

近红外分析中光谱预处理及波长选择 方法进展与应用

褚小立³ 袁洪福 陆婉珍

(石油化工科学研究院 北京 100083)

摘 要 光谱预处理和波长选取方法在近红外光谱分析技术中相当重要。本文综述了常用的 NIR 预处理和波长选取方法及这一领域的最新进展,详细介绍正交信号校正(OSC)、净分析信号(NAS)和小波变换(WT)等新光谱预处理方法以及无信息变量消除(UVE)和遗传算法(GA)等波长选取方法,并给出了这些方法的具体算法和一些应用实例。

关键词 近红外光谱 化学计量学 光谱预处理 正交信号校正 净分析信号 小波变换 遗传算法 无信息变量消除

中图分类号: O657133 文献标识码: A 文章编号: 10052281X(2004)0420528215

Progress and Application of Spectral Data Pretreatment and Wavelength Selection Methods in NIR Analytical Technique

Chu Xiaoli³ Yuan Hongfu Lu Wanzhen

(Research Institute of Petroleum Processing, Beijing 100083, China)

Abstract In the past decade, near infrared spectral analysis technique (NIR) has been quickly developed and widely applied in virtue of the development of chemometrics, in which spectral data pretreatment and wavelength selection methods play an important role. In the paper, the typical and commonly used pretreatment and wavelength selection methods are described. Some newly developed methods in this field such as orthogonal signal correction(OSC), net analyte signal(NAS), wavelet transform(WT), elimination of uninformative variables(UVE) and genetic algorithm(GA) methods are introduced in detail with 170 references. The algorithms and applications in NIR analysis of those methods are given and discussed.

Key words NIR; chemometrics; spectral data pretreatment; OSC; NAS; WT; GA; UVE

一、引 言

近年来,随着计算机技术和化学计量学的发展,现代近红外光谱(NIR)分析技术以其分析速度快、效率高、成本低和易于实现在线分析等特点,在农业、医药、石化、烟草和食品等行业得到广泛应用^[1-3]。NIR 光谱区(700—2500nm)主要是由含氢基团的倍频和组频吸收峰组成,吸收强度弱灵敏度相对较低,吸收带较宽且重叠严重。因此,依靠传统

的建立工作曲线方法进行定量分析是十分困难的。化学计量学的发展为这一问题的解决奠定了数学基础。

化学计量学方法在 NIR 分析中的应用主要包括以下 4 个方面:(1)光谱预处理和波长筛选方法^[4],目的是针对特定的样品体系,通过对光谱的适当处理或变换,减弱以至于消除各种非目标因素对光谱的影响,尽可能地去掉无关信息变量,提高分辨率和灵敏度,从而提高校正模型的预测能力和稳健性;

收稿: 2003 年 5 月, 收修改稿: 2003 年 7 月

³ 通讯联系人 e2mail: cxlyuli@sina.com

(2) 建立 NIR 定量分析模型的多元校正方法^[5],如多元线性回归(MLR)、主成分回归(PCR)、偏最小二乘(PLS)、局部权重回归(LWR)和人工神经网络(ANN)等,目的是建立用于预测未知样品性质或组成的分析模型;(3) 模式识别定性方法,如线性学习机、K2最近邻法和 SIMCA 法等有监督的模式识别方法,以及聚类分析法等无监督模式识别方法,目的是通过近红外光谱数据对不同样本按某些共同的特征进行分类识别,从而发现被量测样本之间的内在联系,获得决策性的信息^[6]。(4) 模型传递方法^[7],如有限脉冲响应算法(FIR)、直接校正算法(DS)、分段直接校正算法(PDS)和 Shenk's 算法等,目的是将在一台仪器上建立的定性或定量校正模型可靠地移植到其它相同或类似的仪器上使用,或将在某一条件建立的模型用于同一台仪器另一条件采集的光谱,从而减少建模所需的时间和费用。除此之外,应用于 NIR 分析的化学计量学方法还有模型界外样品的识别方法、校正样品的选择方法、模型质量控制方法以及模型评价方法等。

在近红外光谱分析使用的这些化学计量学方法中,光谱预处理和波长选取方法是基础,定性和定量分析都是在此基础上进行的。因此,光谱预处理和波长选取方法,尤其是光谱预处理方法对建立预测能力强、稳健性好的分析模型至关重要,有时甚至起决定作用^[8]。本文综述了常用的 NIR 预处理和波长选取方法及这一领域的最新进展,详细介绍正交信号校正(OSC)、净分析信号(NAS)和小波变换(WT)等新光谱预处理方法,以及用于波长选取的遗传算法(GA)和无信息变量消除(UVE)方法,并给出了这些方法的具体算法和一些应用实例。

二、光谱预处理方法与应用

近红外光谱仪所采集的光谱除样品的自身信息外,还包含了其它无关信息和噪音,如电噪音、样品背景和杂散光等。因此,在用化学计量学方法建立模型时,旨在消除光谱数据无关信息和噪音的预处理方法变得十分关键和必要。常用的谱图预处理方法有数据增强变换、平滑、导数、标准正态变量变换、多元散射校正、傅立叶变换等。近几年,小波变换、正交信号校正和净分析信号等一些新方法正在得到发展和应用。

1. 数据增强算法(data enhancement)

在使用多元校正方法建立近红外光谱分析模型时,将光谱的变动(而非光谱的绝对量)与待测性质

或组成的变动进行关联。基于以上特点,在建立 NIR 定量或定性模型前,往往采用一些数据增强(data enhancement)算法^[4]来消除多余信息,增加样品之间的差异,从而提高模型的稳健性和预测能力。常用的算法有均值中心化(mean centering)、标准化(autoscaling)和归一化(normalization)等,其中均值中心化和标准化是最常用的两种方法,在用这两种方法对光谱数据进行处理的同时,往往对性质或组成数据也进行同样的变换。

2. 平滑(smoothing)算法

由光谱仪得到的光谱信号中既含有有用信息,同时也叠加着随机误差(噪声)。信号平滑是消除噪声最常用的一种方法,其基本假设是光谱含有的噪声为零均随机白噪声,若多次测量取平均值可降低噪声提高信噪比。常用的信号平滑方法有移动平均平滑法和 SavitzkyGolay 卷积平滑法^[10,11]。

采用移动平均平滑法,平滑窗口宽度是一个重要参数:若窗口宽度太小,平滑去噪效果将不佳;若窗口宽度太大,进行简单求均值运算,会平滑掉一些有用信息,造成光谱信号的失真。为此 SavitzkyGolay 提出了卷积平滑法。

SavitzkyGolay 卷积平滑法与移动平均平滑法的基本思想是类似的,只是该方法没有使用简单的平均,而是通过多项式来对移动窗口内的数据进行多项式最小二乘拟合,其实质是一种加权平均法,更强调中心点的中心作用。SavitzkyGolay 卷积平滑法是目前应用较广泛的去噪方法,但应注意移动窗口宽度及多项式次数的优化选择。

3. 导数算法(derivative)

光谱的一阶(1st Der)和二阶导数(2nd Der)是 NIR 光谱分析中常用的基线校正和光谱分辨预处理方法。对光谱求导一般有两种方法:直接差分法和 SavitzkyGolay 求导法。

对于分辨率高、波长采样点多的光谱,直接差分法求取的导数光谱与实际相差不大;但对于稀疏波长采样点的光谱,该方法所求的导数则存有较大误差,这时可采用 SavitzkyGolay 卷积求导法计算。

SavitzkyGolay 卷积平滑也可用于求取导数光谱,通过最小二乘可计算得到与平滑系数相似的导数系数,可通过查表得到^[10]。

导数光谱可有效地消除基线和其它背景的干扰,分辨重叠峰,提高分辨率和灵敏度。但它同时会引入噪声,降低信噪比。在使用时,差分宽度的选择是十分重要的:如果差分宽度太小,噪声会很大,影

响所建分析模型的质量;如果差分宽度太大,平滑过度,会失去大量的细节信息。可通过差分宽度与校正标准偏差(SEP)或预测标准偏差(SEC)作图来选取最佳值^[12],一般认为差分宽度不应超过光谱吸收峰半峰宽的 115 倍。

4. SNV 和去趋势(de2trending)算法

标准正态变量变换(standard normal variate transformation, SNV)主要是用来消除固体颗粒大小、表面散射以及光程变化对 NIR 漫反射光谱的影响^[13]。SNV 与标准化算法的计算公式相同,不同之处在于标准化算法对一组光谱进行处理(基于光谱阵的列),而 SNV 算法是对一条光谱进行处理(基于光谱阵的行)。对需 SNV 变换的光谱 $X_{i,k}$ 按式(1)计算:

$$X_{i,SNV} = \frac{X_{i,k} - X_i}{\sqrt{\frac{\sum_{k=1}^m (X_{i,k} - X_i)^2}{m-1}}} \quad (1)$$

式中, X_i 为第 i 样品光谱的平均值(标量), $k=1, 2, \dots, m$, m 为波长点数; $i=1, 2, \dots, n$, n 为校正集样品数。

去趋势算法(de2trending)通常用于 SNV 处理后的光谱,用来消除漫反射光谱的基线漂移。其算法非常直接,首先按多项式将光谱 x_i 的吸光度和波长拟合出一趋势线 d_i ,然后把 d_i 从 x_i 减掉($x_i - d_i$)即可。该算法除了和 SNV 联合使用外,也可以单独使用。应注意的是在使用 SNV 前需要将反射光谱单位转换成 logIPR 的形式。

5. MSC 和 PMSC 算法

多元散射校正(multiplicative scatter correction, MSC)的目的与 SNV 基本相同^[14,15],主要是消除颗粒分布不均匀及颗粒大小产生的散射影响。MSC 算法的属性与标准化相同,是基于一组样品的光谱阵进行运算的。

(1) MSC

MSC 是由 Martens^[14]等人提出的,随后在 NIR 固体漫反射和浆状物透(反)射光谱分析时得到了广泛应用。MSC 的具体算法如下:

计算校正集样品的平均光谱 \bar{x} ($1 \times m$) (理想光谱);

将 x_i 与 \bar{x} 进行线性回归,

$$x_i = 1a_i + \bar{x}b_i, \text{求取 } a_i \text{ 和 } b_i; \quad (2)$$

$$x_{i,MSC} = (x_i - 1a_i)Pb_i; \quad (3)$$

其中, $i=1, 2, \dots, n$, n 为校正集样品数; 1 为 $1 \times m$ 的单位向量, m 为波长点数。

对于校正集外的光谱进行 MSC 处理时则需要用到校正集样品的平均光谱 \bar{x} ,即首先求取该光谱的 a 和 b ,再进行 MSC 变换。对于进行 MSC 处理的反射光谱单位应为 logIPR 或 Kubelka2Munk 形式。因 MSC 校正假定散射与波长及样品的浓度变化无关,所以,对组分性质变化较宽的样品, MSC 的处理效果较差。有文献证明 MSC 与 SNV 是线性相关的^[16],两种方法的处理结果也应是相似的。

除标准 MSC 算法外,还存在一些改进变形算法如反向信号校正(ISC)和扩展反向信号校正(EISC)^[17-21]。但这些方法都不如 MSC 应用广泛。

(2) PMSC

由 MSC 算法可以看出,它消除散射影响的基本假设是每条光谱与“理想光谱” \bar{x} 在全波长范围内存在线性关系,即光散射对每个样品、每个波长点产生的影响是线性的。因此,通过简单的最小二乘法便可对全波长范围的光谱消除由光散射产生的线性基线或背景的影响。但在大多数实际情况中,这样的假设并不存在,如对于不同颗粒大小的样品,光散射引起的背景将十分复杂,仅靠校正集的平均光谱作为标准谱是远远不够的。分段多元散射校正(piecewise multiplicative scatter correction)正是为消除这种非线性的散射而提出的^[22]。与 MSC 算法相比,PMSC 在进行校正时,假设在移动窗口宽度为 $j = (w_1 + w_2 + 1)$ 的波长范围内, x_{ij} 与平均光谱 \bar{x} 存在线性关系,对每一移动窗口分别按式(4)进行一元线性回归,由最小二乘法依次求出每段移动窗口的斜率 b_{ik} 和截距 a_{ik} 。

$$x_{ij} = 1a_{ik} + \bar{x}_j b_{ik} \quad (4)$$

用式(5)便可得到经 PMSC 校正后的光谱 $x_{i,k,PMSC}$:

$$x_{i,k,PMSC} = (x_{i,k} - a_{ik})P b_{ik} \quad (5)$$

式中, $x_{ij} = [x_{i,k-w_1}, x_{i,k-w_1+1}, \dots, x_{i,k}, x_{i,k+w_2-1}, x_{i,k+w_2}]$, x_j 为在窗口宽度为 $(w_1 + w_2 + 1)$ 波长段的平均光谱; 1 为 $1 \times (w_1 + w_2 + 1)$ 的单位向量。

移动窗口大小的选择对处理结果影响较大。若移动窗口过大,区间的线性关系将不存在;若移动窗口过小,会显著减弱不同样品间的光谱差异,使所建模型的预测能力变差。

6. 傅里叶变换(Fourier transform, FT)

傅里叶变换(FT)是一种十分重要的信号处理技术,它能够实现频域函数与时域函数之间的转换^[10,23-26],其实质是把原光谱分解成许多不同频率

的正弦波的叠加和。根据需要可通过 FT 对原始光谱数据进行平滑、插值、滤波、拟合及提高分辨率等运算,或用 FT 频率谱即权系数(傅里叶系数)直接参与模型的建立。在 NIR 光谱分析中,傅立叶变换可用来对光谱进行平滑去噪、数据压缩以及信息的提取。

McClure^[27]利用傅里叶变换系数进行多元逐步线性回归建立模型。Devaux^[28]用 FT 对用于面粉分类的 NIR 反射光谱数据压缩后再进行 PCA 处理,在不牺牲准确度的前提下,大大缩小了运算时间。Wu^[29]将傅里叶变换系数作为模式识别的特征变量,其结果与 PCA 相当。Pasi^[30]利用傅里叶变换与遗传算法相结合,从 NIR 光谱数据中提取相关特征进行多元校正。该方法将遗传算法选定的 PS 谱(pow2 er spectra)用于多元线性回归模型的建立,结果表明,与 GA2MLR 方法、全谱的 PCR 和 PLS 方法以及 UVE2PLS 方法相比,FT 是一个快速有效的信息提取及压缩方法,同时该方法也可用于研究校正集中样品分布情况。

7. 小波变换(wavelet transform, WT)

近年来,小波变换(WT)在信号和图像处理中的应用逐渐广泛和成熟起来。与 FT 相比,WT 具有时频局部化特性^[31,32]。WT 理论在 80 年代得到了迅速发展。其思想起源于工程学、物理学及纯数学领域,被认为是泛函分析、傅里叶分析、样条分析、调和分析、数值分析的完美结晶。自 1989 年以来,由于大量基函数的出现及快速算法的建立,WT 在许多领域得到了广泛的应用。90 年代初,WT 被引入化学领域并形成了化学小波分析。WT 能够将化学信号根据频率的不同,分解成多种尺度成分,并对大小不同的尺度成分采取相应粗细的取样步长,从而能够聚焦于信号中的任何部分,因此,被称为化学信号的“数学显微镜”。在分析化学领域中,WT 已成功应用于流动注射分析、伏安分析、高效液相色谱、红外光谱、质谱、核磁共振谱、可见紫外光谱、光声光谱、扩展 X 射线吸收精细结构(EXAFS)谱等分析化学信号的平滑滤噪、去除背景、数据压缩以及重叠信号解析^[33-38]。

(1) 小波变换的基本概念

与 FT 相比,WT 具有时频局部化特性。WT 的实质是将信号 $x(t)$ 投影到小波 $\psi_{a,b}(t)$ 上,即 $x(t)$ 与 $\psi_{a,b}(t)$ 的内积,得到便于处理的小波系数,按照分析的需要对小波系数进行处理,然后对处理后的小波系数进行反变换得到处理后的信号。

所谓的小波是指满足一定条件的函数 $\psi(t)$ 通过伸缩和平移产生的一个函数族 $\psi_{a,b}(t)$:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left[\frac{t-b}{a}\right] \quad (6)$$

$a, b \in \mathbb{R}, a \neq 0$

其中 a 用于控制伸缩(dilation),称为尺度参数(scale parameter), b 用于控制位置(position),称为平移参数(translation parameter), $\psi(t)$ 小波基或小波母函数。

$\psi(t)$ 必须满足两个条件:

小 (small): $\psi(t)$ 迅速趋向于零或迅速衰减为零

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (7)$$

连续 WT 可表示为: $WT_x(a, b) = \langle x(t), \psi_{a,b}(t) \rangle$

$$WT_x(a, b) = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}(t) dt \quad (8)$$

在分析信号的 WT 处理中,一般使用的是离散小波变换,与离散 FT 不同的是,此处的离散是针对连续的尺度参数 a 和连续参数 b ,而非时间变量 t 的。

离散小波定义 $a = a_0^m$ ($a_0 > 1, m \in \mathbb{Z}$), $b = nb_0 a_0^{-n}$ ($b_0 \in \mathbb{R}, n \in \mathbb{Z}$), 则 $\psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi(a_0^{-m} t - nb_0)$ 。一般取 $a_0 = 2, b_0 = 1$, 此时称为二进小波(dyadic wavelet)。

对于等波长间隔的 k 个离散光谱数据点 x_1, x_2, \dots, x_k , 其离散二进小波变换为:

$$WT_x(m, n) = \langle x_i, 2^{-\frac{m}{2}} (2^{-m} t_i - n) \rangle = \sum_{i=1}^k 2^{-\frac{m}{2}} (2^{-m} t_i - n) x_i \quad (9)$$

上式说明了 WT 实际上是将离散信号在小波基函数上的投影,不同的 m, n 代表不同的分辨率(尺度)和不同的时域(平移),小波函数正是通过不同的 m 和 n 来调节不同的局部时域和不同的分辨率。

与 FT 所用的基本函数(只有三角函数)相比,WT 中用到的小波函数不具有唯一性,即 $\psi(t)$ 具有多样性,同一问题用不同的小波函数进行分析,有时结果相差甚远。因此,小波函数的选用是 WT 在实际应用中的一个难点,目前通常采用经验或不断尝试方法,对比结果来选择最佳的小波函数。

在众的小波基函数家族中,有些小波函数被实践证明是十分有效的,其中在光谱分析中最常用主要有 Haar 小波、Daubechies (dbN) 小波、Coiflet 小波和 Symlets 小波等。

(2) 小波变换的计算方法 (Mallat 算法)

$m_{,n}(t)$ 一般不具有解析表达式,为实现有限离散 WT 计算,数值计算经常采用 Mallat 提出的多分辨率信号分解 (multiresolution signal decomposition, MRSD) 算法或塔式 (pyramid) 算法来实现,又称为 Mallat 算法^[39]。

将 $m_{,n}(t)$ 离散地表示成一对低通滤波器 $H = \{h_l\}$ 和高通滤波器 $G = \{g_l\}$, $(l = Z)$, $\{h_l^3\}$ 和 $\{g_l^3\}$ 为对应的镜像滤波器。对一等波长间隔的 k 个离散光谱数据点 x_1, x_2, \dots, x_k , 表示成 $C(l)$, 则正交离散二进小波分解可以写成:

$$C^j(i) = \sum_{l=Z} h^3(l-2i) C^{j-1}(l) \quad (10)$$

$$D^j(i) = \sum_{l=Z} g^3(l-2i) C^{j-1}(l) \quad (11)$$

式中, $j=0, 1, \dots, J$, J 为最高分解级次。由于分解正交性,通过 C^j 和 D^j 可以重构得到原始信号 C^0 :

$$C^{j-1}(i) = \sum_{l=Z} h(i-2l) C^j(l) + \sum_{l=Z} g(i-2l) D^j(l) \quad (12)$$

尺度参数 a 与 j 的关系为 $a=2^j$, 分辨率定义为 $1/2^j a$, 随着 j 的增加,分解的尺度二进扩展,细节分辨率随之降低。 C^j 和 D^j 分别称为 2^{-j} 分辨率下的离散近似和离散细节,即 C^j 表示频率低于 2^{-j} 的低频分量,而 D^j 表示频率介于 $2^{-j} \sim 2^{-j+1}$ 之间的高频分量。

低通滤波器 $H = \{h_l\}$ 和高通滤波器 $G = \{g_l\}$ 存在以下关系:

$$g_l = (-1)^l h_{l-Z} \quad \text{且} \quad \sum_{l=Z} h_l = \sqrt{2}, \quad \sum_{l=Z} g_l = 0 \quad (13)$$

(3) 小波包变换 (wavelet packet transform, WPT)

在正交离散二进小波分解处理中,只对低通滤波器的输出再进行变换处理,且其尺度是在二进制变化的,所以在高频段其频率分辨率较差,而在低频段时间分辨率较差。与 WT 不同的是,小波包变换对高频部分进行进一步分解,并能够根据被分析信号的特征,自适应地选择相应频带,使之与信号频谱相匹配,从而提高了时频分辨率。小波包变换的具体算法参见相关文献^[40,41]。

(4) 小波变换在 NIR 中的应用

从原理上讲,在 NIR 光谱分析中,用到 FT 的地方一般都可使用 WT,如光谱去噪平滑、光谱数据压缩和化学信息的提取等。

光谱去噪平滑

WT 用于平滑和滤噪的一般步骤为:首先对原始光谱进行 WT 得到高频和低频小波系数矢量;再通过阈值法去除小波系数中被认为是表示噪音的元素(称为滤噪),或去除小波系数中的高频(低尺度)元素(称为平滑);最后用经过处理的小波系数进行反变换即可得到滤噪后的光谱信号。阈值法通常有两种形式:硬阈值法,即把所有低于阈值的小波系数全部置零;软阈值法,即将小于阈值的小波系数置零并从大于阈值的小波系数的绝对值中扣除该阈值。关于阈值的估测方法也有不少报道^[42,43],如简单的软、硬阈值法, SURE 方法, VISU 方法, HYBRID 方法和 MINMAX 方法等。

闵顺耕等^[44]利用 WT 对 52 个烟草样品的漫反射光谱进行滤波处理,并用 PLS 法来计算烟草样品的总氮含量。结果表明,光谱 WT 滤波后,预测集的相对标准偏差由原来的 9.2% 降为 7.4%,提高了模型的预测准确性,其结果优于 FT 和五点三次平滑方法。Walczak^[45]将 WPT 应用到 NIR 光谱数据的模式识别中,其明显优于 SNV 方法的结果。

Walczak 和 Park 等人利用 NIR 光谱的小波细节系数进行模型传递^[46,47]。其基本思想是先对 NIR 光谱进行小波变换,得到小波细节系数;最后利用普通的模型传递算法校正小波系数,再进行小波重构,得到传递后的光谱,从而实现模型传递。由于小波细节系数基本上不含噪音,同时数据点少,因此,这种基于小波系数的模型传递方法具有速度快、抗干扰能力强的优点。Greensill 等人^[48]利用这种方法对建立的测定瓜中可溶固体含量的分析模型在不同二极管阵列近红外光谱仪间进行传递,其结果明显优于其它传递方法。Tian^[49]等人在此基础上提出了更为复杂的小波混合直接校正算法 (wavelet hybrid direct standardization, WHDS)。该方法将离散逼近和离散细节分别重构,然后利用常规模型传递算法,分别校正主仪器和子仪器的重构离散逼近和重构离散细节之间的差别,最后利用小波重构得到传递后的光谱。

光谱数据压缩

通过 WT 对数据进行压缩的基本原理类似于去噪^[50],一般采取如下步骤:先对原始数据进行 WT 得到小波系数;然后用阈值法删除小波系数中足够小而被认为不代表有用信息的系数,并保存处理后的系数;需要时,将其反变换即可得到原始数据。阈值的确定一般采用经验值或通过尝试得到。

Trygg 和 Wold^[51]将快速 Daubechies24 小波用于 NIR2Vis 光谱的数据压缩 (数据压缩比为 30),并将

压缩后的数据进行 PLS 分析,结果表明经过 WT 压缩的数据能够保持原始信号中的所有信息。Jetter 等人^[52]采用 WT 对小麦的漫反射光谱进行去噪和压缩处理,建立了测定水分含量的分析模型,取得了满意的结果。Yoon^[53]用 Harr 或 Daubechies 小波和硬阈值法得到的 NIR 压缩数据,采用 DS 算法实现了不同仪器间基于压缩数据所建分析模型的传递,其结果优于 PDS 算法。Trygg^[54]利用 2D 小波变换的能量谱(wavelet power spectrum, WPS),提出了一种更有效的界外样品检测方法,同时可将 NIR 光谱数据压缩 1 000 倍,较适合在线快速过程控制。

有效信息的提取

原始光谱经 WT 可得到反映不同信息的 WT 系数。根据先验知识或尝试方法,确定与组分有关的小波系数,可以直接利用这些小波系数建立定量或定性校正模型,或人为将此系数乘上大于 1 的倍数,然后进行重构,实现 NIR 光谱信息的提取。

Jouan2Rimbaud 等^[55]结合 WT 和无关变量消除方法(UVE)来提取光谱数据中与待测组分相关信息,用除去噪音和无关信息的光谱数据建立校正模型,提高了模型的预测能力。李华北^[56]采用 Dau2 bechies24 小波对食醋的 NIR 光谱进行了分解,分离得到了还原糖的特征信号频率,从而提高了 NIR 分析食醋中还原糖的预测能力。陈斌^[57]利用同样原理对 NIR 分析方便面中的含油率进行了优化,结果表明,与 1st和 2nd Der 预处理方法相比,WT 预处理效果较好。Depczynski^[58]利用小波变换系数回归与遗传算法相结合,用于近红外光谱的多组分分析,该方法自动完成数据预处理、波长区间选择以及模型的建立,结果优于使用 FT 系数建模的结果。Tan 等人^[59]用 WT 来消除 NIR 光谱中的非恒量背景,并与 1st Der、MSC 和 OSC 等预处理方法进行了比较,WT 的结果最为满意。近期,他们^[60]又提出了用于光谱校正的双域 HLS 和 PCR(DDHLS 和 DDPCR)方法,其校正和预测结果以及模型的稳健性均分别好于常规的 HLS 和 PCR 方法。Mittermary 等人^[61]也用 WT 消除由水分产生的 NIR 光谱背景,通过已知的化学知识选择信息强的小波系数,用 MLR 建立校正模型,其结果优于 FT 和 2nd Der 预处理经 HLS 建立的模型。

在 WT 对光谱进行预处理过程中,需要人为选择一些合适的参数,如小波基函数、压缩中的阈值、去噪中的截断尺度以及分解层次等,目前对这些参数的选择尚没有客观的标准,需要靠经验和尝试来

确定。尽管如此,因 WT 的时频局域性、多分辨率分析和可供选择的大量基函数等特点,使其不失为一种强有力的信号预处理方法。将 WT 与其它新化学计量学方法如 ANN 和 GA 相结合仍是一个重要的研究热点和发展方向。

8. 正交信号校正(orthogonal signal correction)

以上提到的光谱预处理方法,只是对谱图本身数据进行处理,并未考虑浓度阵的影响。所以,在进行预处理时,极有可能损失部分对建立校正模型有用的信息,又可能对噪音消除得不完全,而影响所建分析模型的质量。

正交信号校正方法是近几年来提出的一类新概念谱图预处理方法。目前有三种实现方式:正交信号校正(orthogonal signal correction, OSC)^[62]、直接正交信号校正(direct orthogonal signal correction, DOSC)^[63]和直接正交(direct orthogonalization, DO)^[64,65],其中 OSC 存有多种具体算法。这类预处理方法的基本原理均基于在建立定量校正模型前,将光谱阵用浓度阵正交,滤除光谱与浓度阵无关的信号,再进行多元校正,达到简化模型及提高模型预测能力的目的。

一般当光谱阵与浓度阵相关性不大,或光谱阵背景噪音太大时,用 HLS 或 PCR 方法建立校正模型,前几个主因子对应的光谱载荷往往不是浓度阵信息,而是与浓度阵无关的光谱信号。因此,在建立定量校正模型前,通过正交的数学方法将与浓度阵无关的光谱信号滤除,可减少建立模型所用的主因子数,进一步提高校正模型的预测能力和稳健性。此外,正交信号校正方法还可用于解决多元校正中的模型传递以及奇异点的检测等问题。以下主要介绍用于正交信号校正的几种流行算法。

(1) S. Wold 算法

S. Wold 等人^[62]第一次提出了正交信号校正思想,其具体算法如下:

将原始校正集光谱阵 $X(n \times m)$ 和浓度阵 $Y(n \times 1)$ 进行均值化中心化或标准化处理(n 为校正集样品数, m 为波长点数,下同);

计算光谱阵 X 的第一主成分得分向量 t ;

将 t 对 Y 作正交处理, $t_{\text{new}} = (I - Y(Y^T Y)^{-1} Y^T) t$, I 为单位阵;

计算权重向量 w , w 为 X 与 t_{new} 进行 HLS 或 PCR 运算得到的回归系数;

计算新的 t , $t = Xw$;

判断是否 $t - t_{\text{old}} P t < 10^{-6}$, 如果满足

进行下一步,否则返回步骤(3);

计算载荷向量 $p = t X P(t_{\text{new}})$;

在 X 中将正交信号减去, $X = X - t p$;

返回步骤 ,直至循环完所需的主因子数 f , f 为需正交处理的主因子数;

对于预测向量 x_{new} ,由权重 w 和载荷 p 求出校正后的光谱: $t = x_{\text{new}} W$, $x_{\text{OSC}} = x_{\text{new}} - t p$ 。

(2) Fearn 算法

Fearn 在 S. Wold 算法的基础上提出了一种简单快速的 OSC 算法^[66]:

将原始校正集光谱阵 $X(n \times m)$ 和浓度阵 $Y(n \times 1)$ 进行均值化中心化或标准化处理;

计算 $M = I - X Y(Y^T X Y)^{-1} Y^T X$, I 为单位阵;

计算 $Z = X M$;

对 Z 进行奇异值分解, $[U, S, V] = \text{svd}(Z)$;

取前 f 个需正交处理的特征值 $g = \text{diag}(S_f)$

及对应的载荷矩阵 $C = V_f$;

计算权重向量 $w_i = M X C_i P g_i, i = 1, 2, \dots, f$;

计算得分向量 $t_i = C_i g_i$;

计算载荷向量 $p_i = X t_i P(t_i, t_i)$;

在 X 中将正交信号减去, $X_{\text{OSC}} = X - \sum_{i=1}^f t_i p_i$;

对于预测向量 x_{new} ,由权重 w 和载荷 p 求出校正后的光谱: $t = x_{\text{new}} W$, $x_{\text{OSC}} = x_{\text{new}} - t p$

Feudale 基于 Fearn 算法提出了分段正交信号校正算法 (piecewise orthogonal signal correction, POSC)^[67],用来解决正交无关的局部性特点。对两组 NIR 光谱数据的处理结果表明,其性能略优于 OSC 算法,但同时带来了选择窗口大小的问题。S. Wold 算法与 Fearn 算法的比较可以看出, S. Wold 算法求取 t 和 p 时有数学根据,求 w 却没有; Fearn 算法求 w 有理论基础,而求取 t 和 p 却没有。Li 结合这两种方法提出了一种新的 OSC 算法^[68]。

(3) DOSC 算法

与 S. Wold 算法不同, Westerhuis 提出的 DOSC 算法^[63]首先将光谱阵 X 与 Y 正交,然后再对正交后的 X 进行主成分分析,求取 T 和 P 具体算法如下:

将原始校正集光谱阵 $X(n \times m)$ 和浓度阵 $Y(n \times 1)$ 进行均值化中心化或标准化处理;

计算 $M = X ((X)^{-1}) Y$;

计算 $Z = X - M M^{-1} X$;

对 $Z Z$ 进行主成分分析,取前 f 个需正交处理的主成分得分矩阵 T_f ;

计算权重矩阵 $W_f = X^{-1} T_f$, 广义逆 X^{-1} 通过 HLS 得到;

计算新的 $T_f, T_f = X W_f$;

计算载荷矩阵 $P_f = X T_f P(T_f T_f)$;

$X_{\text{DOSC}} = X - T_f P_f$;

对于预测向量 x_{new} ,由权重 w 和载荷 p 求出校正后的光谱: $T = x_{\text{new}} W$, $x_{\text{OSC}} = x_{\text{new}} - T P$

(4) DO 算法

DO 算法与 S. Wold 算法的差异在于前者是用逆偏最小二乘回归的方法滤除与浓度阵无关的信号,而 DO 算法是直接光谱阵与浓度阵正交来滤除无关的信号^[64,65]。因此, DO 算法比 OSC 算法简单,运行速度快,两种算法对光谱实际预处理的结果有一定的差异。DO 运算步骤如下:

将原始校正集光谱阵 $X(n \times m)$ 和浓度阵 $Y(n \times 1)$ 进行均值化中心化或标准化处理;

计算 $M = X Y(Y^T Y)^{-1}$;

计算 $Z = X - Y M$;

对 Z 进行主成分分析,取前 f 个需正交处理的得分矩阵 T_f 和载荷矩阵 P_f ;

计算新的 $T_f, T_f = X P$;

$X_{\text{OD}} = X - T_f^T P_f$;

对于预测向量 x_{new} ,由载荷 p 求出校正后的光谱: $T = x_{\text{new}} P$, $X_{\text{OD}} = x_{\text{new}} - T P$

在用正交信号校正算法对光谱进行预处理时,应注意以下两个问题:其一是对光谱阵正交处理所用主因子数的选取。用正交方法对光谱预处理都存在主因子的选取,一般选 1—5 个主因子数,但最终确定主因子数还依靠未知样品的预测结果,因此可用主因子数对验证集的预测标准偏差 (SEP) 作图来选取。其二是浓度阵准确性对光谱正交处理结果的影响。浓度参考阵数据的准确性对光谱正交处理的影响至关重要,若参考方法的测定结果不准确,在用该数据对光谱正交处理时,会滤除与浓度阵相关的部分信息,而保留与之无关的信号,从而使校正模型的预测能力变差。因此,在使用 OSC 方法时,一定要保证浓度阵的准确性。

正交信号校正算法提出不久就被用于解决 NIR 分析模型传递问题^[69],随后几乎涉及模型传递问题都将 OSC 算法作为对比方法^[70,71],其中 Geladi^[72]比

较了几种模型传递方法 (FIR、WT、PDS 和 S2G 平滑), 对用湖底沉淀物预测湖水 pH 值的 NIR 分析模型在不同仪器间进行传递, 证明 OSC 预处理方法的传递结果最佳。Blanco 等人^[73] 使用 OSC 方法有效消除了两类 NIR 光谱数据 (固体药物在线和实验室所测光谱) 的差异, 校正和预测结果均优于 1st Der、SNV 和 MSC 预处理方法。

需要说明的是, 由于校正过程中单独 PLS 方法在一定程度上也可以消除非线性和其它的不相关变量, 因此, 在大多数情况下, OSC 算法并未显著提高模型的预测能力, 在本质上也未简化模型所用的主因子数 (OSC 正交的主因子数与 OSC2PLS 的主因子数之和与单独使用 PLS 所用的主因子数基本一致)。Bertran^[74] 试图采用 OSC 算法提高 NIR 分析模型预测低浓度含量的能力, 结果并不理想。但 OSC 算法能够较好直观地解释光谱特征^[75]。目前, 已有含 OSC 算法的商品化近红外光谱 2 化学计量学分析软件^[76]。

最近, Svensson 对不同的 OSC 算法进行比较^[77]。Trygg 等人将 OSC 融入 PLS 回归步骤中, 提出了一些新的预处理和多元校正方法^[78-80]。

9. 净分析信号算法 (net analyte signal, NAS)

净分析信号也是有浓度阵参与的一种预处理算法^[81-83], 最早由 Lorber 提出^[84]。它的基本思想与 OSC 基本相同, 都是通过正交投影除去光谱阵中与待测组分无关的信息。其具体算法如下:

- (1) 将原始校正集光谱阵 $X (n \times m)$ 和浓度阵 $Y (n \times 1)$ 进行均值化中心化或标准化处理;
- (2) 计算 X 中与 Y 正交的部分 Z , $Z = (I - Y Y^T) X$, (I 为单位阵, $n \times n$);
- (3) 对 Z 进行主成分分析, 取前 f 个需正交处理的载荷矩阵 P_f ;
- (4) 计算正交投影矩阵 $R = I - P_f P_f^T$ (I 为单位阵, $f \times f$);
- (5) 计算经 NAS 处理后的 $X_{nas} = X R$;
- (6) 对于预测向量 X_{new} , $X_{nas} = X_{new} R$ 。

经过 NAS 处理后的校正集光谱阵, 一般通过 CLS 建立校正模型, 也可使用 PLS 或 PCR 建立。此外, NAS 还用来计算多元校正的分析性能参数, 如灵敏度、选择性、检测限和置信区间, 也可用来检测奇异点以及选择波长^[85-88]。

Boschetti^[89] 使用 NASPLS 建立了 NIR 测定橡胶中两种添加剂含量的校正模型, 其结果与单独使用 PLS、PCR 的结果相当。Goicoechea^[90] 比较了 NAS 和

OSC 算法的差异, 并对两组样品集分别用 OSCPLS、OSCPPLS、NASPLS 和 NASPPLS 建立了模型, 但 NAS 和 OSC 模型的预测能力均未得到显著提高。Berg2 er^[91-93] 基于 NAS 提出了一种新的多元校正方法——混合线性分析法 (hybrid linear analysis, HLA)。Xu^[94] 等人也基于 NAS 提出了一种不需要选择最佳主因子数的校正方法。Faber^[95] 则使用 NAS 评价了 MSC、1st Der 和 2nd Der 等预处理方法对 NIR 校正模型预测能力的影响。

在涉及浓度阵参与的光谱预处理方法中, 除了 OSC 和 NAS 算法外, 还有独立干扰消除算法 (independent interference reduction, IIR), 该方法主要用来解决 NIR 或 IR 测量低浓度物质的问题^[96], 以及 Ferre 提出的正交化算法^[97]。

三、波长选择方法与应用

在 NIR 结合 PLS 方法建模中, 传统观点认为 PLS 具有较强的抗干扰能力, 可全波长参与多元校正模型的建立^[98]。随着对 PLS 方法的深入研究和应用, 通过特定方法筛选特征波长或波长区间有可能得到更好的定量校正模型^[99]。波长选择一方面可以简化模型, 更主要的是由于不相关或非线性变量的剔除, 可以得到预测能力强、稳健性好的校正模型。

目前, 在多元校正分析中, 波长选择方法主要有相关系数法、方差分析法、逐步回归法、无信息变量的消除法 (UVE)、间隔偏最小二乘法 (interval PLS, iPLS)、遗传算法 (genetic algorithms, GA) 等, 其中 GA 是应用较广泛的一种波长选取方法。

1. 相关系数法和方差分析法

相关系数法是将校正集光谱阵中的每个波长对应的吸光度向量 x_i 与浓度阵中的待测组分浓度向量 y_i 进行相关性计算, 得到波长 2 相关系数 r 图^[100]。对应相关系数越大的波长其信息应越多, 因此, 可结合已知的化学知识给定一阈值, 选取相关系数大于该阈值的波长参与模型建立。相关系数 r 由下式计算:

$$r_j = \frac{\sum_{i=1}^n (x_{i,j} - x_j) (y_i - y)}{\sqrt{\sum_{i=1}^n (x_{i,j} - x_j)^2 \sum_{i=1}^n (y_i - y)^2}}$$

其中, $x_j = \left[\begin{matrix} x_{1,j} \\ \vdots \\ x_{n,j} \end{matrix} \right] / n$, $y = \left[\begin{matrix} y_1 \\ \vdots \\ y_n \end{matrix} \right] / n$, $j = 1, 2, \dots, m$, m 为波长点数; $i = 1, 2, \dots, n$, n 为校正集的样品数。

方差分析法是通过对校正集光谱阵在各波长下

的方差分析,得到波长 λ 标准偏差 s 图,对应标准偏差越大的波长,其光谱变动越显著。与相关系数法相似,可给定一阈值来选择波长区间。由于方差分析法不是针对待测组分优化选取波长,一般不用于定量分析模型,但该方法特别适合于定性分析中的波长选取。

王艳斌^[101]采用相关系数法选取波长,提高了 NIR 测定润滑油粘度指数校正模型的质量。王芳等人^[102]也用相同的波长选择方法,建立了 NIR 测定卷烟样品中常规成分的定量模型,与全谱方法相比,总糖、总植物碱、总氮、氯和蛋白质的 SEP 分别下降了 913%、415%、219%、112%和 219%。

值得注意的是,由于相关系数法是基于线性统计方法建立的,对于非线性相关及校正集样本分布不均匀的问题,通过该方法选取的结果往往不可靠。

2. 逐步回归分析方法 (stepwise regression analysis, SRA)

逐步回归法最初是 MLR 中选择回归变量的一种常用数学方法,即利用逐步回归法按一定显著水平筛选出统计检验显著的波长,再进行多元线性回归计算。后来,该方法发展为其它校正方法如 ANN 选取输入变量^[103]。逐步回归法的基本思想是,逐个选入对输出结果有显著影响的变量,每选入一个新变量后,对选入的各变量逐个进行显著性检验,并剔除不显著变量。如此反复选入、检验、剔除,直至无法剔除且无法选入为止。具体算法参见文献^[104]。

早期的近红外光谱分析大都采用 MLR 方法,逐步回归法在波长选取方面起到了重要的作用。姜能座^[105]采用逐步回归分析方法,选取 8 个波长建立了 NIR 漫反射测定饲料中水分含量的校正模型。ANN 是建立非线性校正模型较好的一种方法,但若将光谱仪采集的上千个数据直接输入 ANN 网络中,不但计算容量大,且会产生“过拟合”的分析模型,反而使其预测适应能力大大下降。齐小明^[106,107]则用逐步回归分析方法直接选取波长或选取主成分回归得到的主成分作为 ANN 网络的输入变量,得到了满意的结果。

在使用逐步回归法时经常遇到的问题是输入变量间具有多重交互作用,输入变量不仅与输出相关,而且彼此相关。在此情况下,模型中的一个输入变量可能会屏蔽其它变量对结果的影响。因此,逐步回归法选取的变量在大多数情况下不是最优的。

3. 无信息变量消除方法 (uninformative variables elimination, UVE)

无信息变量消除 (UVE) 方法^[108]是基于 PLS 回归系数 b 建立的一种波长选取方法。其具体算法如下:

(1) 将校正集光谱阵 $X(n \times m)$ 和浓度阵 $Y(n \times 1)$ 进行 PLS 回归,并选取最佳主因子数 f ;

(2) 人为产生一噪声矩阵 $R(n \times m)$,将 X 与 R 组合形成矩阵 $XR(n \times 2m)$,该矩阵前 m 列为 X ,后 m 列为 R ;

(3) 对矩阵 XR 和 Y 进行 PLS,每次剔除一个样品的交互验证,得到 n 个 PLS 回归系数组成矩阵 $B(n \times 2m)$;

(4) 按列计算矩阵 $B(n \times 2m)$ 的标准偏差 $s(1 \times 2m)$ 和平均值 $mn(1 \times 2m)$,然后计算 $h(i) = mn(i)Ps(i)$, $i = 1, 2, \dots, 2m$;

(5) 在 $[m+1, 2m]$ 区间取 h 的最大绝对值 $h_{\max} = \max(\text{abs}(h))$;

(6) 在 $[1, m]$ 区间去除矩阵 X 对应 $h < h_{\max}$ 的变量,并将剩余变量组成经 UVE 方法选取的新矩阵 X_{UVE} 。

UVE 方法在选取波长时集噪声和浓度信息于一体,且较直观实用。Centner 将其应用于模拟和实际 NIR 光谱数据,并与其它相关方法进行比较,UVE 方法得到的 SEP 最小^[108]。Koshoubu 等人对 UVE 算法中的 PLS 主因子数的选取进行了改进,应用于红外光谱测定水乙醇混合物中的乙醇含量,其结果优于相关系数等方法^[109]。Jouan2Rimbaud 还将 UVE 用于小波变换提取光谱数据中与待测组分相关信息^[55]。此外,基于 PLS 回归系数 b 或权重 w 的波长选择方法还有交互变量选择 (interactive variable selection, IVS) 等方法^[110-113]。

4. 遗传算法 (genetic algorithm, GA)

遗传算法最初是由 Holland 于 1975 年提出的,它借鉴生物界自然选择和遗传机制,利用选择、交换和突变等算子的操作,随着不断的遗传迭代,使目标函数值较优的变量被保留,较差的变量被淘汰,最终达到最优结果^[114]。NIR 自 70 年代提出以来,在分析化学领域得到了较多应用^[115-119],其中在特征变量筛选方面获得了较好的结果。

(1) 遗传算法实现过程

遗传算法的实现主要包括 5 个基本要素:参数编码、群体的初始化、适应度函数的设计、遗传操作设计、收敛判据和变量的选取等。NIR 的具体实现步骤如下^[120-122]:

参数编码。由于 NIR 不便直接处理空间数

据,需通过编码将它们表示成遗传空间的基因串结构数据,一般采用基于 OPI 字符的二进制串形式。对于包含 m 个参数(如波长)的问题,可用一串含有 $n \times k$ 个字符(对应于基因)的向量(对应于染色体)表示, k 表示每个参数需要的基因位数。对于波长选择来说,通常 k 选取 1,即一条染色体中的每个基因对应一个实际参数,若基因为 1 表示其代表的参数被选中,基因为 0 则未被选中。

群体的初始化。随机或根据一定的限制条件产生一个给定大小的初始群体,群体的大小即个体(染色体)的数目可根据参数(基因)的多少选定,一般选 30—100。

适应度函数的设计。NIR 根据适应度函数来评价个体的优劣,作为以后遗传操作的依据。由于在整个搜索进化过程中,只有适应度函数与所解决的具体问题相联系,因此,适应度函数的确定至关重要。对于波长选择,可选用 PLS 交互验证中浓度阵的预测值与实际值的相关系数(r)、SEC 或 SEP 作为适应度函数。

遗传操作设计。遗传操作由选择、交叉和变异构成。

选择:选择算子又称复制算子,通过选择把适应度高的个体直接遗传到下一代或通过交叉或变异产生新的个体再遗传到下一代。选择操作是建立在群体中个体的适应度评估基础上的。选择的目的是为了避开基因缺陷,提高全局收敛性和计算效率。选择方法包括适应度比例、最优保存、确定式采样及排序选择等方法,其中最常用的选择方法为适应度比例方法,也称转轮法,每个个体的选择概率与其适应度成比例。

交叉:交叉运算是指两个相互配对的染色体按某种方式相互交换其部分基因,从而形成两个新的个体。它是 NIR 中最主要的算子,是产生新个体的主要方法,寻优的搜索过程主要是通过它来实现的,因此,它决定了 NIR 的全局搜索能力。交叉算子有随机一点交叉、两点与多点交叉、均匀交叉和算术交叉等,交叉概率一般选择 0.15—0.18。在交叉运算前必须对群体中的个体进行配对,目前常用随机配对策略,即将群体中的 N 个个体随机组成 $N/2$ 对配对个体组,交叉运算在这些配对个体组中的两个个体之间进行。

变异:变异是将个体染色体编码串中的某些基因进行补运算,即 0 变为 1,或 1 变为 0。引入变异算子的目的是维持群体的多样性,防止出现未成熟

收敛现象,此外还改善 NIR 的局部搜索能力。交叉算子和变异算子相结合,共同完成对搜索空间的全局和局部搜索,从而使 NIR 能够以良好的搜索性能完成最优化问题的寻优过程。最简单的变异算子为基本位变异算子,即在个体中随机挑选一个或多个基因以变异概率作变动,变异概率为 0.101—0.11。此外还有均匀变异、非均匀变异、边界变异和高斯变异等变异算子。

收敛判据。常规的数学规划方法在数学上都有比较严格的收敛判据,但 NIR 的收敛判据基本是启发式的。因此,NIR 的判据较多,如计算时间、计算机变量或从解的质量方面等确定判据。选取遗传迭代次数是常用的收敛终止条件,其取值范围一般为 100—1 000。

变量选取。在遗传迭代终止后,所有变量按选取频率重新排列,再由选取变量数与适应度函数作图选定最佳变量数,最终得到所选的变量。

(2) 遗传算法在 NIR 波长选择中的应用

NIR 用于 NIR 波长选择的目的是优化分析模型以提高其预测能力,以及建立抗外界因素如环境温度影响小的稳健分析模型,而且通过所选波长可以更好地解释待测组分对应的光谱区域。

Rimbaud^[123]将 GA 与 MLR 相结合用于 NIR 测定不同种类(不同化学结构)聚醚多元醇的羟值,其结果与全谱 PLS 或 PCR 相当。Ding^[124]考察了不同仪器分辨率对 GA 选取波长结果的影响,随着分辨率的下降,GA 对波长优化选择的优势逐渐变强。Ban2galore 等人^[125]针对三组光谱交叠严重且浓度含量在检测限附近的 NIR 光谱数据,用 GA 选取与待测物相关性强的波长通过 PLS 建立分析模型,结果较全谱 PLS 显著提高。Roger^[126]使用 GA 对短波 NIR 测定樱桃中白利糖度的模型进行优化,其结果要比平滑、归一化和导数等预处理方法所得的 SEP 低约 3 个白利糖度,同时还比较了不同交互验证方法对波长选择的影响。Leardi^[127,128]先后用多个 NIR 校正集(包括大豆、牛奶、小麦、汽油、粮食和有机物混合物等)10 余种组成和性质对 GA 选取波长优化 PLS 分析模型的有效性进行了考察,均取得了较好的结果。最近,Leardi^[129]又将 GA 用于选取红外光谱波长,结合 PLS 测定聚合物膜中的两种添加剂,不仅提高了模型的预测能力,而且有助于解释光谱。褚小立^[130]利用 GA 对 NIR 测定石油产品有关组成的波长变量进行筛选,结果表明,通过 GA 选取波长在简化 PLS 模型的同时也增强了所建立模型的预测能力,尤其

适用于单纯 PLS 较难校正关联的体系。王宏等人^[131]的研究结果也表明,GA 优选的波长与样品中被测成分有关,即使样品复杂程度发生变化这种有效性也未变,可提高 NIR 无创性检测人体血糖浓度的精度。Smith^[132]则将 GA 波长选择与 PCA2MD 方法结合用于 NIR 模式识别,完全将两种极为相似的微晶纤维素进行分类,而使用全波长的正确识别率只有 85.7%。van den Broek^[133]则将类似的方法用于红外光谱成像识别中。

选取信息强且对外界因素变化不敏感波长是建立稳健 NIR 分析模型的一个有效途径^[134,135]。褚小立等人^[75]利用 GA 选取对样品温度不敏感的波长,建立了样品温度稳健的测定重整汽油芳烃组成和辛烷值的 NIR 校正模型。Durmus^[136-139]则采用基于 GA 选取波长的遗传回归方法 (genetic regression) 来建立稳健模型以及解决模型在不同仪器间传递的问题。所谓的遗传回归是通过将 GA 选取的波长进行组合以及简单的数学运算后,利用最小二乘回归来实现的。

GA 除选择波长外,还可用于模型建立中其它变量的优化选取。Shaffer 等人^[140]在使用 NIR 对牛血清白蛋白、人血清和牛血中的葡萄糖进行测定时,采用 GA 对带通数字滤波的位置、带宽、光谱范围以及 PLS 主因子数进行了优化,同时考察了不同适应度函数对结果的影响。Ding^[141]也采用 GA 对带通傅里叶数字滤波的位置和带宽进行选择,使 NIR 测定水中微量有机污染物 (1—160ppm) 成为可能。Depczynski^[142]利用 GA 选取小波系数建立 NIR 定量分析模型。Barros^[143]和 Frost^[144]采用 GA 对 PLS 或 PCR 所用的主成分进行优化选取组合,所建模型的预测能力优于传统按顺序选取前几个主成分来回归的结果。

综上所述,GA 以其全局最优、易实现等特点,成为目前最常用且最有效的一种波长选择方法^[145,146]。但在实际使用时应注意以下问题:其一,由于 NIR 的初始群体是随机选取的,选择、交叉和变异也带有较强的随机性,所以不能保证每次波长选取结果的一致性;其二,根据经验,在使用 GA 时,校正集中波长变量与样品数的比值一般要小于 4,否则得到的结果是不可靠的;其三,选择的合适的适应度函数对 GA 尤其重要,不同的适应度函数得到的结果将大相径庭。

5. 其它波长选取方法

模拟退火算法 (simulated annealing algorithm,

SAA) 也是一种常用的变量选取方法^[147],提出于本世纪 80 年代初,其思想源于固体退火过程:将固体加温至充分高,再让其徐徐冷却。加温时,固体内部粒子随温升变为无序状,内能增大;而徐徐冷却时粒子渐趋有序,在每个温度都达到平衡态,最后在常温时达到基态,内能减为最小。根据 Metropolis 准则,粒子在温度 T 时趋于平衡的几率为 $e^{-\Delta E/(kT)}$,其中 E 为温度 T 时的内能, ΔE 为其改变量, k 为 Boltzmann 常数。SAA 解决组合优化问题的步骤是,将内能 E 模拟为目标函数值 f ,温度 T 演化成控制参数 t ,即得到解组合优化问题的模拟退火算法:由初始解 p 和控制参数初值 t 开始,对当前解重复“产生新解 计算目标函数差 判断是否接受 接受或舍弃”的迭代,并逐步衰减 t 值,算法终止时的当前解即为所得近似最优解。退火过程由冷却进度表控制,包括控制参数的初值 t 及其衰减因子 α 、每个 t 值时的迭代次数 L 和停止条件 S 。SAA 用于波长选取的具体算法及应用可参见有关文献^[148,149]。

间隔偏最小二乘方法 (interval PLS, iPLS)^[150] 是 Norgaard 提出的一种波长区间选择方法:首先将 NIR 全谱分割成等长的多个区间,用 PLS 回归得到的 SEC 评价出最优的一个区间;然后再以该区间为中心单向或双向扩充 (或消减) 波长变量,得到最佳的波长区间。该方法还可以与 GA 方法结合,选择几个最优的波长区间组合。此外,与 iPLS 相似的光谱区间选择方法还有窗口移动 PLS 方法^[151,152]。

除以上提到的波长选择方法外,用于光谱波长选取的方法还有统计学方法^[159]、多链方法 (multiple chain method, MCM)^[154,155] 及连续投影波长选取方法等^[156-158]。

四、结束语

光谱预处理和波长选取方法在近红外光谱分析中的重要地位已是不言而喻的,其直接决定着所建分析模型的预测能力和长期可靠性。目前,文献涉及的光谱预处理和波长选取方法多达几十种,且每种方法如小波变换又有不同的参数和函数。所以,在实际应用时,就会遇到如何选取最优方法的问题。尽管有一定的规律可寻,如导数方法一般用于基线校正,MSC、SNV 和二阶导数方法用于漫反射 NIR 光谱以消除颗粒分布不均匀引起的光散射,小波变换可以有效消除光谱背景,提高模型的稳健性,如果使用得当波长选择方法总能简化模型、提高预测能力等。但在具体应用时,不同的分析体系及所解决问

题的不同,最佳预处理方法也不尽相同,仍需要对一些可能的方法进行比较,以确定最佳结果^[159-167]。

若分析体系相对复杂,仅用一种光谱预处理方法往往不能得到较好的结果,这时可将不同预处理和波长选取方法结合使用,以获得预期的结果。但不同预处理方法的组合及运算顺序仍需要尝试优化。早期,有文献提出采用因子设计方法来解决预处理方法的组合问题^[168],但这一问题有待进一步探讨。在2003年召开的11届国际近红外光谱会议上,光谱预处理方法作为一个专题进行了研讨。其中,FernándezCaban 等人将7种预处理方法组合成56种方法,对NIR测定饲料中的粗蛋白和粗脂肪^[169]以及测定猪脂肪中的各种脂肪酸^[170]的分析模型进行了优化,得到了满意的结果。

将预处理和波长选取方法融入到多元校正步骤中形成新的校正和预处理方法,而非在校正之前单独使用,是该领域未来的一个重要发展方向。可以相信,随着对预处理和波长选取方法的深入研究和理解,势必会推动近红外光谱分析技术的不断发展和应用。这些预处理和波长选取方法对其它分析技术也有一定的借鉴意义。

参考文献

- [1] 陆婉珍,袁洪福,徐广通. 现代近红外光谱分析技术. 北京:中国石化出版社,2000
- [2] Kemény G J. Handbook of Near-Infrared Analysis. Marcel Dekker Inc., 2001
- [3] 冯新沪,史永刚. 近红外光谱及其在石油产品分析中的应用. 中国石化出版社,2002
- [4] Mobley P R, Kowalski B R. Appl. Spectrosc. Rev., 1996, 31: 347—368
- [5] 梁逸曾. 白灰黑复杂多组分分析及其化学计量学算法. 湖南科技出版社,1996
- [6] ASTM E1655200, Standard Practices for Infrared Multivariate Quantitative Analysis
- [7] 褚小立(Chu X L),袁洪福(Yuan H F),陆婉珍(Lu W Z). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2001, 21: 881—885
- [8] Seasholtz M B, Kowalski B R. Anal. Chim. Acta, 1993, 277: 165—177
- [9] Seasholtz M B, Kowalski B R. J. Chemom., 1992, 6: 103—111
- [10] 梁逸曾,俞汝勤. 分析化学手册(10)——化学计量学. 北京:化工出版社,2001
- [11] 刘树深,易忠胜. 基础化学计量学. 北京:科学出版社,1999
- [12] 徐广通(Xu G T),袁洪福(Yuan H F),陆婉珍(Lu W Z). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2000, 20: 619—622
- [13] Barnes R J, Dhanoa M S. Appl. Spectrosc., 1989, 43: 772—777
- [14] Isaksson T, Naes T. Appl. Spectrosc., 1988, 42: 1273—1284
- [15] Chen J Y, Iyo C, Teradab F. J. Near Infrared Spectrosc., 2002, 10: 301—307
- [16] Dhanoa M S, Liste S J, Sanderson R. J. Near Infrared Spectrosc., 1994, 2: 43—47
- [17] Helland I S, Naes T, Isaksson T. Chemom. Intell. Lab. Syst., 1995, 29: 233—241
- [18] Martens H, Stark E. J. Pharmaceutical and Biomedical Analysis, 1991, 9: 625—635
- [19] Martens H, Pram N J, Balling E S. Anal. Chem., 2003, 75: 394—404
- [20] Pedersen D K, Martens H, Nielsen J P. Appl. Spectrosc., 2002, 56: 1206—1214
- [21] Stark E, Martens H. US 5 568 400, 1996
- [22] Isaksson T, Kowalski B. Appl. Spectrosc., 1993, 47: 702—709
- [23] 鲍伦军(Bao L J),莫金垣(Mo J H),唐祖英(Tang Z Y). 化学学报(Acta Chim. Sinica), 1997, 55: 907—914
- [24] Small G W, Arnold M A, Marquardt L A. Anal. Chem., 1993, 65: 3279—3289
- [25] 王洪(Wang H),司圣柱(Si S Z). 分析仪器(Analytical Instrumentation), 1997, 2: 46—49
- [26] 何锡文(How X W),李鼎(Li D). 分析化学(Chin. J. Anal. Chem.), 1994, 22: 94—100
- [27] McClure W F. Appl. Spectrosc., 1988, 38: 322—329
- [28] Devaux M, Bertrand D, Robert P. J. Chemom., 1987, 1: 103—110
- [29] Wu W, Walczak B, Penninckx W, Massart D L. Anal. Chim. Acta, 1996, 331: 75—83
- [30] Pasti L, Jouan Rimbaud D, de Noord O E. Anal. Chim. Acta, 1998, 364: 253—263
- [31] 卢小泉(Lu X Q),莫金垣(Mo J H). 分析化学(Chin. J. Anal. Chem.), 1996, 24: 1100—1106
- [32] Fearn T, Davies A M C. J. Near Infrared Spectrosc., 2003, 11: 333—339
- [33] 高志明(Gao Z M),李井会(Li J H),高礼让(Gao L R)等. 化学进展(Prog. Chem.), 2000, 12: 179—191
- [34] 邵学广(Shao X G),庞春艳(Pang C Y),孙莉(Shun L). 化学进展(Prog. Chem.), 2000, 12: 233—244
- [35] 秦侠(Qin X),沈兰荪(Shen L S). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2000, 20: 892—897
- [36] 郭怀忠(Guo H Z),杨准(Yang Z),张尊建(Zhang Z J). 药理学进展(Prog. Pharm.), 2000, 1: 5—9
- [37] Cai C S, Harrington P D B. J. Chem. Inf. Comput. Sci., 1998, 38: 1161—1170
- [38] Bakshi B R. J. Chemom., 1999, 13: 415—434
- [39] Mallat S G. IEEE Trans. Pattern Anal. Mach. Intell., 1989, 11: 674—693
- [40] Walczak B, Massart D L. Chemom. Intell. Lab. Syst., 1997, 36: 81—94
- [41] 邵学广(Shao X G),蔡文生(Cai W S). 高等学校化学学报(Chem. J. Chin. Univ.), 1999, 20: 42—46

- [42] Barclay V J, Bonner R F, Hamilton I P. *Anal. Chem.*, 1997, 69: 78—90
- [43] Alsberg B K, Woodward A M, Winson M K. *Analyst*, 1997, 122: 645—652
- [44] 阎顺耕(Yan S G), 谢秀娟(Xie X J). *分析化学*(*Chin. J. Anal. Chem.*), 1998, 26: 34—37
- [45] Walczak B, van den Bogaert B, Massart D L. *Anal. Chem.*, 1996, 68: 1742—1747
- [46] Walczak B, Bouveresse E, Massart D L. *Chemom. Intell. Lab. Syst.*, 1997, 36: 41—51
- [47] Park K S, Ko Y H, Lee H. *Chemom. Intell. Lab. Syst.*, 2001, 55: 53—65
- [48] Greensill C V, Wolfs P J, Spiegelman C H. *Appl. Spectrosc.*, 2001, 55: 647—653
- [49] Tian H W, Brown S D. *J. Chemom.*, 2001, 15: 647—663
- [50] Chau F T, Gao J B, Shih T M, Wang J. *Appl. Spectrosc.*, 1997, 51: 649—659
- [51] Trygg J, Wold S. *Chemom. Intell. Lab. Syst.*, 1998, 42: 209—220
- [52] Jetter K, Depczynski U, Molt K, Niemüller A. *Anal. Chim. Acta*, 2000, 420: 169—180
- [53] Yoon J, Lee B, Han C H. *Chemom. Intell. Lab. Syst.*, 2002, 64: 1—14
- [54] Trygg J, Wold N K, Lars W. *J. Chemom.*, 2001, 15: 299—319
- [55] Jouan2Rimbaud D, Walczak B, Poppi R J. *Anal. Chem.*, 1997, 69: 4317—4323
- [56] 李华北(Li H B), 陈斌(Chen B). *农业工程学报*(*Acta Agri. Engineer. Sinica*), 2000, 16: 114—117
- [57] 陈斌(Chen B). *农业机械学报*(*Acta Agri. Machine Sinica*), 2001, 32: 84—87
- [58] Depczynskia U, Jettera K, Moltband K. *Chemom. Intell. Lab. Syst.*, 1999, 47: 179—187
- [59] Tan H W, Brown S D. *J. Chemom.*, 2002, 16: 228—240
- [60] Tan H W, Brown S D. *J. Chemom.*, 2003, 17: 111—122
- [61] Mittermery C R, Tan H W, Brown S D. *Appl. Spectrosc.*, 2001, 55: 827—833
- [62] Wold S, Antti H, Lindgren F. *Chemom. Intell. Lab. Syst.*, 1998, 44: 175—185
- [63] Westerhuis J A, De Jong S, Smilde A K. *Chemom. Intell. Lab. Syst.*, 2001, 56: 13—25
- [64] Andersson C A. *Chemom. Intell. Lab. Syst.*, 1999, 47: 51—63
- [65] FernandezPiera J A, Massart D L. *Chemom. Intell. Lab. Syst.*, 2001, 55: 101—108
- [66] Fearn T. *Chemom. Intell. Lab. Syst.*, 2000, 50: 47—52
- [67] Feudale R N, Tan H W, Brown S D. *Chemom. Intell. Lab. Syst.*, 2002, 63: 129—138
- [68] Li B B, Morris A J, Elaine B M. *J. Chemom.*, 2002, 16: 556—561
- [69] Sjoblom J, Svensson O, Josefson M. *Chemom. Intell. Lab. Syst.*, 1998, 44: 229—244
- [70] Fearn T. *J. Near Infrared Spectrosc.*, 2001, 9: 229—244
- [71] Feudale R N, Woody N A. *Chemom. Intell. Lab. Syst.*, 2002, 64: 181—192
- [72] Geladi P, B3ring H, D.hakk E. *J. Near Infrared Spectrosc.*, 1999, 7: 251—264
- [73] Blanco M, Coello J, Montoliu I. *Anal. Chim. Acta*, 2001, 434: 125—132
- [74] Bertran E, Iturriaga H, Maspoeh S, Montoliu I. *Anal. Chim. Acta*, 2001, 431: 303—311
- [75] 褚小立(Chu X L), 袁洪福(Yuan H F), 王艳斌(Wang Y B), 陆婉珍(Lu W Z). *光谱学与光谱分析*(*Spectroscopy and Spectral Analysis*), 2004, (accepted)
- [76] The Unscrambler 0 Accessory Pack for Spectroscopy. CAMO Technologies Inc., USA, 2003
- [77] Svensson O, Kourti T, Macgregor J F. *J. Chemom.*, 2002, 16: 176—188
- [78] Trygg J, Wold S. *J. Chemom.*, 2002, 16: 119—128
- [79] Trygg J. *J. Chemom.*, 2002, 16: 283—293
- [80] Trygg J, Wold S. *J. Chemom.*, 2003, 17: 53—64
- [81] Lorber A, Faber K, Kowalski B R. *Anal. Chem.*, 1997, 69: 1620—1626
- [82] Faber N M. *Anal. Chem.*, 1998, 70: 5108—5110
- [83] Ferre J, Brown S D, Rius F X. *J. Chemom.*, 2001, 15: 537—553
- [84] Lorber A. *Anal. Chem.*, 1986, 58: 1167—1172
- [85] Faber N M. *Anal. Chem.*, 1998, 70: 2972—2982
- [86] Ferre J, Rius F X. *Anal. Chem.*, 1998, 70: 1999—2007
- [87] Gicoechea H C, Olivieri A C. *Analyst*, 1999, 124: 1999—2007
- [88] Boque R, Rius F X. *Chemom. Intell. Lab. Syst.*, 1996, 32: 11—23
- [89] Boschettia C E, Olivieri A C. *J. Near Infrared Spectrosc.*, 2001, 9: 245—254
- [90] Gicoechea H C, Olivieri A C. *Chemom. Intell. Lab. Syst.*, 2002, 63: 129—138
- [91] Berger A J, Koo T W, Itzkan I, Feld M S. *Anal. Chem.*, 1998, 70: 623—627
- [92] 亓云鹏(Qi Y P), 吴玉田(Wu Y T), 李通化(Li T H), 柴逸峰(Chai Y F). *分析化学*(*Chin. J. Anal. Chem.*), 2002, 30: 401—405
- [93] Gicoechea H C, Olivieri A C. *Anal. Chem.*, 1999, 71: 4361—4368
- [94] Xu L, Schechter I. *Anal. Chem.*, 1997, 69: 3722—3730
- [95] Faber N M. *Anal. Chem.*, 1999, 71: 557—565
- [96] Hansen P W. *J. Chemom.*, 2001, 15: 123—131
- [97] Ferre J, Brown S D. *Appl. Spectrosc.*, 2001, 55: 708—714
- [98] Thomas E V, Haaland D M. *Anal. Chem.*, 1990, 62: 1091—1099
- [99] Thomas E V. *Anal. Chem.*, 1994, 66: 795A—804A
- [100] Wu W, Walczak B, Massart D L. *Anal. Chim. Acta*, 1995, 315: 243—255
- [101] 王艳斌(Wang Y B), 袁洪福(Yuan H F), 陆婉珍(Lu W Z). *润滑油*(*Lubricating Oil*), 2001, 16: 53—56
- [102] 王芳(Wang F), 陈达(Chen D), 邵学广(Shao X G). *烟草科技*(*Tobacco Tech.*), 2002, 5: 23—26
- [103] Ramadan Z, Song X H, Hopke P K. *Anal. Chim. Acta*, 2001, 446: 233—244

- [104] 许祿. 化学计量学方法. 北京: 科学出版社, 1995
- [105] 姜能座(Jiang N Z), 杨德辉(Yang D H). 台湾海峡(Taiwan Strait), 1996, 15(supplement): 100—103
- [106] 齐小明(Qi X M), 张录达(Zhang L D). 北京农学院学报(J. Beijing Agriculture College), 1999, 14: 45—49
- [107] 齐小明(Qi X M), 张录达(Zhang L D). 北京农学院学报(J. Beijing Agriculture College), 1999, 14: 47—52
- [108] Centner V, Massart D L, de Noord O E. Anal. Chem., 1996, 68: 3851—3858
- [109] Koshoubu J, Iwata T, Minami S. Appl. Spectrosc., 2000, 54: 148—152
- [110] Alsberg B K, Kell D B, Godacre R. Anal. Chem., 1998, 70: 4126—4133
- [111] Lindgren F, Celadi P, Wold S. J. Chemom., 1994, 8: 349—363
- [112] Osborne S D, Jordan R B. Analyst, 1997, 122: 1531—1537
- [113] Frank I E. Chemom. Intell. Lab. Syst., 1987, 1: 233—242
- [114] 陈国良, 王煦法, 庄镇泉. 遗传算法及其应用. 北京: 人民邮电出版社, 1996
- [115] 蔡文生(Cai W S), 邵学广(Shao X G), 赵贵文(Zhao G W), 张懋森(Zhang M S). 分析化学(Chin. J. Anal. Chem.), 1997, 25: 231—237
- [116] 邓勃(Deng B), 刘嘉(Liu J). 分析科学学报(Anal. Sci. Acta), 1997, 13: 160—168
- [117] 章元(Zhang Y), 朱尔一(Zhu E Y), 庄峙厦(Zhuang S X), 王小如(Wang X R). 高等学校化学学报(Chem. J. Chin. Univ.), 1999, 20: 1371—1375
- [118] 刘芳(Liu F), 王俊德(Wang J D). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2001, 21: 331—335
- [119] 何险峰(He X F), 周家驹(Zhou J J). 化学进展(Prog. Chem.), 1998, 10: 312—318
- [120] Leardi R. J. Chemom., 2001, 15: 559—569
- [121] Lucasius C B, Kateman G. Chemom. Intell. Lab. Syst., 1993, 19: 1—33
- [122] Hibbert D B. Chemom. Intell. Lab. Syst., 1993, 19: 277—293
- [123] Rimbaud D J, Massart D L, Leardi R, De Noord O E. Anal. Chem., 1995, 67: 4295—4301
- [124] Ding Q, Small G W, Arnold M A. Anal. Chem., 1998, 70: 4472—4479
- [125] Bangalore A S, Shaffer R E, Small G W, Arnold M A. Anal. Chem., 1996, 68: 4200—4212
- [126] Roger J M, Bellon-Maurel V. Appl. Spectrosc., 2000, 54: 1313—1320
- [127] Leardi R. J. Chemom., 2000, 14: 643—655
- [128] Leardi R. Chemom. Intell. Lab. Syst., 1998, 41: 195—207
- [129] Leardi R, Seasholtz M B, Pell R J. Anal. Chim. Acta, 2002, 461: 189—200
- [130] 褚小立(Chu X L), 袁洪福(Yuan H F), 王艳斌(Wang Y B), 陆婉珍(Lu W Z). 分析化学(Chin. J. Anal. Chem.), 2001, 29: 437—442
- [131] 王宏(Wang H), 李庆波(Li Q B), 刘则毅(Liu Z Y). 分析化学(Chin. J. Anal. Chem.), 2002, 30: 779—783
- [132] Smith B M, Gemperline P J. Anal. Chim. Acta, 2000, 423: 167—177
- [133] van den Broke W, Wienke D, Melssen W J. Appl. Spectrosc., 1997, 51: 1210—1217
- [134] Swierenga H, de Groot P J, de Weijer A P. Chemom. Intell. Lab. Syst., 1998, 41: 237—248
- [135] Swierenga H, Wülfert F, de Noord O E. Anal. Chim. Acta, 2000, 411: 121—135
- [136] Molsley M, Williams R. Appl. Spectrosc., 1998, 52: 1197—1202
- [137] Ozdemir D, Molsley M, Williams R. Appl. Spectrosc., 1998, 52: 1203—1209
- [138] Ozdemir D, Williams R. Appl. Spectrosc., 1998, 52: 210—217
- [139] Paradkar R P, Williams R R. Appl. Spectrosc., 1997, 51: 92—97
- [140] Shaffer R E, Small G W, Arnold M A. Anal. Chem., 1996, 68: 2663—2675
- [141] Ding Q, Boyd B L, Small G W. Appl. Spectrosc., 2000, 54: 1047—1054
- [142] Depczynski U, Jetter K, Molt K, Niemüller A. Chemom. Intell. Lab. Syst., 1999, 47: 179—187
- [143] Barros A S, Rutledge D N. Chemom. Intell. Lab. Syst., 1998, 40: 65—81
- [144] Frost V J, Molt K. J. Near Infrared Spectrosc., 1998, 6: 185A—190A
- [145] Gicochea H C, Olivieri A C. J. Chem. Inf. Comput. Sci., 2002, 42: 1146—1153
- [146] Guchardi R, da Costa Filho P A, Poppi R J. J. Near Infrared Spectrosc., 1998, 6(1P4): 333—339
- [147] 朱尔一, 杨梵原. 化学计量学技术与应用. 北京: 科学出版社, 2001
- [148] Brenchley J, Horchner U, Kalivas J. Appl. Spectrosc., 1997, 51: 689—699
- [149] Kalivas J H, Roberts N, Sutter J M. Anal. Chem., 1989, 61: 2024—2030
- [150] Norgaard L, Saudland A. Appl. Spectrosc., 2000, 54: 413—419
- [151] Archibald D D, Akin D E. Vibrational Spectroscopy, 2000, 23: 169—180
- [152] Jiang J H, Berry R J, Siesler H W. Anal. Chem., 2002, 74: 3555—3565
- [153] 高建波(Gao J B), 胡鑫尧(Hu X Y), 胡东成(Hu D C). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2001, 21: 599—602
- [154] McShane M J, Cameron B D, Côté G L. Appl. Spectrosc., 1999, 53: 1575—1581
- [155] McShane M J, Cameron B D, Côté G L. Anal. Chim. Acta, 1999, 388: 251—264
- [156] Araújo M C U, Saldanha T C B, Galvão R K H. Chemom. Intell. Lab. Syst., 2001, 57: 65—73
- [157] Spiegelman C H, McShane M J, Cote G L. Anal. Chem., 1998, 70: 35—44

- [158] Gregory A B, Tracy P H, John H K. *Chemom. Intell. Lab. Syst.*, 1999, 45: 225—239
- [159] de Groot P J, Postma G J, Melssen W J. *Appl. Spectrosc.*, 2001, 55: 173—181
- [160] Blanco M, Coello J, Iturriaga H. *Appl. Spectrosc.*, 1997, 51: 240—246
- [161] Bertran E, Blanco M, Maspoch S, Ortiz M C. *Chemom. Intell. Lab. Syst.*, 1999, 49: 215—224
- [162] Pravdova V, Walczak B, Massart D L. *Anal. Chim. Acta*, 2001, 450: 131—141
- [163] Stordrange L, Libnau F O, Malthe-Sørensen D. *J. Chemom.*, 2002, 16: 529—541
- [164] Swierenga H, Haanstra W G, de Weijer A P. *Appl. Spectrosc.*, 1998, 52: 7—16
- [165] Artursson T, Hagman A, Bjork S. *Appl. Spectrosc.*, 2000, 54: 1222—1230
- [166] Eriksson L, Trygg J, Johansson E. *Anal. Chim. Acta*, 2000, 420: 181—195
- [167] Pizarro Consuelo, Esteban D éz I, González J M. 11th International Conference on NIR, Córdoba, Spain, 2003. 4212
- [168] Olsson, Ralf J O. 5th International Conference on NIR, New York, USA, 1992
- [169] FernándezCaban á V M, GarridoVaro A, PérezMar í M D. 11th International Conference on NIR, Córdoba, Spain, 2003. 4210
- [170] FernándezCaban á V M, GarridoVaro A, PérezMar í M D. 11th International Conference on NIR, Córdoba, Spain, 2003. 4211