

经验模态分解法在近红外无创血红蛋白检测中的应用研究

樊奕辰^{1,2}, 卢启鹏^{1*}, 丁海泉¹, 高洪智¹, 陈星旦¹

1. 中国科学院长春光学精密机械与物理研究所, 应用光学国家重点实验室, 吉林 长春 130033

2. 中国科学院研究生院, 北京 100049

摘要 为提高近红外血红蛋白预测模型的稳健性, 分别应用 Savitzky-Golay 平滑、移动窗口平滑以及经验模态分解(EMD)方法对原始光谱进行去噪处理, 以提高数据信噪比。采集了 81 例临床志愿者的手指指端血流容积脉搏波光谱数据, 同时获取相应的血红蛋白浓度值临床化验结果。剔除异常样品, 确定 78 例样品为研究对象, 建立反向传播神经网络(BP-ANN)定量分析模型并预测。结果表明, 经 EMD 处理后的模型预测效果最优, 预测相关系数由 0.74 提高至 0.87, 误差均方根由 $12.85 \text{ g} \cdot \text{L}^{-1}$ 减小至 $8.08 \text{ g} \cdot \text{L}^{-1}$ 。实验证明应用 EMD 方法能够获得高信噪比的容积脉搏信号, 提高血红蛋白浓度预测模型的准确性, 有利于推动近红外无创血红蛋白检测技术的进一步发展。

关键词 近红外光谱; 血红蛋白浓度; 经验模态分解; 无创检测

中图分类号: O657.3 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2013)02-0349-05

引言

血红蛋白(Hemoglobin, Hb)是血液运输氧的载体。人体血红蛋白浓度是诊断贫血症(Hb 偏低)以及真性血红蛋白增多症(Hb 偏高)的可靠依据之一^[1,2]。目前临床采用的血红蛋白测定方法为有创方法, 需要对患者采血, 因而有感染风险, 且血液采集与分析过程之间存在延时, 难以对重症患者进行实时监测^[1-4]。因此, 实现无创血红蛋白检测具有非常重要的临床意义。

近红外光谱分析技术(near infrared spectroscopy, NIRS)具有无损快速、无需试剂、多组分同时分析等特点^[3], 且近红外光对人体组织的穿透能力较强, 已成为无创血红蛋白检测研究领域关注的热点之一^[1,4,5]。然而, 由于人体各种生理组织背景信息混杂, 对近红外血液光谱干扰严重, 导致近红外无创血红蛋白测量精度尚未达到临床要求^[1,6,7]。

为消除人体复杂的生理组织背景干扰, 陈星旦^[6]提出血流容积光谱相减方法。通过将不同血流容积下测得的两幅光谱相减, 可扣除皮肤、肌肉等组织背景的干扰, 从而提取血流容积变化量对应于血液成分信息的“纯净”光谱。光谱相减法虽能去除背景差异等影响, 获得有效光谱信号, 但同时也

会导致信号强度降低^[5,6]。因此, 选择合适的方法去除原始数据噪声, 获得高信噪比的脉搏波信号是实现血流容积光谱相减方法的重要前提。目前, 光谱去噪常用的算法主要有 Savitzky-Golay 平滑、移动窗口平滑、傅里叶变换和小波变换等。其中平滑方法可以有效去除高频噪声, 提高数据信噪比, 但容易造成信号失真; 傅里叶变化方法以信号不变的频率特性为前提, 主要应用于平稳信号的去噪处理; 小波变换方法虽然对非平稳信号有良好的去噪效果, 但运算复杂, 实现速度较慢^[8]。经验模态分解(empirical mode decomposition, EMD)方法是由 Huang 等^[9,10]于 1998 年提出的一种数据消噪算法, 具有分解模态少、不用选择基函数等优点, 已越来越多地应用于非线性、非平稳信号的线性化和平稳化处理。

为解决 NIRS 无创血红蛋白检测中噪声干扰的问题, 以平滑算法为参照, 重点研究了应用 EMD 算法去除光谱噪声的可行性及有效性。无创伤测量了 81 名志愿者的指端近红外光谱, 考察了不同去噪方法对血流容积脉搏信号的噪声抑制能力; 对比分析了数据预处理前后 NIRS 血红蛋白浓度模型的预测精度, 讨论了 EMD 算法应用于脉搏光谱数据去噪处理的优点。实验结果为提高无创血红蛋白检测精度提供一种新的技术途径。

收稿日期: 2012-06-01, 修订日期: 2012-08-20

基金项目: 国家自然科学基金项目(60938002), 应用光学国家重点实验室开放基金项目(Y1Q03FQ113)和中国科学院知识创新工程领域前沿项目(Y00232Q100)资助

作者简介: 樊奕辰, 1988 年生, 中国科学院长春光学精密机械与物理研究所硕士研究生 e-mail: fsmfyc@gmail.com

* 通讯联系人 e-mail: luqipeng@126.com

1 实验部分

1.1 仪器

基于动脉血液对光的吸收量随动脉搏动而变化的原理,设计搭建了无创血红蛋白测量系统,利用透射检测方式采集

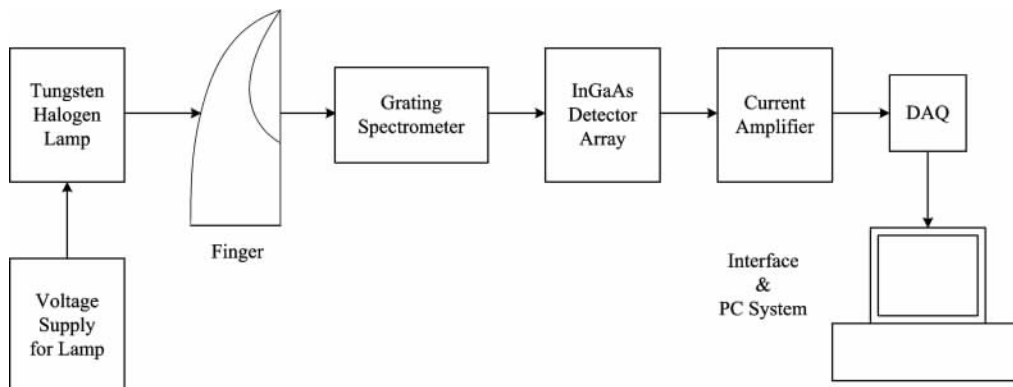


Fig 1 Noninvasive hemoglobin measurement system

1.2 测量对象及过程

实验测量对象为参加常规体检的 81 名受检者,其中男性 46 人,女性 35 人,空腹。检测时测量对象食指指端完全遮盖测量系统的数据采集孔,测量过程中接触压力保持基本稳定,采集时间 30 s,数据保存至计算机。每采集完一组血流容积脉搏波数据,立即由医院专业人员对被测者进行静脉抽血经生化分析获取其血液参数临床化验值。

1.3 异常样品剔除

所谓异常样品,是指由于操作不当、测量环境改变等各种原因导致的光谱数据值或临床分析值存在较大误差的样品。异常样品的存在将严重影响 NIRS 血红蛋白浓度模型的预测精度,应该将其剔除。本文采用主成分分析对采集的血液容积脉搏波数据进行分析,在样品光谱矩阵的主成分空间利用改进的马氏距离算法剔除异常样品,由图 2 所示。在 81 份样品中,18, 35, 81 号样品有较大的得分距离(score distance, SD)或正交距离(orthogonal distance, OD),判别为异常样品。

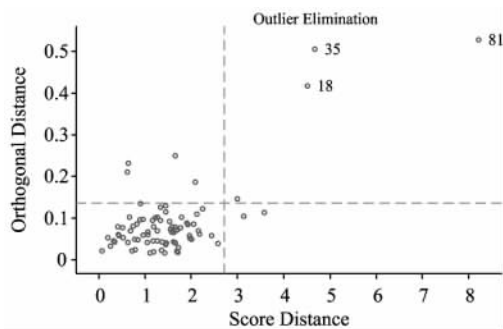


Fig 2 Distances of PC subspace

1.4 定标集与预测集划分

剔除异常样品后剩余 78 份容积脉搏波数据,根据浓度

手指指端吸光特性曲线,结构如图 1 所示。光源采用 75 W 卤钨灯;分光部分选用光栅光谱仪;探测系统采用日本 HAMAMATSU 公司的 16 元 InGaAs 光电探测器;光谱范围为 1 100~1 400 nm;数据采集部分选择美国 NI 公司的数据采集卡 NI6281。

范围均匀分布的原则将其划分为定标集与预测集,其中定标集样品 61 份,预测集样品 17 份,如表 1 所示。

Table 1 Statistic properties of calibration set and prediction set

Statistic property	Calibration set	Prediction set
Number of samples	61	17
Minimum/(g · L ⁻¹)	113	122
Maximum/(g · L ⁻¹)	180	176
Mean/(g · L ⁻¹)	150.75	150.41
Standard deviation/(g · L ⁻¹)	16.06	16.07

1.5 EMD 算法原理^[9, 10]

EMD 算法能够将复杂信号分解为有限个表征原始信号特征时间尺度的固有模态函数(intrinsic mode function, IMF),通过去除反映噪声信息的 IMF 分量实现去噪目的。每个 IMF 分量都必须满足两个条件:(1)整个数据段内的极值点与零点个数相等或相差 1;(2)信号关于时间轴对称。

对给定信号 $x(t)$ 进行 EMD 去噪的具体步骤如下:(1)确定 $x(t)$ 所有极值点,分别应用三次样条曲线连接极大值与极小值点,得到 $x(t)$ 的上、下包络线;(2)计算上下包络线的平均值 m_1 ,求取其与 $x(t)$ 之差 h_1 : $h_1 = x(t) - m_1$;(3)判断 h_1 是否符合 IMF 条件:若是,则直接将 h_1 赋值给 c_1 ;若不是,则将 h_1 作为新的 $x(t)$ 重复上两步操作,至新的 h_1 符合 IMF 条件,记作 c_1 ;所得的 c_1 即为 $x(t)$ 的第一个 IMF 分量;(4)将 $x(t)$ 与 c_1 的差值 $r_1 = x(t) - c_1$ 作为新的 $x(t)$,重复上述操作,至第 n 阶差值 r_n 为单调或直流信号为止。

由此可知,EMD 算法能够将任一复杂信号 $x(t)$ 分解为 n 个 IMF 分量与残余信号的和

$$x(t) = \sum_{i=1}^n c_i + r_n \quad (1)$$

各 IMF 分量代表信号从高到低不同频率段的成分, 其中阶数小的 IMF 对应于数据的高频成分, 一般包括信号的高频部分以及噪声; 阶数大的 IMF 对应于信号的低频成分, 受噪声影响很小。

2 结果与讨论

2.1 脉搏信号的平滑去噪处理

分别采用 S-G 平滑以及移动窗口平滑算法对采集到的容积脉搏信号进行去噪处理。图 3(a) 为某一被测人员中心波长为 1 552 nm 处的手指指端吸光特性曲线, 受噪声影响比较严重。通过调整平滑算法的多项式次数及滤波带宽, 确定

7 点 2 次 S-G 滤波模型以及 5 点移动窗口均值滤波效果最佳, 去噪后的脉搏波光谱如图 3(b) 和 (c) 所示。对比图 3(a) 可知, 去噪后的光谱平滑性显著增强, S-G 平滑以及移动窗口平滑算法均能有效去除原始数据中的高频噪声。然而, 经平滑去噪后的信号, 尤其是移动窗口平滑后的信号较原始信号有所失真, 算法在去除数据噪声的同时也损失了原始数据中的部分有用信息。

2.2 脉搏信号的 EMD 去噪处理

为有效去除数据噪声并保留原始数据的特性, 采用 EMD 算法对图 3(a) 中的容积脉搏信号进行去噪处理。对该信号进行 EMD 分解, 自适应地得到 6 阶固有模态函数 IMF1 ~ IMF6 和残余函数 RES, 如图 4 所示。

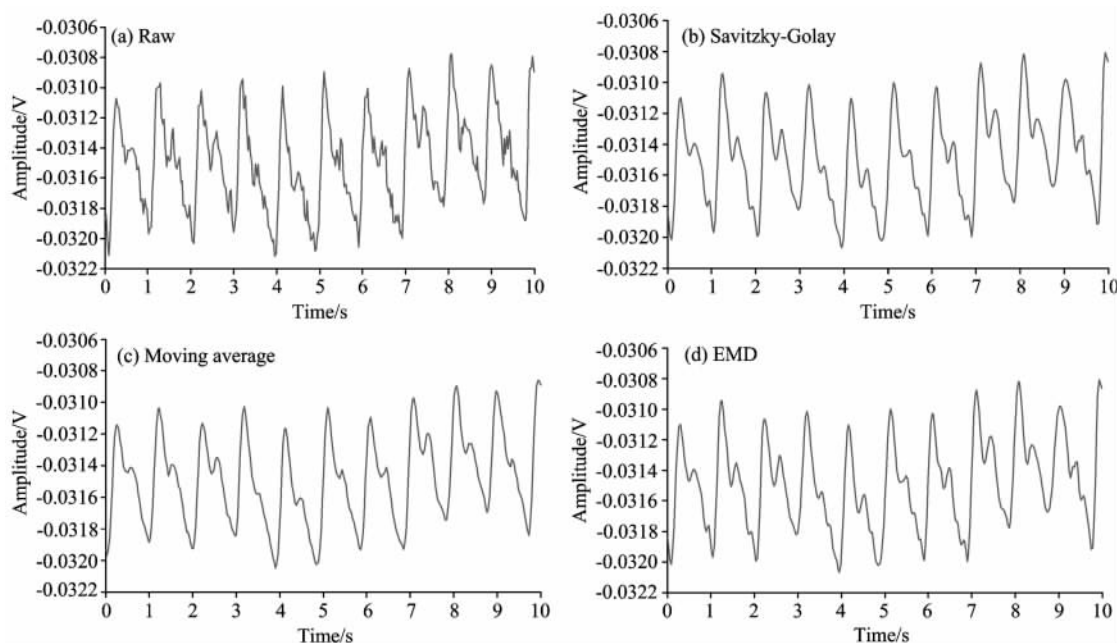


Fig 3 Original and denoised signal of single wave absorption curve

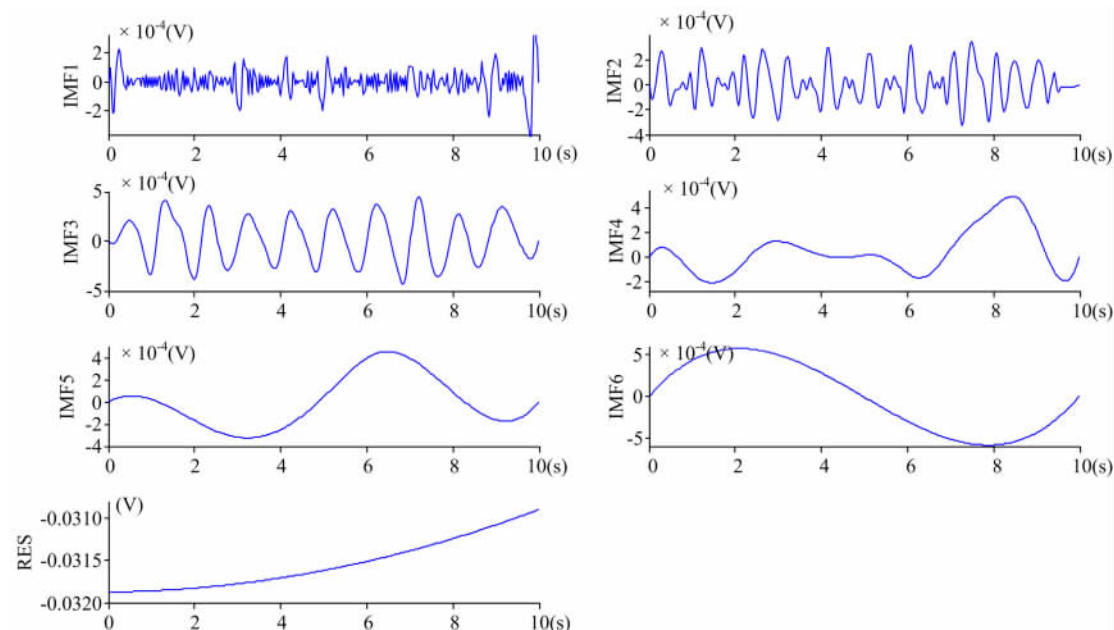


Fig 4 Decomposition of fingertip absorption curve

从图 4 可以看出, IMF1 分量频率较高且幅值较小, 主要由噪声引起, 而 IMF2~IMF4 主要表现为信号的高频分量。为有效去除噪声并保留数据本身的特性, 利用 EMD 分解的时空尺度滤波特性, 对 IMF1 分量进行滤波, 将 IMF2~IMF6 及残余分量 RES 按式(1)重构, 得到去噪后的血流容积脉搏信号, 如图 3(d)所示。对比图 3(a), 去噪后的光谱平滑性增强, 噪声明显降低; 且重构信号和原始信号极为相似, 去噪后的信号失真很小, 保留了数据本身的特性, 效果优于平滑算法。

2.3 BP-ANN 模型的建立

应用 BP-ANN 方法, 建立 NIRS 血红蛋白浓度的定量分析模型, 对比不同数据预处理方法对模型的优化能力。人工神经网络是一种应用类似于大脑神经突触联接的结构进行信息处理的数学模型, 具有高度非线性逼近能力, 目前应用最多的是 BP 人工神经网络。BP 神经网络一般由 3 个神经层组成, 即输入层、输出层和隐含层。数据由输入层输入, 施以权重传输到隐含层, 经过权值、阈值和激励函数运算后, 由输出层给出神经网络的预测值。网络经训练后误差达

最小值^[11]。

建立了一个三层的 BP 人工神经网络结构, 隐含层采用对数 Sigmoid 型传递函数, 输出层采用线性传递函数, 同时选用量化共轭梯度法 trainscg 函数训练网络。以 16 个分立波长下的脉搏信号作为神经网络的输入, 即输入层单元数为 16, 网络的输出单元为 1, 通过调整隐含层节点数优化网络模型, 建立最优定标模型, 定标结果如表 2 所示。对预测集 17 份样品的预测结果如图 5 所示。

Table 2 Results of calibration and prediction by PLS and BP-ANN

	R_c^*	RMSEC/(g · L ⁻¹)	R_p^*	RMSEP/(g · L ⁻¹)
Raw	0.90	13.86(9.19%)	0.74	12.85(8.49%)
Savitzky-Golay	0.90	13.23(8.75%)	0.79	10.97(7.27%)
Moving Average	0.89	13.94(9.22%)	0.77	11.76(7.80%)
EMD	0.93	11.69(7.76%)	0.87	8.08(5.34%)

* R_c : 定标集相关系数; R_p : 预测集相关系数

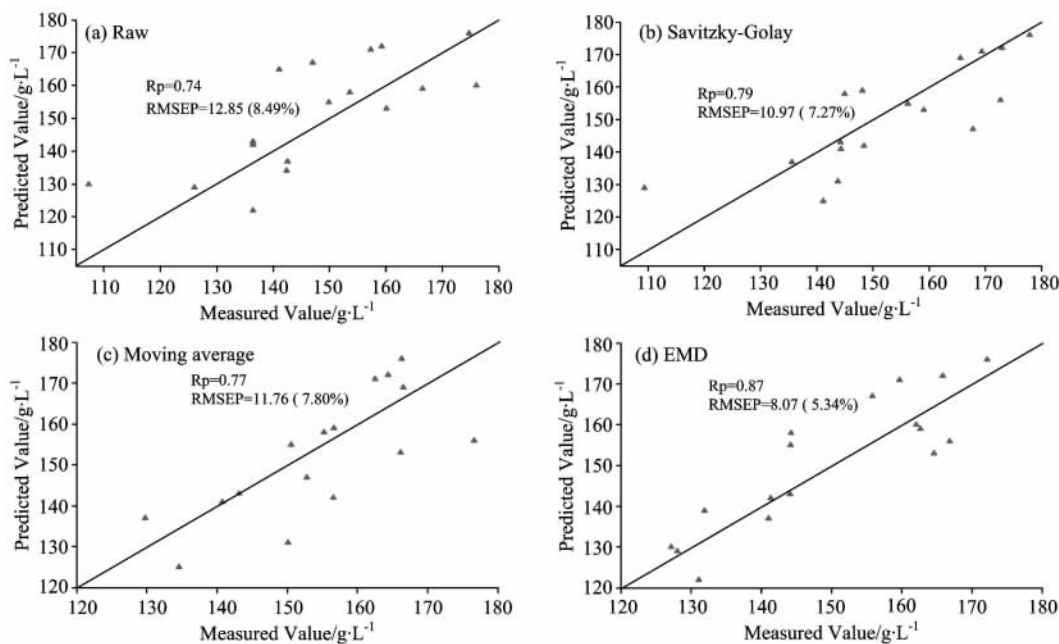


Fig 5 Prediction results by BP-ANN

由表 2 可知, NIRS 无创血红蛋白浓度模型对待测样品的预测精度最大相对误差均不超过 10%, 证实了该技术的可行性。对比分析不同数据预处理方法所建模型下的预测结果, 原始数据建模的预测相关系数只有 0.74, 误差均方根为 12.85 g · L⁻¹; S-G 平滑以及移动窗口平滑算法均对模型预测效果有所提高, 但效果不显著; 经 EMD 预处理后, 模型的预测相关系数可达 0.87, 误差均方根减小到 8.08 g · L⁻¹, 明显优于去噪前模型, 与平滑算法相比, 对模型的优化能力更强。结果表明, 通过 EMD 方法可以有效去除数据噪声, 得到高信噪比的血流容积脉搏信号, 从而提高 NIRS 无创血红蛋白预测模型的预测精度。

3 结 论

目前 NIRS 无创血红蛋白检测技术的分析精度尚未能达到临床要求, 选择合适方法有效去除数据噪声、提高定标模型稳健性, 已成为该领域的研究热点。本文无创采集了 81 例临床志愿者的手指指端血流容积脉搏信号, 通过剔除异常样品, 最终确定 78 例样品为研究对象。分别应用 S-G 平滑、移动窗口平滑及 EMD 分解方法进行数据预处理, 有效地抑制了噪声, 提高了脉搏波数据的信噪比。建立 BP-ANN 定量分析模型并预测, 经 EMD 去噪后所建模型预测效果最优: 预测相关系数由 0.74 提升到 0.87、误差均方根则由 12.85 g ·

L^{-1} 减小至 $8.08 \text{ g} \cdot L^{-1}$ 。实验证明应用 EMD 方法在脉搏波光谱去噪方面比平滑算法更具优势,能够有效提高血流容积脉搏信号的信噪比,为推动近红外无创血红蛋白检测技术的进一步发展提供了技术支持。

References

- [1] Timm U, Lewis E, McGrath D, et al. *Procedia Engineering*, 2010, 5: 488.
- [2] Zhou Y, Zheng C, Cao H, et al. *Biochemical and Biophysical Research Communications*, 2012, 420: 205.
- [3] Causey M W, Miller M D S, et al. *The American Journal of Surgery*, 2011, 201: 592.
- [4] Jeon Kyejin, Kim Sujin, Park kunkook. *Journal of Biomedical Optics*, 2002, 7(1): 45.
- [5] LU Qi-peng, CHEN Cong, PENG Zhong-qi(卢启鹏, 陈 丛, 彭忠琦). *Optics and Precision Engineering(光学精密工程)*, 2012, 20(4): 873.
- [6] CHEN Xing-dan(陈星旦). *Optics and Precision Engineering(光学精密工程)*, 2008, 16(5): 759.
- [7] LU Wan-zhen(陆婉珍). *Modern Near-Infrared Spectroscopy Analytical Technology*, 2nd ed(现代近红外光谱分析技术, 第2版). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2006.
- [8] CHEN Cong, LU Qi-peng, PENG Zhong-qi(陈 丛, 卢启鹏, 彭忠琦). *Acta Optica Sinica(光学学报)*, 2012, 32(5): 0530001-1.
- [9] Huang N E, Shen Z, Long S R, et al. *Proceedings of the Royal Society A*, 1998, 454: 903.
- [10] Huang N E, Shen Z, Long S R. *Annual Review of Fluid Mechanics*, 1999, 31: 417.
- [11] Bao L, Dai P H. *Journal of Systems Engineering*, 1999, 14(4): 338.

Study on the Application of Empirical Mode Decomposition to Noninvasive Hemoglobin Measurement by NIRS

FAN Yi-chen^{1,2}, LU Qi-peng^{1*}, DING Hai-quan¹, GAO Hong-zhi¹, CHEN Xing-dan¹

1. State Key Lab of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract To increase the signal-to-noise ratio (SNR) of human near infrared (NIR) spectra, so as to improve the stability and precision of calibration model, the empirical mode decomposition (EMD) method was applied. Eighty-one fingertip absorption curves were collected, with the corresponding clinical examination results obtained immediately. By means of outliers detection and removal, finally 78 samples were determined as the research objects. A three-layer back-propagation artificial neuron network (BP-ANN) model was established and worked for prediction. The results turned out that, through EMD method, the prediction correlation coefficient increased greatly from 0.74 to 0.87. RMSEP was reduced from 12.85 to $8.08 \text{ g} \cdot L^{-1}$. Other indexes were also obviously improved. The overall results sufficiently demonstrate that it is feasible to use EMD method for high SNR pulse wave signals, thus improving the performance of noninvasive hemoglobin calibration models. The application of EMD method can help promote the development of noninvasive hemoglobin monitoring technology.

Keywords Near infrared spectroscopy; Hemoglobin concentration; Empirical mode decomposition; Noninvasive measurement

(Received Jun. 1, 2012; accepted Aug. 20, 2012)

* Corresponding author