

文章编号: 0253-2239(2010)03-0854-07

# 基于 O2-PLS 方法的血糖无损检测实验研究

李亚萍 张广军 李庆波

(北京航空航天大学精密光机电一体化技术教育部重点实验室, 北京 100191)

**摘要** 将 O2-PLS 方法应用于近红外血糖无损检测的基础实验中, 将光谱矩阵和浓度矩阵中的目标因素和非目标因素分离, 只取两矩阵中的目标因素建立校正模型。分别选用成分逐渐复杂的二成份葡萄糖溶液、四成份葡萄糖溶液和人体血浆样品作为研究对象, 采用分组验证法对所建模型进行验证。结果显示, 模型的预测相关系数均达到 0.999, 预测均方根误差(RMSEP)分别为 1.446, 1.931, 2.274 mg/dL (1 dL=0.1 L), 预测精度均高于传统偏最小二乘回归(PLS)预测。而且相对于传统 PLS 模型, O2-PLS 模型的预测精度受样品复杂程度的影响较小。结果表明, O2-PLS 方法更适用于成分复杂的样品检测。

**关键词** 光谱学; 近红外光谱; O2-PLS; 血糖; 无损检测

中图分类号 O657.33 文献标识码 A doi: 10.3788/AOS20103003.0854

## Application of O2-PLS in Experimental Study on Non-invasive Measurement of Blood Glucose

Li Yaping Zhang Guangjun Li Qingbo

(Key Laboratory of Education Ministry in Precision Optical Electromechanical Integration Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

**Abstract** The O2-PLS method was applied to basic research on non-invasive measurement of human blood glucose. The O2-PLS method can respectively divide the spectral matrix and concentration matrix into aim factors and aimless factors, only using aim factors to build calibration model. Two-component glucose solution, four-component glucose solution and human plasma samples are taken as research objects, and test set validation method is used to evaluate prediction results. The results show that prediction correlation coefficients are all 0.999 and root mean square errors of prediction (RMSEP) were 1.446, 1.931, 2.274 mg/dL respectively. The prediction accuracy of O2-PLS models are higher than that of PLS models. Moreover, the complexity of samples has less effect on the prediction accuracy of O2-PLS models compared with traditional partial least squares (PLS) models. The experimental results indicate that O2-PLS method is applicable to complex samples detection.

**Key words** spectroscopy; near-infrared spectra; O2-PLS; blood glucose; non-invasive measurement

## 1 引言

传统的血糖检测需要首先采集患者血液, 然后应用化学方法进行处理, 在给患者带来痛苦的同时, 使得血液疾病的传染机会增加, 而且不能实现连续

的检测。近红外血糖无损检测技术具有无疼痛、无感染以及无耗材等优点, 已成为生物医学检测领域的一个研究热点<sup>[1~3]</sup>。在近红外血糖无损检测中, 血液中多种生化成分对近红外光的强烈吸收, 严重掩盖光谱中葡萄糖的信息; 另外, 仪器的状态和测量

收稿日期: 2009-04-27; 收到修改稿日期: 2009-06-16

基金项目: 国家自然科学基金(60708026), 长江学者和创新团队发展计划(IRT0705)和北京市优秀人才培养基金(20081D1600600348)资助课题。

作者简介: 李亚萍(1980—), 女, 博士研究生, 主要从事生物光谱检测和光学精密测量等方面的研究。

E-mail: liyaping\_mail@sina.com

导师简介: 张广军(1965—), 男, 教授, 博士生导师, 主要从事光电测量与动态跟踪、光机电一体化、计算机视觉与模式识别、生物光谱检测等方面的研究。E-mail: gjzhang@buaa.edu.cn(通信联系人)

的环境等均会随时间发生波动和变化,这些因素都会影响血糖无损检测的精确度<sup>[4~6]</sup>。目前,一些简单的近红外数据处理方法通常在消除以上非目标因素的同时,也不可避免地去除了与被测量有关的目标因素,对近红外血糖无损检测的精确度提高意义不大<sup>[7~9]</sup>。O2-PLS 是一种新型的回归建模方法<sup>[10,11]</sup>,与传统的方法相比,该方法不仅能去除光谱数据中的干扰信号,而且可以去除浓度数据中的干扰信号,并尽可能地保持两矩阵中最大的有效信息。目前尚无使用 O2-PLS 测定液体中葡萄糖浓度的报道。本文选用成分逐渐复杂的人体血糖检测基础实验样品——二成份葡萄糖溶液(只包含葡萄糖和水)、四成份葡萄糖溶液(包含牛血红蛋白、牛白蛋白、葡萄糖和水)以及人体血浆为研究对象,采用 O2-PLS 方法建立葡萄糖与其近红外光谱之间的校正模型,计算分析结果满意,可望推广应用于人体近红外血糖无损检测。

## 2 算法简介

O2-PLS 方法的基本思想为<sup>[10,11]</sup>:利用正交信号校正(OSC)分别将光谱矩阵  $\mathbf{X}_{n \times p}$  和响应矩阵  $\mathbf{Y}_{n \times q}$  分成两部分,一部分作为有用信息用来建模,另一部分作为干扰信息被滤掉。其中  $n$  为样本个数, $p, q$  分别为光谱矩阵和响应矩阵变量维数。这样 O2-PLS 模型可以写成

$$\mathbf{X} = \mathbf{T}\mathbf{W}^T + \mathbf{T}_{Y\text{-ortho}}^T + \mathbf{P}_{Y\text{-ortho}}^T \mathbf{E}, \quad (1)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^T + \mathbf{U}_{X\text{-ortho}} \mathbf{P}_{X\text{-ortho}}^T + \mathbf{F}, \quad (2)$$

$$\mathbf{T} = \mathbf{U}\mathbf{B}_U + \mathbf{H}_{TU}, \quad (3)$$

$$\mathbf{U} = \mathbf{T}\mathbf{B}_T + \mathbf{H}_{UT}, \quad (4)$$

$$\mathbf{Y}_R = \mathbf{T}\mathbf{B}_T \mathbf{C}^T, \quad (5)$$

$$\mathbf{X}_R = \mathbf{U}\mathbf{B}_U \mathbf{W}^T, \quad (6)$$

(1)式和(2)式中, $\mathbf{T}$ 为矩阵 $\mathbf{X}$ 中有用成分的得分矩阵, $\mathbf{W}$ 为其相应的载荷矩阵; $\mathbf{T}_{Y\text{-ortho}}$ 为矩阵 $\mathbf{X}$ 中与 $\mathbf{Y}$ 正交成分的得分矩阵, $\mathbf{P}_{Y\text{-ortho}}$ 为其相应的载荷矩阵; $\mathbf{U}$ 为矩阵 $\mathbf{Y}$ 中有用成分的得分矩阵, $\mathbf{C}$ 为其相应的载荷矩阵; $\mathbf{U}_{X\text{-ortho}}$ 为矩阵 $\mathbf{Y}$ 中与 $\mathbf{X}$ 正交成分的得分矩阵, $\mathbf{P}_{X\text{-ortho}}$ 为其相应的载荷矩阵。 $\mathbf{E}$ 和 $\mathbf{F}$ 分别为矩阵 $\mathbf{X}$ 和矩阵 $\mathbf{Y}$ 的残差矩阵。矩阵 $\mathbf{X}$ 和矩阵 $\mathbf{Y}$ 的内在联系由(3)式和(4)式给出,矩阵 $\mathbf{X}$ 和矩阵 $\mathbf{Y}$ 的预测 $\mathbf{X}_R$ 和 $\mathbf{Y}_R$ 分别由(5)式和(6)式给出。在矩阵 $\mathbf{T}_{Y\text{-ortho}}$ 和 $\mathbf{U}_{X\text{-ortho}}$ 中各列相互正交,但是在它们相应的载荷矩阵 $\mathbf{P}_{Y\text{-ortho}}$ 和 $\mathbf{P}_{X\text{-ortho}}$ 中各列并不相互正交。

## 3 材料与方法

### 3.1 实验材料

实验中采用二成份葡萄糖溶液样品、四成份葡萄糖溶液样品和人体血浆样品作为研究对象,并将这三种样品的近红外光谱作为实验数据。

二成份葡萄糖溶液样品共 40 份,葡萄糖质量浓度范围为 35~1215 mg/dL(1 dL=0.1 L),分布均匀,浓度标准差为 351.392 mg/dL。随机选取 30 份样品组成校正集,其余的 10 份样品组成测试集。该样品的配置采用稀释的方法,其配置步骤为:首先配置母液,并将母液放入容量瓶中,加去离子水至 1000 mL,然后使用移液管提取每个样品所需的母液体积,转移到容量瓶中,再次使用去离子水稀释至 100 mL。

四成份葡萄糖溶液样品共 44 份,葡萄糖质量浓度范围为 100~5000 mg/dL,分布均匀,浓度标准差为 1455.726 mg/dL。随机选取 33 份样品组成校正集,其余的 11 份样品组成预测集。该样品的配置过程与二成份葡萄糖溶液的配置过程完全相同。

人体血浆样品共 39 份,葡萄糖质量浓度范围为 10.4~44.4 mg/dL,分布均匀,浓度标准差为 8.520 mg/dL。随机选取 30 份样品组成校正集,其余的 9 份样品组成预测集。该样品的配置步骤为:在全血中加入肝素抗凝剂,在离心机中经过 1500 r, 10 min 后分离出血浆,并在血浆中加入葡萄糖使其具有一定浓度变化的血糖值。

### 3.2 实验方法

采用美国 Perkin-Elmer 公司 Spectrum GX 傅里叶变换红外光谱仪、液态氮冷却的 InSb 检测器、1 mm 石英样品池、蠕动泵自动进样系统和恒温装置进行光谱测量。首先扫描空样品池作为背景,然后分别测量以上三种样品溶液。光谱测量范围为 4000~10000  $\text{cm}^{-1}$ ,扫描分辨率为 1  $\text{cm}^{-1}$ ,扫描温度为 25  $^{\circ}\text{C}$ 。每次测量均扫描 16 次后取平均,以降低随机噪声。随机抽取被测样品,而不按照成份含量的高低顺序进行测量。

## 4 结果与讨论

### 4.1 样品的近红外吸收光谱

纯水中的 O—H 键伸缩振动的一级基频区在 5155  $\text{cm}^{-1}$  附近,合频区在 6940  $\text{cm}^{-1}$  附近,这两个波长附近是水分吸收的敏感区,选择光谱波长范围应尽量避开纯水吸收峰的特征波长区。又因为葡萄糖的合频吸收峰在 4750, 4400, 4300  $\text{cm}^{-1}$  附近,所

以本研究中选择了  $4800\sim 4200\text{ cm}^{-1}$  波数区间作为研究光谱区。葡萄糖在研究光谱区的吸收系数曲线如图 1 所示。图 2(a)、图 3(a)和图 4(a)分别为二成份葡萄糖溶液样品、四成份葡萄糖溶液样品和人体血浆样品的原始吸光度光谱图。从图 2(a)中可以看出,虽然不同样品的葡萄糖质量浓度存在较大差别,但其谱线却几乎完全重合,这主要是因为样品中水的含量占绝对优势,严重掩盖了葡萄糖的信息。图 3(a)和图 4(a)的情况与图 2(a)类似。图 2(b)、图 3(b)和图 4(b)分别为以上三种样品经 O2-PLS 方法校正后的吸光度光谱图。可以看出,每个图中样品光谱产生明显的差异信息,在糖的吸收波长位置均有吸收峰,每条谱线均明显地反映了葡萄糖的吸收信息。

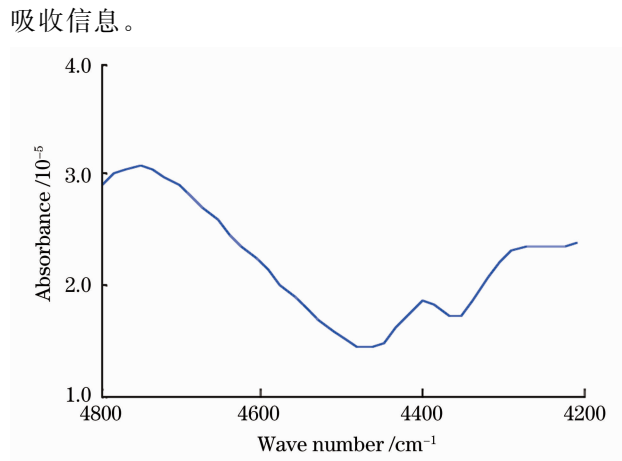


图 1 葡萄糖吸收系数曲线

Fig. 1 Absorption-coefficient curve of glucose

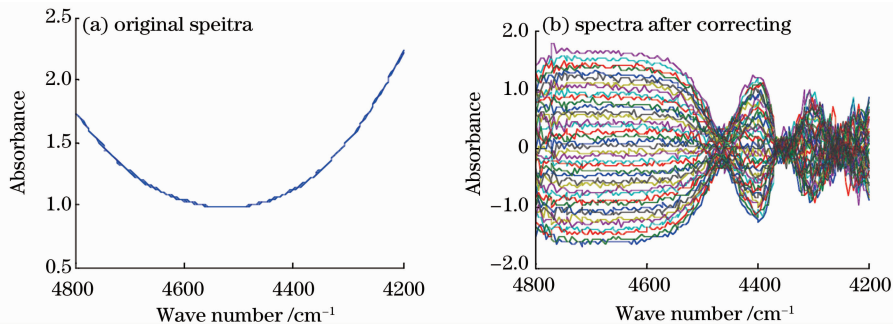


图 2 二成份葡萄糖溶液样品吸光度光谱图

Fig. 2 Absorption spectra of two-component glucose solution

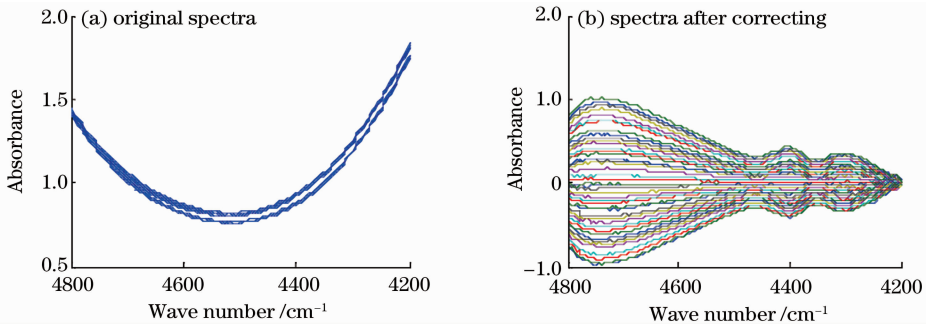


图 3 四成份葡萄糖溶液样品吸光度光谱图

Fig. 3 Absorption spectra of four-component glucose solution

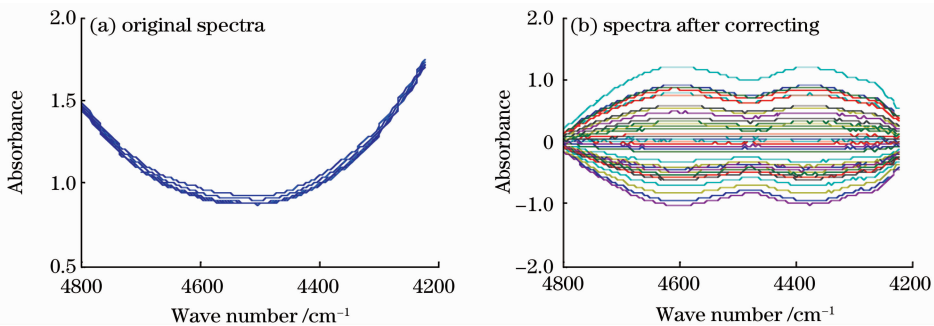


图 4 人体血浆样品吸光度光谱图

Fig. 4 Absorption spectra of human plasma

## 4.2 模型预测能力比较与分析

应用 O2-PLS 方法分别建立二成份葡萄糖溶液样品、四成份葡萄糖溶液样品以及人体血浆样品的多变量校正模型,通过分析其校正模型的性能参数来考察 O2-PLS 方法的有效性。同时,还选用传统的 PLS 方法建模,以作对照比较。

图 5(a),图 5(b)分别为二成份葡萄糖溶液样品

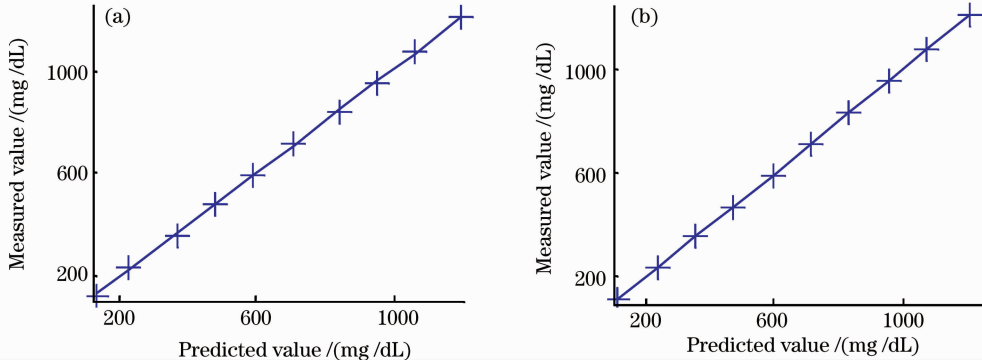


图 5 二成份葡萄糖溶液样品预测结果。(a) PLS 模型; (b) O2-PLS 模型

Fig. 5 Predictions of two-component glucose solution. (a) PLS model; (b) O2-PLS model

表 1 PLS 模型与 O2-PLS 模型的性能参数比较

Table 1 Comparison of performance parameters between PLS model and O2-PLS model

	PLS model			O2-PLS model		
	RMSEP /(mg/dL)	R	factor	RMSEP /(mg/dL)	R	factor
Two-component glucose solution	3.701	0.982	3	1.446	0.999	1
Four-component glucose solution	7.653	0.926	5	1.931	0.999	1
Human plasma	16.600	0.750	7	2.274	0.999	1

图 6 为四成份葡萄糖溶液样品的预测结果。如表 1 所示,四成份葡萄糖溶液样品 PLS 模型的相关系数  $R$  为 0.926,预测均方根误差为 7.653 mg/dL。该样品 O2-PLS 模型的相关系数  $R$  为 0.999,比 PLS 模型的相关系数提高 7.307%;预测均方根误差为 1.931 mg/dL,比 PLS 模型的预测均方根误差降低 74.768%。

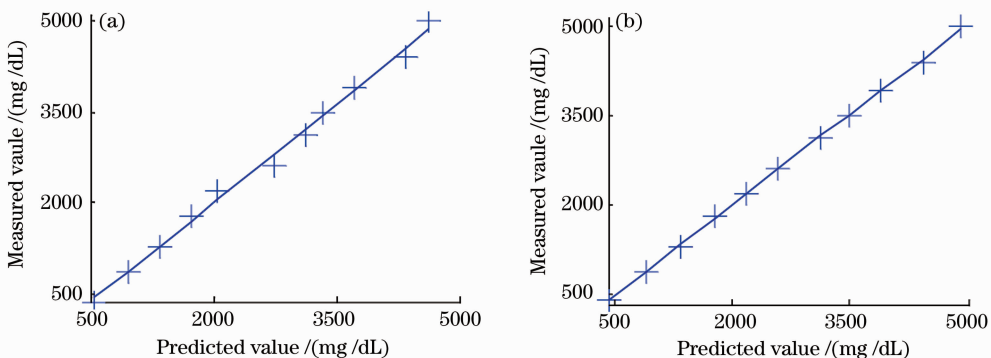


图 6 四成份葡萄糖溶液样品预测结果。(a) PLS 模型; (b) O2-PLS 模型

Fig. 6 Predictions of four-component glucose solution. (a) PLS model; (b) O2-PLS model

PLS 模型和 O2-PLS 模型的预测结果。如表 1 所示,二成份葡萄糖溶液样品 PLS 模型的相关系数  $R$  为 0.982,预测均方根误差 (RMSEP) 为 3.701 mg/dL。该样品 O2-PLS 模型的相关系数  $R$  为 0.999,比 PLS 模型的相关系数提高 1.731%;预测均方根误差为 1.446 mg/dL,比 PLS 模型的预测均方根误差降低 60.930%。

图 7 为人体血浆样品的预测结果。如表 1 所示,人体血浆样品 PLS 模型的相关系数  $R$  为 0.750,预测均方根误差为 16.600 mg/dL。该样品 O2-PLS 模型的相关系数  $R$  为 0.999,比 PLS 模型的相关系数提高 33.2%;预测均方根误差为 2.274 mg/dL,比 PLS 模型的预测均方根误差降低 86.301%。

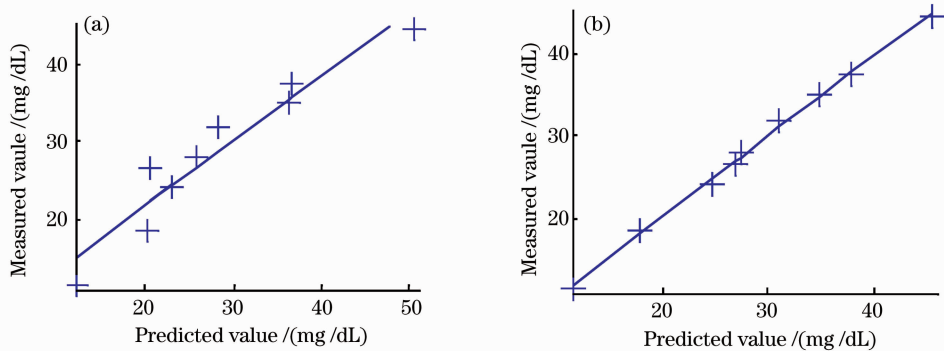


图 7 人体血浆样品预测结果。(a)PLS 模型；(b)O2-PLS 模型

Fig. 7 Predictions of human plasma. (a)PLS model; (b)O2-PLS model

表 1 为 PLS 模型与 O2-PLS 模型性能参数的比较。从表 1 可以看出, O2-PLS 模型的预测精度明显高于 PLS 模型。同时还可以看出, 当样品溶液的成分比较简单时, PLS 模型预测值和测定值之间具有良好的线性关系, 其预测效果与 O2-PLS 模型相差不大。然而随着样品复杂度的增加, PLS 模型相关系数  $R$  会逐渐变小, 预测均方根误差则会逐渐变大。相反, O2-PLS 模型的性能参数受样品复杂程度的影响较小, 而且模型的性能参数值均优于 PLS 模型所对应的参数值。上述分析说明, O2-PLS 方法具有更精确的建模效果, 更适合应用于成分复杂的样品检测。

### 4.3 模型载荷曲线比较与分析

在主成分分析 (PCA) 中, 载荷反映的是波长变量与主成分的相关性。波长变量在某个主成分中的载荷值越大, 说明它与该主成分的相关性越大<sup>[12]</sup>。利用主成分载荷曲线的峰值波长位置则可判断出该

主成分所表征的组分信息。图 8~图 10 分别为二成分葡萄糖溶液样品、四成分葡萄糖溶液样品以及人体血浆样品的主成分载荷曲线。在这些曲线中, 如果某一个主成分载荷曲线的峰值波长等于或者接近葡萄糖的吸收波长, 则可判断出该主成分表征的是葡萄糖的信息, 进而可利用该主成分的贡献率来解释 O2-PLS 模型预测精度高于 PLS 模型的原因。

图 8(a) 为二成份葡萄糖溶液样品 PLS 模型的主成分载荷曲线。该样品成份比较简单, 根据预测残差平方和 (PRESS) 方法确定的最佳主成分的个数为 3 个, 其中主成分 1 的载荷曲线与图 1 所示的葡萄糖吸收系数曲线的波峰位置相同, 因此可以判断该成分表征的是葡萄糖信息。该成分对光谱信息的贡献率为 79%。图 8(b) 为二成份葡萄糖溶液样品 O2-PLS 模型的主成分载荷曲线。最佳主成分的个数只有 1 个, 由载荷曲线的波峰位置同样可以判断出该成分表征的是葡萄糖信息, 该成分对光谱信息的贡献率为 100%。

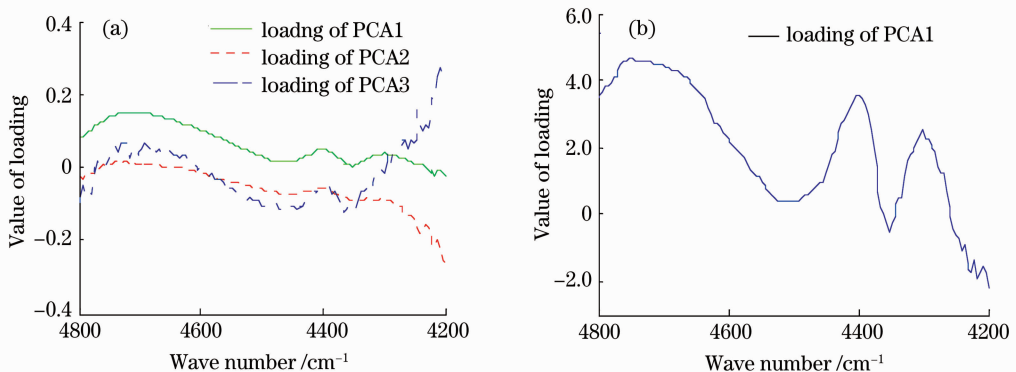


图 8 二成份葡萄糖溶液样品的主成分载荷曲线。(a)PLS 模型；(b)O2-PLS 模型

Fig. 8 Principal-component loadings of two-component glucose solution. (a)PLS model; (b)O2-PLS model

图 9(a) 为四成份葡萄糖溶液样品 PLS 模型的主成分载荷曲线。该样品成份比二成份葡萄糖溶液复杂, 根据 PRESS 方法确定的最佳主成分个数为 5 个, 图中只取了前 3 个主成分的载荷曲线。由图中

可以看出, 主成分 2 的载荷曲线与葡萄糖吸收系数曲线的波峰位置相同, 因此可以判断该成分表征的是葡萄糖信息。该成分对光谱信息的贡献率为 7%。图 9(b) 为四成份葡萄糖溶液样品 O2-PLS 模

型的主成分载荷曲线。最佳主成分的个数只有 1 个,由载荷曲线的波峰位置同样可以判断出该成分

表征的是葡萄糖信息,该成分对光谱信息的贡献率为 100%。

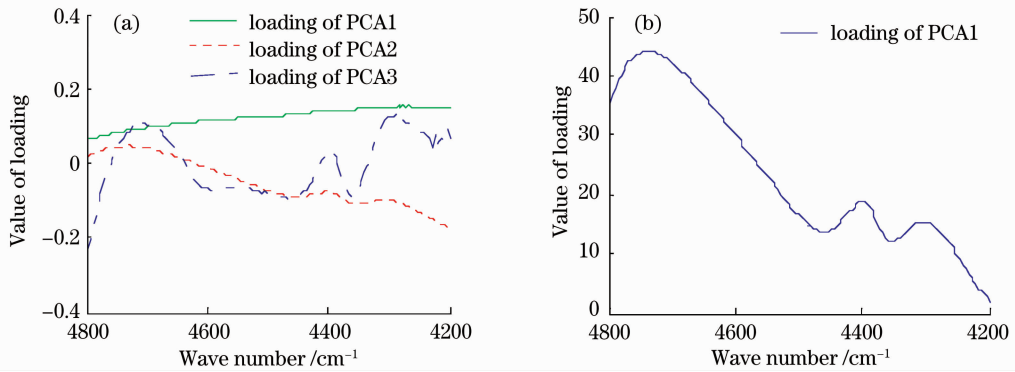


图 9 四成份葡萄糖溶液样品的主成分载荷曲线。(a)PLS 模型；(b)O2-PLS 模型

Fig.9 Principal-component loadings of four-component glucose solution. (a)PLS model; (b)O2-PLS model

图 10(a)为人体血浆样品 PLS 模型的主成分载荷曲线。该样品成份比前两种样品更加复杂,根据 PRESS 方法确定的最佳主成分个数为 7 个,图中只取了前 3 个主成分的载荷曲线。由图中可以看出,主成分 3 的载荷曲线与葡萄糖吸收系数曲线的波峰位置相同,因此可以判断该成分表征的是葡萄糖信

息。该成分对光谱信息的贡献率为 1%。图 10(b)为人体血浆样品 O2-PLS 模型的载荷曲线。最佳的主成分个数只有 1 个,由载荷曲线的波峰位置同样可以判断出该成分表征的是葡萄糖信息,该成分对光谱信息的贡献率为 100%。

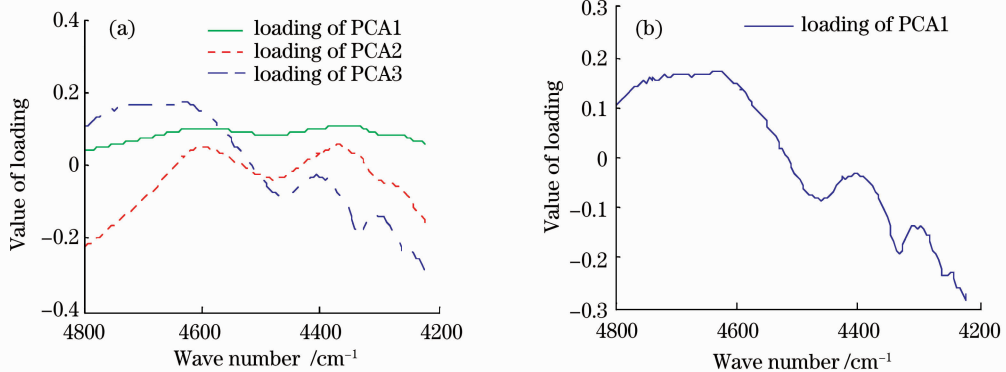


图 10 人体血浆样品的主成分载荷曲线。(a)PLS 模型；(b)O2-PLS 模型

Fig. 10 Principal-component loadings of human plasma. (a)PLS model; (b)O2-PLS model

由以上分析可以看出,样品成分越复杂,建立 PLS 模型所需的主成分个数越多,表征葡萄糖信息的主成分对光谱的贡献率就越低,因而随着样品复杂程度的增加,PLS 模型的预测精度会逐渐减小。然而,O2-PLS 模型的主成分个数不受样品复杂程度的影响,且均为 1 个。第一主成分对光谱信息的贡献率均为 100%,完全反映了葡萄糖的信息,因此 O2-PLS 模型的预测精度较 PLS 模型有显著提高。

还选用传统的 PLS 方法建模,以作对照比较。结果表明,随着样品复杂程度的增加,代表葡萄糖信息的成分对光谱信息的贡献率逐渐减小,因而造成 PLS 模型的预测值与测定值相关系数逐渐减小,预测均方根误差逐渐变大。O2-PLS 模型由于剔除了与样品成分浓度含量无关的主成分,所提取的成分对浓度数据有最大的贡献率,因而 O2-PLS 校正模型预测精度更高,更加适用于复杂成分样品。

### 5 结 论

选用了三种复杂程度递增的样品作为近红外血糖无损检测基础实验研究的对象,应用 O2-PLS 回归建模的方法分别对三种样品建立校正模型,同时

致谢 感谢天津大学精密仪器与光电子工程学院生物光学与应用光谱研究组在实验环节和数据共享中对我们的帮助和支持。



## 参 考 文 献

- 1 He Zhonghai, Luo Yunhan, Gu Xiaoyu *et al.*. A Study of blood glucose concentration measurement by scattering coefficient[J]. *Acta Optica Sinica*, 2006, **26**(4): 591~594  
贺忠海, 罗云瀚, 谷筱玉等. 利用散射系数变化测量血糖浓度的研究[J]. *光学学报*, 2006, **26**(4): 591~594
- 2 Li Qingbo, Zhang Lu, Wang Yan *et al.*. Near-infrared spectropic measurement of glucose concentrations in an aqueous matrix-Fundamental study on noninvasive measurement method of human blood glucose concentration [J]. *Chinese Journal of Biomedical Engineering*, 2004, **23**(6): 602~606  
李庆波, 张 鲁, 汪 曦等. 葡萄糖水溶液浓度的近红外光谱测量法——人体血糖浓度无创伤测量方法基础研究[J]. *中国生物医学工程学报*, 2004, **23**(6): 602~606
- 3 Xu Kexin, Gao Feng, Zhao Huijuan. Biomedical Photonics[M]. Beijing: Science Publishing Company, 2007. 145~177  
徐可欣, 高 峰, 赵会娟. *生物医学光子学*[M]. 北京: 科学出版社, 2007. 145~177
- 4 Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, **26**(1): 136~140  
赵杰文, 张海东, 刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. *光学学报*, 2006, **26**(1): 136~140
- 5 Lü Lina, Liu Rong, Zhou Dingwen. Application of O-PLS in fundamental study of non-invasive measurement of human blood glucose concentration with near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(12): 1950~1954  
吕丽娜, 刘 蓉, 周定文. O-PLS在近红外无损人体血糖浓度测量基础研究中的应用[J]. *光谱学与光谱分析*, 2005, **25**(12): 1950~1954
- 6 Han Donghai, Wang Jiahua. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy[J]. *Chinese J. Lasers*, 2008, **35**(8): 1123~1131  
韩东海, 王加华. 水果内部品质近红外光谱无损检测研究进展[J]. *中国激光*, 2008, **35**(8): 1123~1131
- 7 Svante Wold, Henrik Antti, Fredrik Lindgren *et al.*. Orthogonal signal correction of near-infrared spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 1998, **44**: 175~185
- 8 Johan Trygg, Svante Wold. Orthogonal projections to latent structures(O-PLS)[J]. *J. Chemometrics*, 2002, **16**: 119~128
- 9 Chu Xiaoli, Yuan Hongfu, Lu Wangzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, **16**(4): 528~542  
褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, **16**(4): 528~542
- 10 Johan Trygg. O2-PLS for qualitative and quantitative analysis in multivariate calibration [J]. *J. Chemometrics*, 2002, **16**: 283~293
- 11 Johan Trygg, Svante Wold. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter [J]. *J. Chemometrics*, 2003, **17**: 53~64
- 12 Yuan Weiqi, Qu Xiaofeng, Ke Li *et al.*. PCA reconstruction error palmprint recognition [J]. *Acta Optica Sinica*, 2008, **28**(10): 1903~1909  
苑玮琦, 曲晓峰, 柯 丽等. 主成分分析重建误差掌纹识别方法[J]. *光学学报*, 2008, **28**(10): 1903~1909