定律,某待测对象的浓度范围是可知的,而吸光系数 ε 一般不变,因此只有选择最佳的光程长来达到最佳吸光度,才能提高检测准确度.早期研究主要集中于如何选择最佳的吸光度及相应的光程长[6,7],但研究发现,实际测量中很难实现最佳精确光程长,可根据光程与灵敏度之间的定量关系来大概选择光程[8].

根据比尔定律可知,由于不同待测量浓度范围不同,最佳光程长也不同,但近红外测量是根据同一光谱计算多个成分的含量.因此,有人提出采用双光程组合和多光程建模[9,10],从而兼顾多个成分的最佳光程长.但如何提取多光程光谱的信息,尚缺乏行之有效的方法.

小波变换作为一种"显微镜"式的信息分析技术,已经被广泛用于近红外光谱的去噪、数据压缩及模型传递[11].在小波域,信号被分解成频率不同的小波系数,从而可以区分有用信号、高频噪声及低频基线漂移.因此,小波变换是一种有力的信息提取方法.

近红外光谱已广泛应用于苹果汁的检测中,Segtnan等人[6]对不同测量模式下果汁近红外检测的光程进行了研究,研究结果表明对于有散射特性的果汁透反射优于透射.本研究以苹果汁为研究对象,在不同光程下采集其透反射近红外光谱,探讨光程对苹果汁透反射光谱的影响,并应用多光程光谱来建立模型,探讨应用小波变换同时提取多光程光谱信息的方法.

1 材料与方法

1.1 样品准备

从水果市场采集 28个嘎拉、42个美国加利果.洗净后用气压裹包榨汁机(AF2000,东兴食品机械有限公司)榨汁,然后在 3600r/min的速度下离心 10min(TD-5型离心机,上海安亭科学仪器厂).滤去悬浮的上层杂质将澄清液在 -18℃条件下贮存 3个月,空气自然解冻后再过滤取澄清液进行测量.

1.2 光谱采集及参考值的测定

用 CCD近红外光谱仪(AvaSpec-2048, arantes, The Netherlands)自带的可变光程的透射式浸入型光纤探头采集苹果汁的透反射光谱.光程长分别设为 5mm、10mm、15mm、20mm.光源为石英卤素灯 12W/12V,检测器为 2048像元线性阵列 CCD检测器,光谱范围 580～1100nm,CCD积分时间为 2ms,扫描次数为 100.以空气作参比.可溶性固形物(SSC)的测定采用阿贝折光仪(WYA型,上海精密科学仪器有限公司),参照国标 GB/T 12143.1的规定执行.pH值采用 pH计(PHS-25型,上海雷磁仪器厂)测定.

1.3 小波变换及信息提取的基本原理

有关小波变换的详细原理可参考文献[12],下面只作简单的介绍.设函数 $\Psi(t)\in L^2(R)$ 是小波函数,$\Phi(t)$ 是对应的尺度函数,则信号 $f(t)\in L^2(R)$ 的连续小波变换为:

$$WT_f(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty}f(t)\Psi\left(\frac{t-\tau}{a}\right)dt=f\langle,\Psi_{a,\tau}\rangle \quad , \quad (1)$$

式中,a 和 τ 分别是尺度参数和平移参数.对 $\Psi(t)$ 和 $\Phi(t)$ 按 $a=2^j$,$\tau=k2^j$ 离散化,得到二进离散小波函数族 $\Psi_{j,k}(t)$ 和尺度函数族 $\Phi_{j,k}(t)$,它们构成希尔伯特空间的完备基,a 可简称为尺度 j $\Phi_{j,k}(t)$ 描述 $f(t)$ 大于等于尺度 j 范围内的轮廓特征,即信号的低频特征;$\Psi_{j,k}(t)$ 描述 $f(t)$ 在尺度 j 下的细节特征,即信号在某一高频段的特征,尺度越小反映的信号频率越高.这样就可用 $\Psi_{j,k}(t)$ 和 $\Phi_{j,k}(t)$ 按照尺度来分割原始信号,也即对原始信号作小波级数展开:

$$f(t)=\sum_{k\in Z}f\langle,\Phi_{j,k}(t)\rangle\Phi_{j,k}(t)+\sum_{j=1}^{J}\sum_{k\in Z}f\langle,\Psi_{j,k}(t)\Psi_{j,k}(t)\rangle$$

$$=\sum_{k\in Z}c_{J,k}\Phi_{J,k}(t)+\sum_{j=1}^{J}\sum_{k\in Z}d_{j,k}\Psi_{j,k}(t) \quad , \quad (2)$$

公式(2)又叫重构公式,其中 J 为任意设定的最大分解尺度,$c_{J,k}$ 与 $d_{j,k}$ 分别代表低频信号的近似系数和高频分量的细节系数.根据需要对近似系数和细节系数进行处理,就可以从原始信号中提取出我们感兴趣的频率信息供进一步分析.

1.4 数据处理

异常样本的存在对模型有很大影响,因此在建立各模型之前,综合利用外在学生残差-杠杆值图和COOK值-杠杆值图来剔除各光程下的异常样本,用剩余的57个样本进行建模.将不同光程下的近红外光谱按波长点依次展开,得到展开光谱,采用离散小波变换对展开光谱进行分解,得到近似系数和细节系数.由于近似系数主要反映了近红外光谱的基本信息,而细节系数主要反映了光谱的细微信息及噪声信息,因此采用近似系数作为多光程光谱的有效信息和偏最小二乘法(PLS)的输入,参与建立近红外光谱与苹果汁品质参数之间的校正模型,最佳主因子数(LV)由舍一交互验证确定.小波变换及PLS利用wavelet toolbox 3.0及自编的matlab程序在Matlab2006a下完成.采用校正相关系数 r、交互验证标准差SECV及相对分析误差RPD(SD/SECV)来评价模型效果.

2 结果与分析

2.1 不同光程下的苹果汁近红外光谱及品质参数特征

经异常样本剔出后,苹果汁的SSC和pH值统计参数见表1.在模型计算时去掉了噪声较大及CCD响应较弱的波段,选取780～1000nm波段进行分析.图1为苹果汁分别在5mm、10mm、15mm和20mm光程下的平均光谱.可见,随着光程的增加,光谱的吸光度值逐渐变大.光谱的形状基本相似,但并不是简单的纵向平移,这可能是由于果汁光谱存在非线性的特征.比尔定律要求测量对象为透明的真溶液,并且入射光为平行光.而苹果汁存在散射、折射[6],光程越长,散射越严重,因此不同光程下等效吸收系数不同,同一波长不同光程长处测得的苹果汁光谱也不是严格的线性增长关系,而存在一定的非线性信息.在970nm附近,有一个明显的吸收峰,为O-H的3倍频振动,它反映了苹果汁中糖及水分的信息,由于pH值表征的是游离态氢的浓度,因此,该峰为pH值的预测奠定了基础.

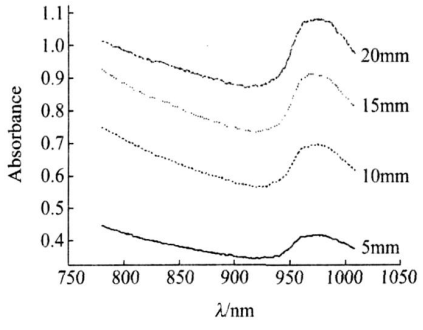

图1 苹果汁在四种不同光程下的平均光谱

Fig 1 The mean spectra of apple juice at four different optical path lengths

表1 苹果汁参考值的统计特征

Table 1 The statistic characteristic of the reference values of apple juice

	平均值	范围	标准偏差
可溶性固形物(°Brix)	11.86	9.1～14.0	1.15
pH值	4.05	3.84～4.36	0.14

2.2 单一光程分别建模

利用5mm、10mm、15mm和20mm光程下的近红外漫透射光谱分别建立SSC和pH值的校正模型,结果见表2.从表2可以看出,采用5mm光程所建的SSC模型效果最好,而采用20mm光程所建的pH模型效果最好,说明苹果汁中的不同成分对应着不同的最佳光程长.若能针对每个成分,确定其最佳光程长,再扫描光谱进行建模,测量准确度当然最好.但近红外测量中一般根据同一光谱同时计算多个成分,另外,单一成分的最佳光程长也很难准确确定.因此,根据经验或经简单预实验所确定的单一光程建模,很难使多个成分的模型都达到最优.

2.3 光谱直接展开

多光程相对于单一光程来讲,能提供更多的光谱信息,其中不同光程下与浓度相关的光谱信息呈线性关系,它们是由化学因素引起的,而由于果汁散射、光源不是严格的平行光等所有杂散光因素引起的光谱信息呈非线性,则是由物理因素引起的.光程 L 太小,所得吸光度 A 就小,灵敏度就低;光程太大,散射严重,测量误差也较大.因此只有将多光程光谱混合在一起建立校正模型,才能充分利用各光程下的信息,更加接近各待测量的最佳光程长,从而提高模型性能.另外,各波长处物理因素的影响趋势也不一致,近红外多元校正方法本身能一定程度上校正物理干扰,采用合适的预处理方法,能更进一步消除散射的影响.

表 2 单一光程近红外光谱所建立的 SSC和 pH值模型

Table 2 The SSC and pH model constructed by NIR spectra with single pathlength

光程	可溶性固形物 主因子数 (LV)	相关系数 (r)	交互验证标准差 (SECV)	相对分析误差 (RPD)	建模时间 (Time (s))	pH值 主因子数 (LV)	相关系数 (r)	交互验证标准差 (SECV)	相对分析误差 (RPD)	建模时间 (Time (s))
5mm	3	0.8607	0.6883	1.6706	4.6	3	0.7352	0.1105	1.2405	4.7
10mm	3	0.7951	0.7747	1.4841	4.8	4	0.8304	0.1192	1.1501	4.7
15mm	4	0.8638	0.7891	1.4570	4.7	4	0.8236	0.1064	1.2890	4.5
20mm	4	0.7336	0.8868	1.2965	4.8	5	0.9117	0.0832	1.6477	4.7

最简单的多光程建模方法是将不同光程扫描得到的光谱按波长展开并进行纵向堆砌 (称之为“展开光谱”),然后利用 PLS建立展开光谱与参考值之间的校正模型. 苹果汁在 5mm、10mm、15mm、20mm光程下的展开光谱见图 2(a),可见随着光程的增加,吸光度值增加,噪声也逐渐变大. 采用光谱直接展开法所建立的模型结果见表 3,表 3也列出了采用单一光程所建立的最优模型. 可见光谱直接展开所建模型明显优于单一光程的建模效果, SECV显著下降, SSC和 pH模型的 RPD值分别达到了 2.4178和 1.7156,说明采用多光程混合建模能提高模型的稳健性.

2.4 各光程光谱的小波系数展开

虽然采用展开光谱直接建模能增加模型的稳健性,但由于展开后的光谱其变量数成倍增加,模型的复杂度增加,且计算量迅速膨胀,建模时间由单一光程所需的 4s多迅速增加到 18s以上 (表 3),对于数据量很大的傅里叶近红外光谱,采用展开光谱直接建模几乎不可能. 因此如何高效提取多光程光谱所含的有效信息,是采用多光程建模的核心. 本研究尝试采用小波变换来提取多光程光谱的信息. 首先对不同光程下的光谱分别进行离散小波变换,提取其近似系数,再将不同光程对应的近似系数展开,用

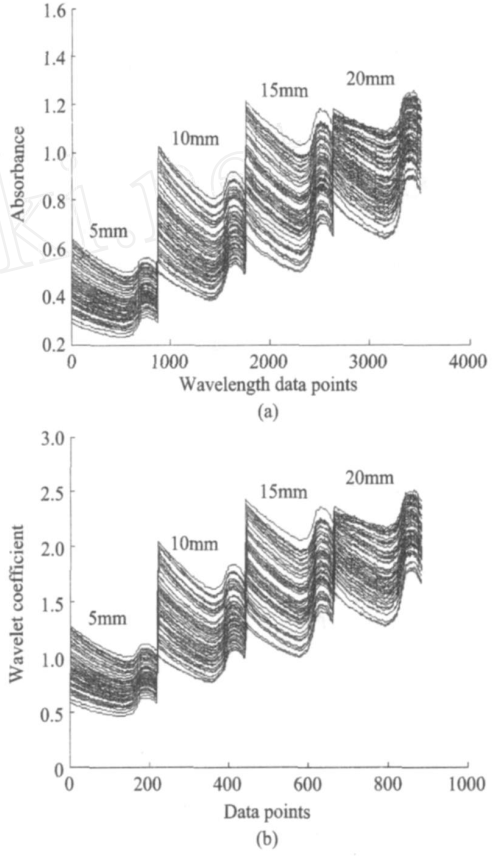

图 2 (a)四种光程下的苹果汁近红外展开光谱 (b)展开光谱的小波系数图

Fig 2 (a) The unfolded NIR spectra of apple juice with four path lengths (b) the plot of wavelet coefficient for unfolded NIR spectra

表 3 多光程光谱所建立的 SSC和 pH值模型

Table 3 The SSC and pH model constructed by NIR spectra with multi-path lengths

	信息提取方法	主因子数 (LV)	相关系数 (r)	交互验证标准差 (SECV)	相对分析误差 (RPD)	建模时间 (Time (s))
可溶性固形物 SSC	单一最佳光程	3	0.8607	0.6883	1.6707	4.6
	光谱直接展开	9	0.9661	0.4755	2.4178	19.0
	小波系数展开	9	0.9652	0.4803	2.3938	5.1
	展开光谱的小波系数	9	0.9656	0.4761	2.4148	4.7
pH值	单一最佳光程	5	0.9117	0.0832	1.6477	4.8
	光谱直接展开	10	0.9849	0.0799	1.7156	19.1
	小波系数展开	10	0.9834	0.0806	1.7009	5.1
	展开光谱的小波系数	10	0.9842	0.0799	1.7152	4.8

PLS建立近似系数和参考值之间的校正模型. 若对原始光谱进行 j次小波分解,则提取出的近似系数其数据长度为原来的 2^{-j},因此用于建模的变量数显著减少. 在小波变换过程中,如何选择合适的小波基、确定合适的分解层数是很关键的,需要根据检验反复实验,直至找到最佳的小波变换参数. 本研究采用 db2小波基进行 2层分解,对各光程光谱进行小波变换后,其小波系数的展开谱图所建立的模型见表 3. 可见相对于原始光谱直接展开,小波系数展开

的建模时间显著减少,达到 5s左右,但建模效果与原始光谱直接展开相比要稍差一些,说明在利用小波变换提取多光程光谱的信息时,有部分信息损失.

2.5 展开光谱的小波系数

由于对各光谱提取的小波系数展开后会带来部分信息损失,因此本研究提出第二种信息提取策略.即先将不同光程下的光谱展开,再对展开后的光谱进行小波变换,用提取的展开光谱的小波系数建模.图 2(b)为展开光谱的小波系数图,可见小波系数图与原始光谱图 (图 2(a))形状非常相似,保留了原始光谱的有效信息.小波系数图的纵坐标范围比原始光谱中吸光度的范围更大,更有利于区分不同样本的特征.另外,小波系数图的数据点显著减少,从 3600多点减少至 800多点,与单一光程对应光谱的数据点数一样,这有利于模型的简化.本研究采用 bior1.3小波基对展开光谱进行 2层分解,利用其近似系数建立的 PLS模型见表 3,其 SECV和 RPD值与展开光谱所建模型基本相同,但建模所需时间明显下降,与采用单一光程所需时间基本相同,说明提取小波系数的时间很短,显著提高了模型性能.

3 结论

在应用近红外光谱检测苹果汁的 SSC和 pH值时,两者的最佳光程长不同.相对于单一光程下所建的模型,采用多个光程下的光谱建模可以兼顾各待测量的浓度范围,兼顾各组分的最佳光程长,从而得到更优的模型性能.提出了三种多光程光谱信息的提取方法.其中采用原始光谱直接展开所建的模型虽然能有效利用多光程光谱的信息,但模型复杂度增加,建模时间很长.提出的两种基于小波变换的信息提取方法,它们在高效提取多光程信息的同时,能显著缩短建模时间并简化模型.基于展开光谱的小波近似系数建立的模型性能最优,SSC和 pH值模型的 SECV值分别达到 0.4761°Brix和 0.0779.

REFERENCES

[1] YAN Yan-Lu, ZHAO Long-Lian, HAN Dong-Hai, *et al*. *Elements and Application of Near-Infrared Spectra Analysis* [M]. Beijing: China Light Industry Press(严衍禄,赵龙莲,韩东海,等.近红外光谱分析基础与应用.北京:中国轻工业出版社), 2005: 10—12.

[2] YU Hai-Yan, YING Yi-Bin, XIE Li-Juan, *et al*. Influence of optical path length on NIR analysis results for trace metal determination in chinese rice wine [J]. *Spectroscopy and Spectral Analysis*(于海燕,应义斌,谢丽娟,等.光程对黄酒金属元素近红外透射光谱分析精度的影响.光谱学与光谱分析), 2007, **27**(6): 1118—1120.

[3] Jensen P S, Bak J. Near-infrared transmission spectroscopy of aqueous solutions: Influence of optical pathlength on signal-to-noise ratio [J]. *Applied Spectroscopy*, 2002, **56**(12): 1600—1606.

[4] Willard H, Merrit L, Dean J. *Instrumental Methods of Analysis* [M]. New York: van Nostrand Co., 1983: 73—75.

[5] Howard L, Mark peter, Grifiths R. Analysis of noise in Fourier transform infrared spectra [J]. *Applied spectroscopy*, 2002, **56**(5): 633—639.

[6] Segtnan V H, Isaksson T. Evaluating near infrared techniques for quantitative analysis of carbohydrates in fruit juice model systems [J]. *Journal of Near Infrared Spectroscopy*, 2000, **8**: 109—116.

[7] Hazen K H, Arnold M A, Small G W. Measurement of glucose in water with first-overtone near-infrared spectra [J]. *Applied Spectroscopy*, 1998, **52**(12): 1597—1603.

[8] WANG Yan, LU Yan-Hui, WANG Rui, *et al*. Influence of pathlength on the error of spectral measurement [J]. *Chinese Journal of Tianjin University*(汪嘵,卢延辉,王蕊,等.光程长对光谱测量误差的影响.天津大学学报), 2004, **37**(10): 906—909.

[9] HE Jin-Cheng, YANG Xiang-Long, WANG Li-Ren. Pathlength selection of determing the chemical oxygen demand (COD) in wastewater by using near-infrared transmission spectra [J]. *J. Infrared Millim. Waves*(何金成,杨祥龙,王立人.近红外光谱透射法测量废水化学需氧量 (COD)的光程选择.红外与毫米波学报), 2007, **26**(4): 318—320.

[10] LI Gang, LIU Yu-Liang, LIN Ling, *et al*. Application of multi-optical path length modeling in the quantitative analysis of human whole blood [J]. Chinese Journal of Analytical Chemistry (李刚,刘玉良,林凌,等.采用多光程长建模方法检测血液成分含量.分析化学), 2007, **35**(10): 1495—1498.

[11] TIAN Gao-You, YUAN Hong-Fu, LIU Hui-Yin, *et al*. The application of wavelet transform in near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*(田高友,袁洪福,刘慧颖,等.小波变换在近红外光谱分析中的应用进展.光谱学与光谱分析), 2003, **23**(6): 1111—1114.

[12] LU Xiao-Quan, LIU Hong-De. *The Wavelet Analysis Technology in Analytical Chemistry* [M]. Beijing: Chemistry Industry Press(卢小泉,刘宏德.分析化学中的小波分析技术.北京:化学工业出版社), 2006.