

相关检测技术在近红外光谱分析中的应用

刘青格, 邵定蓉, 李署坚

北京航空航天大学电子与信息工程系, 北京 100083

摘要 运用信号分析的理论, 研究了把相关检测技术应用于近红外光谱分析的方法。通过作药品甲硝唑纯品光谱及混合品光谱的相关函数值, 并对相关曲线进行分析, 结果表明通过相关处理, 光谱信息得到了分离和增强, 可以作为一种光谱信息特征提取的有效方法, 用于定性和定量分析, 对解决近红外光谱检测中光谱叠加、信息率低的难题提供了一种新的思路和探索。

关键词 相关检测; 特征提取; 光谱分析; 近红外辐射

中图分类号: TN911.7 **文献标识码**: A **文章编号**: 1000-0593(2005)12-1978-04

引言

近红外光谱是20世纪90年代以来发展最快、最引人注目的光谱分析技术^[1]。由于几乎所有的有机物的一些主要结构和组成都可以在它们的近红外光谱中找到信号, 而且谱图稳定, 获取光谱容易, 测试过程无污染, 检测成本低, 因此近红外光谱法(NIRS)被誉为分析的巨人^[2], 广泛应用于诸如石油、化工、食品、农业等各领域^[3]。然而, 由于近红外光谱的信息是分子内部振动的倍频和合频, 近红外光谱有其固有的弱点和难点, 主要是: (1)吸收强度较弱, 灵敏度低; (2)测定不经过预处理的样品光谱容易受样品状态、测量条件等影响, 光谱不确定性较大, 测定背景复杂, 光谱中谱峰重叠; (3)建模工作难度大。这些困难的核心是作为信息源的近红外光谱中有效信息率低, 对复杂样品进行近红外光谱分析是从复杂、重叠、变动的光谱中提取微弱信息, 难度很大^[4]。本文以药品甲硝唑为研究对象, 从信号检测的角度, 对于把相关检测技术应用于近红外光谱的信息提取作了尝试。

1 基本原理

1.1 基本原理

相关检测技术是应用信号周期性和噪声随机性的特点, 通过自相关或互相关运算, 达到去除噪声的一种技术。自相关函数是用来度量同一个随机过程前后的相关性, 由于自相关函数实际上是互相关分析的一种特殊情况, 这里只讨论互相关。互相关函数度量两个随机过程间的相关性, 表示两组数据之间依赖关系的相关统计量^[5]。

这里主要利用相关函数的叠加性质, 设实测信号为: $y(t) = x(t) + n(t)$, 则可以构造另一函数 $z(t)$, 作 $z(t)$ 与

$y(t)$ 的互相关函数 $R_{zy}(t)$ 。

设 $z(t)$ 与 $x(t)$ 的互相关函数为 $R_{zx}(t)$, 与 $n(t)$ 的互相关函数为 R_{zn} , 则有

$$R_{zy}(t) = R_{zx}(t) + R_{zn}(t) \quad (1)$$

1.2 相关检测技术^[6]

1.2.1 自相关检测

设混有随机噪声的信号

$$f_i(t) = S_i(t) + n_i(t) \quad (2)$$

则 $f_i(t)$ 的自相关函数为

$$\begin{aligned} R(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f_i(t) f_i(t-\tau) dt \\ &= R_{ss}(\tau) + R_{nn}(\tau) + R_{sn}(\tau) + R_{ns}(\tau) \\ &= R_{ss}(\tau) + R_{nn}(\tau) \end{aligned} \quad (3)$$

随着 τ 的增大, 噪声的自相关函数 $R_{nn}(\tau)$ 趋于零, 其结果使信号的自相关函数 $R_{ss}(\tau)$ 显示出来。

1.2.2 互相关检测

设带有随机噪声的信号为

$$f_1(t) = S_1(t) + n(t)$$

现有一与上述信号频率相同的“干净的”信号

$$f_2(t) = S_2(t)$$

则互相关函数

$$\begin{aligned} R_{12}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f_1(t) f_2(t-\tau) dt \\ &= R_{S_1 S_2}(\tau) + R_{n S_2}(\tau) \\ &= R_{S_1 S_2}(\tau) \end{aligned} \quad (4)$$

噪声与信号 $S_2(t)$ 间没有依附性, 互相关函数 $R_{n S_2}(\tau)$ 应等于零。这是从理论上分析, 实际中只有在计算相关函数求平均值的时间比较长时, $R_{n S_2}(\tau)$ 才趋于零。

由近红外多组分样品分析(NIRDRS)理论, 混合物样品

中任一组分浓度 c_j 与各组分吸光度近似满足线性关系^[4]。根据相关检测原理, 通过作甲硝唑纯品光谱与其他 3 个成分纯品光谱的互相关以及甲硝唑与不同成分含量的混合品光谱的互相关, 反映出不同混合谱与纯品谱之间的关系, 从而反映出混合谱与其成分含量的关系。此过程可以有效降低噪声的影响, 并达到光谱特征信息分离和增强的目的。

1.3 计算方法

计算相关函数的方法有两种, 即标准方法和间接方法: 标准方法直接计算数据值之间的平均乘积, 以此作相关函数的估计; 间接方法则先用 FFT 计算采样数据的功率谱密度, 再根据维纳——辛钦关系作傅里叶变换求得相关函数, 由于 FFT 的计算效率高, 用 FFT 方法计算相关函数比直接方法快, 相同的情况下, 数据点 N 越大, 相对计算速度越快。

2 试验方法

2.1 试验设备与仪器参数

本研究使用由北京第二光学仪器厂生产的 WQF-400N 型傅里叶快速变换近红外光谱分析仪。仪器参数: 采样光谱范围: $10\ 000\sim 4\ 000\ \text{cm}^{-1}$; 光源能量: 5 W; 本底材料: 硫酸钡; 采样分辨率: $2\ \text{cm}^{-1}$; 采样次数: 64 次; 重复试验次数: 3 次。

2.2 试验样品

甲硝唑药品主要由甲硝唑(a)、淀粉(b)、糊粉(c)、硬脂酸镁(d)组成。本试验采用 4 个纯品和 4 个不同成分含量的甲硝唑混合品进行检测。各混合品含量组成见表 1。

Table 1 The component content of different samples

样品号	a 甲硝唑/%	b 淀粉/%	c 糊粉/%	d 硬脂酸镁/%
1	66.993 8	31.108 6	1.103 7	0.793 9
2	47.991 4	1.771 1	49.380 8	0.805 8
3	57.992 7	1.820 6	39.406 9	0.779 8
4	66.989 3	2.505 0	29.722 0	0.783 7

2.3 试验结果

4 个纯品及 4 个不同成分含量组成的混合品吸收光谱分别如图 1 和图 2 所示。由于篇幅限制, 这里只列出样品 2 的吸收光谱。

3 相关计算与结果分析

3.1 相关计算

本文采用互相关函数的无偏估计公式和间接方法来计算^[7]

$$R_{xy}(m) = \frac{1}{N-|m|} \sum_{n=0}^{N-|m|-1} x(n)y^*(n+m) \quad (5)$$

式中 * 表示共轭。用 MATLAB 语言来实现。分别作了甲硝唑纯品与其他纯品的互相关曲线图(见图 3)。图 4 是甲硝唑与不同成分含量组成的 4 个混合品的互相关曲线图。图 5 是图 4 的局部放大图。必须说明的是, 由于利用(5)式进行

互相关函数的无偏估计, 在序列两端会出现很大的均方差, 因此, 如图所示, 在相关曲线两端有极大的变异。

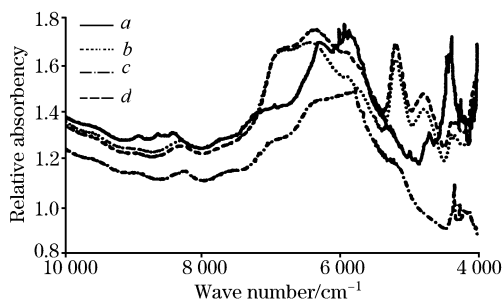


Fig. 1 The pure components' FTIR spectra

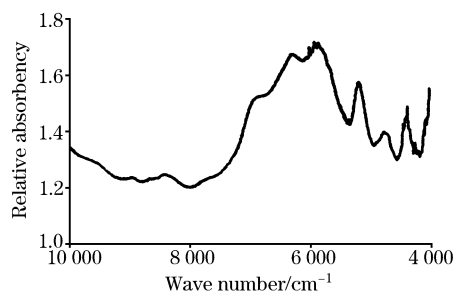


Fig. 2 The 2nd sample's FTIR spectrum

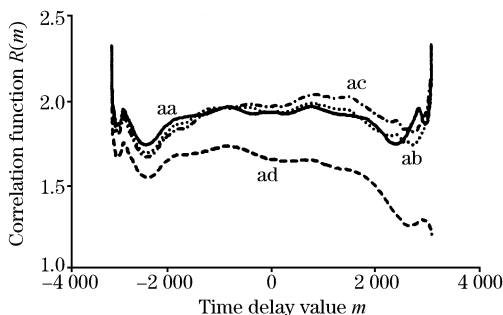


Fig. 3 The correlation between(a) and pure components

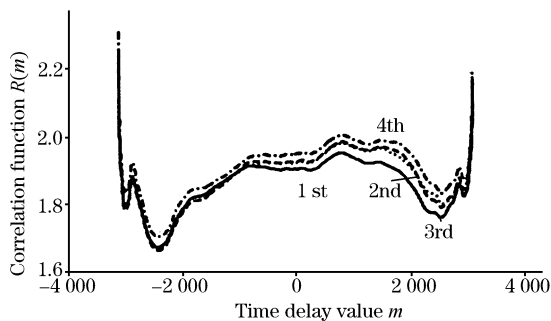


Fig. 4 The correlation between(a) and each sample

3.2 结果分析

(1)由图 3 可知, 甲硝唑与淀粉和糊粉的互相关曲线与其自相关曲线整体上相差不大。说明甲硝唑与淀粉和糊粉的相关性较大, 而与硬脂酸镁的相关性较小。与不同成分的相

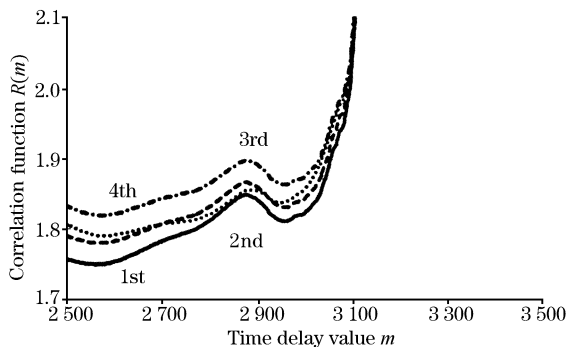


Fig. 5 The local enlarged results of Fig. 4

关曲线不同,使得甲硝唑与混合品互相关时,不同的光谱叠加导致不同的相关函数值,从而反映含量不同的信息。

(2)从局部来分析图 3,在约 $m=2800$ 附近,信息反映得非常清晰。甲硝唑自相关函数取极大值,与糊粉和淀粉的互相关函数取极小值,与硬脂酸镁的互相关则没有明显的极值,且各函数值相差很大。则甲硝唑与不同成分含量混合品相关时,各纯品光谱自相关的峰值大小不同,峰值位置不同,从而时移量 m 相同时,叠加结果呈现不同的相关函数值和不同趋势。相关函数极值大小不同,极值点位置不同,这便为提取和分离光谱信息提供了依据。因此可选择该点附近的信息点,作为分析混合品光谱的特征信息点。

(3)图 4 所示的相关曲线,都与图 3 所示的甲硝唑的自相关曲线形状相同,可得出结论:混合品中含有甲硝唑,达到定性分析的目的。

(4)图 5 是图 4 中相应于信息特征点附近的局部放大。由式 3 可知,与混合品的互相关可以看作纯品与不同成分的互相关函数值的叠加。在 $m=2800$ 附近,比较 2, 3, 4 号样

品曲线,可以看出,随着甲硝唑含量的增加,极值增大。1 号和 4 号样品,尽管甲硝唑含量基本相等,相关结果却相差很大。这是由于甲硝唑与糊粉的相关性较大,且大于其与淀粉的相关性,糊粉含量相差较大,由 1.103 7% 到 29.722 0% 时,将会对相关曲线带来很大影响,由于极小值的存在,降低了相关值。而对于 2, 3, 4 号样,随甲硝唑含量的增加,糊粉含量减小,由于甲硝唑与糊粉的相关值比甲硝唑自相关值小,因此相关值呈现增加的趋势。这说明混合品中各成分含量的不同反映在混合谱与纯谱的关系里,而这个关系则可以通过作纯品光谱与混合谱的互相关来定量反映。

4 结 论

根据结果分析,可总结得出以下结论。

(1)对相关检测技术提取近红外光谱中的信息作了试验分析,结果表明,相关分析作为一种光谱信息的特征提取手段,效果十分显著。

(2)利用相关分析,不仅可以定性的分析光谱中各成分的存在与否,而且可以从相关值的大小来定量分析混合品中各成分的含量。该方法可以减少在光谱处理中噪声的影响,有效的增强和提取叠加光谱中的信息,对以后进一步的光谱分析及建立模型、提高模型精度将有重大意义。

(3)该方法不仅适用于近红外光谱,对其他光谱,诸如紫外光谱、红外光谱等都适用。

(4)此外,还可以利用相关检测原理,综合运用自相关和互相关,来提取混合品中未知成分的信息,从而实现了检测的另一途径,同时,也可解决相关分析中找不到纯品的难题。

(5)将相关检测技术用于近红外光谱分析有一前提,那就是需得到纯品光谱,这是此方法的不足。

参 考 文 献

- [1] WANG Duo-jia, ZHOU Xiang-zu, JIN Tong-ming, HU Xiang-na, ZHONG Jiao-e(王多加, 周向祖, 金同铭, 胡祥娜, 钟娇娥). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(4): 447.
- [2] McClare W F. Anal. Chem. A, 1994, 66(1): 43.
- [3] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Analysis Technique of NIR(现代近红外光谱分析技术). Beijing: Sinica Petro-Chemistry Press(北京: 中国石化出版社), 2000.
- [4] WU Jin-guang(吴瑾光). The Technology and Application of Fourier Transform Infrared Spectroscopy (Book 2)(近代傅里叶变换红外光谱技术及应用·下). Beijing: Scientific and Technical Documents Publishing House(北京: 科学技术文献出版社), 1994.
- [5] WU Miao(吴 淼). Modern Engineering Signal Process and Application(现代工程信号处理及应用). Beijing: China Mining University Press(北京: 中国矿业大学出版社), 2000.
- [6] ZENG Qing-yong(曾庆勇). Faint Signal Detection(微弱信号检测). Second Edition(第 2 版). Hangzhou: Zhejiang University Press(杭州: 浙江大学出版社), 1994. 47.
- [7] ZHENG Hong-xing, YAO Ji-huan, ZHANG Cheng(郑宏兴, 姚纪欢, 张 成). The Application Skills of MATLAB 5. X Toolbox(MATLAB 5. X 工具箱使用技巧). Wuhan: Huazhong University of Science and Technology Press(武汉: 华中科技大学出版社), 2001.

The Application of Correlation Detection to Near-Infrared Spectroscopy

LIU Qing-ge, SHAO Ding-rong, LI Shu-jian

Beijing University of Aeronautics and Astronautics, Beijing 100083, China

Abstract Correlation detection was applied to near-infrared spectroscopy (NIRS) to extract the signal feature in the present paper. With metronidazole tablets as the object, the correlation both between each pure components and between pure metronidazole and different component-content mixtures were obtained. The results show that through correlation processing, the information feature of spectra could be separated and strengthened, and so better fitted to realize qualitative and quantitative spectrochemical analysis. Besides, the impact of noise could be lowered in this process. The conclusion is that correlation detection is an effective way to extract feature information and can be considered as a new solution to the problems of spectral superposition and low information rate in NIRS.

Keywords Correlation detection; Feature extraction; Spectroanalysis; NIR radiation

(Received Aug. 2, 2004; accepted Dec. 18, 2004)

第十四届全国分子光谱学学术会议征文通知

“第十四届全国分子光谱学术会议”由中国化学会和中国光学学会联合举办,预定于 2006 年 7 月在长春召开。届时会议将邀请国内外著名学者到会作报告,同时进行专题讨论和报告交流。现向国内外同行征集论文。

1. 征文内容:分子光谱理论研究;红外光谱、拉曼光谱、紫外光谱、荧光光谱等在化学化工、轻工、医药卫生、环境保护、地质矿物、农业、林业等领域的应用研究和基础研究的创造性论文。已经在期刊杂志或其他全国或国际会议上发表的论文不予受理。

2. 征文要求:提交应征论文的中文详细摘要(1500~1800 字)两份。摘要为 A4 规格复印纸二页,请用中文 Word 软件录入(生成 .doc 文件),用激光打印机输出。输出的版芯规格为 245 mm×165 mm。摘要内容应包括:文题(三号黑体居中)、作者(四号仿宋居中)、作者单位(小四号宋体居中,含城市名称、邮政编码和电子邮件地址(如有),并用逗号分开)、主要学术贡献、研究的意义与结果(五号宋体字)、主题词(自版芯左起)和主要参考文献(自版芯左起)。摘要需按上述顺序撰写,可以安排主要论据的图、表和照片。对应征论文要求和正式投稿要求一样,图注、表头一律用英文。

应征者需向会议提供摘要磁盘,稿件也可以附件形式通过 E-mail 传送。需附中文全文及英文(不小于 800 字符)摘要。附单位介绍信,注明论文联系人及电话、传真、电子信箱。

3. 收稿地址及联系人,来稿请寄:吉林省长春市前卫路 10 号,吉林大学超分子结构与材料教育部重点实验室,赵冰收。信封上请注明“第十四届全国分子光谱学学术会议征文”字样。

邮政编码:130012 电话:0431-5168486, 5168478

E-mail:zhaobing@mail.jlu.edu.cn

4. 截止日期:2006 年 1 月 30 日,以邮戳为凭。

5. 论文处理:筹备组将邀请和组织国内有关专家对应征论文进行评审。2006 年 4 月以前将发出第二轮通知,通知对论文的处理意见。会议召开的具体时间和具体问题,将在第二轮通知中公告,您对开好会有什么建议,请告 0431-5168486, 5168478 或 010-62181070。

6. 会议筹备工作进展和进一步详情,请登录网址:<http://supramol.jlu.edu.cn>

中国化学会物理化学委员会

中国光学学会光谱委员会

吉林大学超分子结构与材料教育部重点实验室

2005 年 4 月