

提高近红外光谱法检测人体血液等复杂溶液成分准确度的研究进展

韩广^{1,3}, 王小燕¹, 陈思琪¹, 王慧泉^{1,3}, 王金海^{1,3}, 赵喆^{2,3*}

1. 天津工业大学生命科学院, 天津 300387
2. 天津工业大学电子与信息工程学院, 天津 300387
3. 天津市光电检测技术与系统重点实验室, 天津 300387

摘要 近红外光谱常被用于含氢有机物质等的物化参数测量, 可以提供丰富的结构和组成信息在复杂溶液的光谱定量分析中更是被广泛应用。然而在人体血液等复杂溶液的近红外光谱分析中, 强大的背景信息造成的噪声干扰和冗余变量的存在, 严重影响着样品的光谱测量和分析, 影响着分析的效率和准确度。如何消除背景噪声等的干扰来提高分析准确度已经引起高度重视, 近几年来, 国内外学者提出了许多基于化学计量学方法的相关方法。从光谱预处理、变量优化和建模分析三方面, 以传统的化学计量学方法出发, 总结和分析这些方法在人体血液等复杂溶液的近红外光谱定量分析的应用和各自的特点, 为提高光谱定量分析准确度的研究提供参考。

关键词 近红外光谱; 背景干扰; 变量优化; 模型; 准确度

中图分类号: O433.4 **文献标识码:** R **DOI:** 10.3964/j.issn.1000-0593(2021)07-1993-05

引言

近红外光谱主要是由于分子振动的非谐性使分子振动从基态向高能级跃迁时产生的, 带有丰富的物质成分信息, 适合于含氢化合物的分析^[1]。随着化学计量学技术的发展, 近红外光谱被广泛应用在物质成分的定量分析中, 在液体食品^[2]、食用油^[3]和血液^[4]等方面的光谱定量分析中, 有快速、非侵入和准确性高的特点, 对人们的生产和生活有着重要的意义。

杂散光、背景变动和冗余变量等引起的干扰导致光谱信噪比低等, 严重影响着分析的准确度, 提高人体血液等复杂溶液的近红外光谱定量分析准确度已经成为非常有意义的研究。近几年来, 相关研究已取得丰富成果。本文对提高分析准确度的最新策略进行综述, 分析杂散光和背景变动等干扰的预处理方法, 探讨优化变量对提高光谱定量分析准确度的影响, 概括在建模方面用来提高溶液的近红外光谱定量分析准确度的最新策略, 为人体血液等复杂溶液的近红外光谱定量分析提供参考。

1 杂散光和背景变动等干扰预处理方法

在复杂溶液的近红外光谱定量分析中, 背景噪声和冗余变量的消除是提高分析准确度的关键, 因此光谱预处理是必不可少的。传统的滤波平滑算法被不断的改进, 除此之外, 新的处理方法也被逐渐应用。例如李刚等人研究分析物与干扰物之间的光谱多重共线性, 采用光谱差异系数作为验证该共线性程度的指标, 通过仿真实验肯定该结论的普适性, 随后又提出基于净信号的波长优选新方法, 并通过乙醇溶液的光谱定量实验证明该方法提高分析准确度的有效性^[5]。此外还研究基于“M+N”理论, 系统地阐述M个组分和N个干扰因子引起的系统误差^[6], 并在阐述散射引起的非线性干扰问题中提到使用预处理和非线性校正方法相结合建立稳健的模型来减少散射引起的非线性影响^[7]。

传统的消除光谱仪机器噪声等无用信息的光谱预处理方法有导数法、平滑法和乘法信号校正(MSC)法等数学滤波算法, 与OSC、MSC等滤波去噪方法不同的是信号平滑算法, 传统的有移动平均平滑法和Savitzky-Golay(S-G)卷积平滑算法, 其中用的最广泛的是S-G平滑算法。在2019年, Amneh A Al-Mbaideen等研究移动平均滤波器在近红外光谱定量分

收稿日期: 2020-07-07, 修订日期: 2020-11-02

基金项目: 国家自然科学基金项目(61905176, 81901789), 博士后科学基金项目(2019M651036)资助

作者简介: 韩广, 1987年生, 天津工业大学生命科学院讲师 e-mail: hanguang@tiangong.edu.cn

* 通讯作者 e-mail: zhaozhe@tiangong.edu.cn

析中的应用,使用指数 MA 滤波器,实验结果表明该方法与其他类型的 MA 和 S-G 滤波器相比,PLS 模型预测的均方根误差明显降低,模型预测准确度提高^[8]。Osamah Alrezi 等提出使用带阻滤波(BSF)法,并将该方法与传统的带通滤波(BPF)法进行比较,结果表明,在近红外区该算法与 PLS 模型结合,将预测均方根误差从 $100 \text{ mg} \cdot \text{dl}^{-1}$ 降到 $34 \text{ mg} \cdot \text{dl}^{-1}$,最佳血糖预测决定系数为 0.982^[9]。两种算法都是将滤波器应用到近红外光谱定量分析中,并且实验结果表明,两种滤波算法都能有效的降低 PLS 模型预测的均方根误差,提高模型预测的准确度。然而移动窗口的宽度决定平滑去噪的效果,窗口大小不当会造成光谱信号失真等影响。

近年来,基于小波分析的去噪算法也逐步用于化学分析。2018 年,周凤波基于小波分析的去噪算法的主要进展,提出基于提升小波变换的阈值改进算法,实验结果表明该算法优于 Donoho 的软硬阈值去噪算法,有效地提高光谱信噪比和降低均方根误差,且得出该方法在紫外可见光谱信号分析中具有较好的应用前景的结论^[9]。2019 年,Shao 等提出基于连续小波变换的高阶导数分析方法,实验结果表明,该方法与 FT 法和 S-G 法一样有效,该方法的一大优势在于四阶导数能提高分辨率,降低噪声和旁瓣效应^[11]。

人体血液成分检测受背景变动等干扰更为显著,个体差异和测量条件的影响使得在实际的临床应用阶段进展缓慢。为克服强大的背景干扰,研究者开始尝试使用不同于传统的测量方法。李刚等在 2006 年对动态光谱法(DS)在近红外无创血液成分检测的应用进行理论分析,并验证该方法提高血液成分检测准确度的可行性^[12],此后对该方法进行一系列的改进,先后采取使用多次谐波分量^[13];小波变换^[14];进行 Boxcar 积分、降低波长分辨率、均衡 DS 信噪比^[15];基于 PPG 信号的补偿拟合提取 DS 数据^[16]和 PPG 预处理和 DS 提取联合应用^[17]的方法,该团队对无创血糖的研究从理论到实验,都证明基于动态光谱法的研究能够提高无创血糖近红外测量准确度,推动无创血糖的临床应用。

徐可欣等首先讨论净信号在评价测量准确度方面的重要作用^[18];然后提出基于背景光谱构建噪声子空间提取葡萄糖净信号的方法,使相关系数达到 0.9,并且突出葡萄糖的特异信号^[19];对标准样本校正和参考波束设计进行研究,评估单光束和双探测器双光束测量的信噪比,证明双光束设计更优于前者^[20];双光路测量结合净信号处理的参考测量方法的 PLS 模型预测结果表明,该方法的效果优于单光路扣除背景光谱,更有效地提取到葡萄糖的特征信息从而提高预测准确度^[21]。

在近几年无创血糖的近红外光谱定量分析中,基于浮动参考方法的相关策略得到较好的发展,旨在消除复杂的人体背景和波动的测量条件带来的强大背景干扰。Jyoti Yadav 等在研究无创血糖的测量时设计出一种低成本、无痛、无创的血糖测量系统,并通过对人体皮肤产生的背景噪声进行浮动或内参照,消除背景噪声的干扰,并采用基于 LM 的神经网络模型进行标定,测量葡萄糖浓度,实验证明该方法的高效性^[22]。刘蓉等在研究基于近红外光谱的血糖无损检测时,考虑到葡萄糖的特异性信号较弱和背景变化问题,采用特殊的

源-探测器距离作为浮动参考位置,进行相对测量和背景校正。研究结果表明该方法显著增强漫反射系数与葡萄糖浓度的相关性,有效地消除背景变化的影响^[23]。而后又在浮动参考位置的基础上,探讨半无限浑浊介质中浮动参考位置的理论基础,采用搭建的基于定制连续移动空间分辨双光纤近红外测量系统,分别进行蒙特卡罗和体外实验,模拟结果和实体验证结果均表明该研究对浮动基准测量光纤的设计具有重要意义^[24]。

徐可欣等提出一种基于浮动基准位置和测量位置的差分校正方法,旨在消除背景干扰引起的光谱变化,并进行体内和体外实验验证,实验结果肯定该方法的有效性,在无创血糖测量中有很好的应用前景^[25]。而后又提出采用新的方法来校正光谱漂移问题,以浮动参考位置的相对变化作为基线参考,定量描述不同源-探测距离分离处的信号漂移,建立一个统一的方程来描述浮动参考位置变化之间的关系,利用该方法校正近红外光谱漂移^[26]。以上的研究均是以浮动参考测量方法为基础展开的相关研究,仿真实验和在体实验都验证该方法的有效性,与传统的方法相比,该方法更好地消除个体差异和测量条件变化引起的干扰,从很大程度上提高无创血糖的近红外光谱定量分析准确度,更推动无创血糖的临床研究。

2 波长变量选取方法

使用光谱定量分析溶液时,变量个数往往大于样本个数,造成信息的冗余,剔除多余变量是简化模型和提高预测准确度必不可少的。在传统的方法中,无信息变量消除(UVE)法是基于 PLS 回归系数分析的变量选择方法,而将蒙特卡罗采样和 PLS 回归系数结合的变量选择方法在近几年的研究中也广泛被应用,本节探讨基于蒙特卡罗采样和无信息变量消除法结合的波长变量选取的最新策略。

Han 等将蒙特卡罗(MC)方法引入到无信息变量消除(UVE-PLS)中,即改进的 MC-UVE-PLS,从模拟数据和实际数据得到的结果来看,该方法提高多元校正模型的预测能力^[27]。朱红求等研究一种基于紫外可见光谱的多金属离子浓度检测方法,提出基于 Monte Carle 的无信息变量消除方法与指数衰减函数相结合的改进型算法^[28]。

竞争自适应加权采样(CARS)算法是蒙特卡罗方法的典型代表,Li 等在 2009 年提出该算法,用 MC 进行 N 次采样,随机抽取校正集部分样本建立 PLS 模型,采用基于指数递减函数(EDF)的强制波长选择和基于自适应重加权采样(ARS)的竞争波长选择方法去除模型回归系数绝对值相对较小的点,筛选出回归系数绝对值大的波长点作为特征变量,结果证明该方法与之前的 MC-UVE-PLS 和移动窗口偏最小二乘(MWPLSR)相比,模型的预测结果更好^[29]。张华秀等将 CARS 变量筛选方法应用在液态奶中蛋白质和脂肪含量的近红外光谱测定中,结果表明该方法可减小模型预测的均方根误差,提高相关系数,有效地简化模型^[30]。Tang 等在 2014 年提出竞争自适应重加权抽样逐次投影算法(CARS-SPA),并将其应用于烟草叶片中尼古丁的近红外反射数据和农药制

剂中有效成分的近红外透射数据,结果证明该方法选取的变量建立的 MLR 模型比 SPA 和 UVE-SPA 具有更好的预测效果^[31]。随后该团队在 2016 年采用 MIR 和 CARS 对苦参碱水溶液中的有效成分进行测定,结果证明了 CARS 消除频谱冗余的有效性^[32]。Jan PM Andries 等提出重复 UVE 算法,即 GME-UVE-PLS 算法,实验结果表明该算法优于 UVE-PLS 算法,能够消除更多的变量,且模型的预测性能更高^[33]。在 2019 年, Li 等提出新的基于竞争自适应重加权采样集合上的移动窗口平滑技术的新的算法(MWS-ECARS),该算法首先进行 CARS 集合,然后进行 MWS 搜索高频变量周围的有效变量,实验结果表明该算法具有更好的预测性能^[34]。Yang 等也提出一种优于 MC-UVE 算法的移动窗口偏最小二乘法与采样误差剖面分析法(SEPA-MWPLS)法,显著改善预测模型的性能^[35],以上的文献均证明 CARS 方法筛选特征变量的有效性和优越性。

变量筛选同样地在近红外高光谱技术中也是关键的步骤,由于高光谱图像通常具有高维数据,其中冗余数据颇多,严重影响近红外高光谱定量分析的准确度,因此使用有效的方法筛选出变量显得尤为重要,CARS 法就是其中方法之一。在 2014 年, Xie 等使用近红外高光谱成像技术鉴定不同品种的芝麻油时,讨论了 CARS, SPA 和 x-LW 三种筛选变量的方法,并分别建立 SVM 和 LDA 模型,结果表明三种变量选择方法都有效地提高模型预测准确度,得到满意的结果^[36]。Xia 等在 2017 年研究近红外高光谱成像技术对产地麦冬的鉴别性能时,比较包括 UVE、SPA 和 CARS 等在内的 10 中波长选择方法,结果证明波长选择方法对建模分析的重要性^[37]。Zhang 等在 2019 年使用近红外高光谱成像技术对苹果品质进行无损检测,并且比较 CARS、SPA、随机青蛙(RF)和 CARS-SPA、CARS-RF 五种变量筛选算法,实验结果表明 CARS-SPA 算法选择波长对苹果可溶性固形物含量的在线检测具有很大的潜力^[38]。

3 建模策略的研究进展

使用分子光谱法对溶液成分进行定量分析时,除传统的线性回归方法,如偏最小二乘、主成分分析和多元线性回归法外,还有可以直接对非线性信息建模的非线性回归法,如人工神经网络法、支持向量机法等,局部建模法也被广泛应用,本节就近几年成分的光谱定量分析在建模方面采取的策略进行讨论。

Chen 等使用近红外光谱技术快速检测水污染,采用 LSSVM 方法建立近红外光谱定量测定水污染的化学需氧量的校正模型,并研究不同核函数嵌入 LSSVM 的效果,提出利用深度学习方法进行参数优化的核函数方法,且结果证明该建模方法的可行性^[39]。Zhao 等为提高楔形样品池中复杂溶液成分检测的准确性,采用多位置、多波长检测方法,获

得包括样品在内的多径非线性特征的多维谱,然后采用多维光谱融合方法建立可视化的光强分布图像^[40]。而后该团队为检测浑浊介质的光学性质和化学成分,开发入射面不平行于出射面的异性样品池,获得浑浊介质的非周向对称高光谱图像的散射信息^[41]。

李刚团队使用近红外漫射透射光谱和偏最小二乘法(PLS-DA)进行血型鉴别,实验结果证明该模型有很好的内部和外部分类能力^[42]。而后又研究分辨人血与非人血液的可见光漫反射光谱的最佳波长选择,使用等间距组合多元线性回归方法来选择最佳波长,其结果表明该方法选取的波长变量构建的 PLS-DA 模型更有效、更准确^[43]。该团队证明使用可见光漫反射光谱法可以实现非接触式人体和非人体血液识别后,又采用偏最小二乘判别分析(PLSDA)法和最小二乘支持向量机(LSSVM)法,其结果证明 LSSVM 法比 PLSDA 法更可靠更稳健^[44]。在测量血浆中的总胆红素时,基于“M+N”理论,采用紫外光谱、可见光谱和近红外光谱相结合的方法进行建模,即使用多波段联合建模方法,实验结果证明该方法可以显著提高光谱分析的准确度^[45]。后来李刚团队在研究血液成分的非创光谱检测时,针对散射引起的非线性干扰,提出根据血红蛋白的含量进行分组建模,将其与非分组建模方法的预测性能比较,发现该方法有效地减小非线性因素的影响,是解决散射引起的非线性影响的新方法^[46]。而后尝试利用径向基函数神经网络对任意非线性函数的逼近能力,从拟合的角度逼近谱与分量浓度之间的非线性关系,而实验结果也证明该方法的可行性^[47]。该团队还研究提高近红外模型鲁棒性的建模方法,采用两位置谱建模方法来抑制测量条件的影响,结果表明该建模方法的 RMSEP 下降 76%^[48]。又根据柔性容器与复杂溶液的吸收光谱差异,提出一种利用 N 路偏最小二乘法(NPLS)建立被测物质浓度与小波变换提取的光谱特征向量之间定量模型的方法,来抑制柔性容器中差异的影响^[49]。

4 结 论

为提高人体血液等复杂溶液成分光谱定量分析的准确度,分别从消除噪声和背景干扰、优化变量、建模分析三方面进行综述,对光谱预处理阶段采取的一系列策略进行重点探讨。在光谱预处理阶段,消除背景带来的干扰是必不可少的。尤其是在人体无创血糖的光谱检测中,寻找合适的在体参考测量方法对于消除背景干扰带来的影响十分重要,而基于动态光谱法和浮动基准测量法的研究是推动无创血糖测量的重要实践。在未来针对人体血液等复杂溶液成分的研究中,降低背景干扰和样本差异带来的测量误差,增强有效信号的灵敏度和提高预测模型的准确性是需要解决的几个关键问题。

References

- [1] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined With Chemometrics and Its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011.
- [2] Wang L, Sun D, Pu H. Critical Reviews in Food Science and Nutrition, 2017, 57(7): 1524.
- [3] Rohman A. Applied Spectroscopy Reviews, 2017, 52(7): 589.
- [4] Jernelv I L, Milenko K, Fuglerud S S. Applied Spectroscopy Reviews, 2018, 54(7): 543.
- [5] Liu H, Wang M, Li X. Analytical Methods, 2016, 8(23): 4648.
- [6] Wan X, Li G, Li T. Applied Spectroscopy Reviews, 2020, 55(2): 87.
- [7] Wan X, Li G, Zhang M. Applied Spectroscopy Reviews, 2020, 55(5): 351.
- [8] Al-Mbaideen A A. Journal of Analytical Chemistry, 2019, 74(7): 686.
- [9] Alrezi O, Benaissa M, Alshebeili S A. Journal of Chemometrics, 2020, 34(3): e3206.
- [10] Zhou F, Zhu H, Li C. Optik, 2019, 182: 786.
- [11] Shao X, Cui X, Wang M. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 213: 83.
- [12] LI Gang, WANG Yan, LI Qiu-xia(李刚, 王焱, 李秋霞). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2006, (5): 345.
- [13] Wang H, Li G, Zhao Z, et al. Transactions of the Institute of Measurement and Control, 2013, 35(1): 16.
- [14] Li G, Zhou M, He F. Review of Scientific Instruments, 2011, 82(9): 879.
- [15] Zhou M, Li G, Lin L. Journal of Biomedical Optics, 2013, 18(5): 57003.
- [16] Peng Y, Li G, Zhou M. RSC Advances, 2017, 7(18): 11198.
- [17] Liu A, Li G, Yan W. Infrared Physics & Technology, 2018, 92: 436.
- [18] LIU Rong, LÜ Li-na, CHEN Wen-liang(刘蓉, 吕丽娜, 陈文亮). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(9): 1042.
- [19] LIU Rong, CHEN Wen-liang, GU Xiao-yu(刘蓉, 陈文亮, 谷筱玉). Nanotechnology and Precision Engineering(纳米技术与精密工程), 2008, (3): 207.
- [20] Min X, Liu R, Hu Y. Analytical Methods, 2014, 6(24): 9831.
- [21] MIN Xiao-lin, LIU Rong, FU Bo(闵晓琳, 刘蓉, 傅博). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(9): 2943.
- [22] Yadav J, Rani A, Singh V. Journal of Research, 2018, 64(1): 116.
- [23] Min X, Liu R, Fu B. Applied Spectroscopy, 2017, 71(9): 2076.
- [24] Min X, Liu R, Fu B. Optics & Laser Technology, 2017, 91: 7.
- [25] Han G, Yu X, Xia D. Applied Spectroscopy, 2017, 71(9): 2177.
- [26] Han G, Han T, Xu K. Journal of Biomedical Optics, 2017, 22(7): 77001.
- [27] Han Q, Wu H, Cai C. Analytica Chimica Acta, 2008, 612(2): 121.
- [28] ZHU Hong-qiu, CHEN Jun-ming, YIN Dong-hang(朱红求, 陈俊名, 尹冬航). Journal of Chemical Industry and Engineering(化工学报), 2017, 68(3): 998.
- [29] Li H, Liang Y, Xu Q. Analytica Chimica Acta, 2009, 648(1): 77.
- [30] ZHANG Hua-xiu, LI Xiao-ning, FAN Wei(张华秀, 李晓宁, 范伟). Journal of Instrumental Analysis(分析测试学报), 2010, 29(5): 430.
- [31] Tang G, Huang Y, Tian K. Analyst, 2014, 139(19): 4894.
- [32] Tang G, Hu J, Yan H. Optik, 2016, 127(3): 1405.
- [33] Andries J P M, Vander Heyden Y, Buydens L M C. Analytica Chimica Acta, 2017, 982: 37.
- [34] Li Q, Huang Y, Song X. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 214: 129.
- [35] Yang W, Wang W, Zhang R. Analytical Sciences, 2020, 36(3): 303.
- [36] Xie C, Wang Q, He Y. PLOS ONE, 2014, 9(5): e98522.
- [37] Xia Z, Zhang C, Weng H. International Journal of Analytical Chemistry, 2017, 2017: 1.
- [38] Zhang D, Xu Y, Huang W. Infrared Physics & Technology, 2019, 98: 297.
- [39] Chen H, Xu L, Ai W. The Science of the Total Environment, 2020, 714: 136765.
- [40] Zhao Z, Yin H, Yan W. Infrared Physics & Technology, 2019, 97: 444.
- [41] Zhao Z, Yan W, Yin H. Microchemical Journal, 2019, 149: 104009.
- [42] Zhang L, Zhang S, Sun M. Infrared Physics & Technology, 2016, 76: 587.
- [43] Zhang L, Sun M, Wang Z. Analytical Methods, 2016, 8(2): 381.
- [44] Zhang L, Li G, Sun M. Infrared Physics & Technology, 2017, 86: 116.

- [45] Lin L, Zhang Q, Zhou M. RSC Advances, 2016, 6(114): 113322.
- [46] Liu A, Li G, Fu Z. Scientific Reports, 2018, 8(1): 8564.
- [47] Liu A, Li G, Yan W. Infrared Physics & Technology, 2018, 92: 436.
- [48] Luo Y, Yang S, Tian H. Infrared Physics and Technology, 2020, 104: 103053.
- [49] Hou X, Shi D, Li G. Infrared Physics & Technology, 2019, 99: 277.

Research Progress on Improving the Accuracy of Near Infrared Spectroscopy Detection of Human Blood and Other Complex Solution Components

HAN Guang^{1,3}, WANG Xiao-yan¹, CHEN Si-qi¹, WANG Hui-quan^{1,3}, WANG Jin-hai^{1,3}, ZHAO Zhe^{2,3*}

1. School of Life Sciences, Tiangong University, Tianjin 300387, China

2. School of Electronic and Information Engineering, Tiangong University, Tianjin 300387, China

3. Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin 300387, China

Abstract Near-infrared spectroscopy has rich structure and composition information and is often used to measure hydrogen-containing organic substances' physical and chemical parameters. In recent years, it has been widely used in the quantitative analysis of complex solutions. However, the near-infrared spectroscopy analysis of complex solutions such as human blood, noise interference and redundant variables caused by strong background information seriously affect the spectral measurement and analysis of the sample itself, and affect the efficiency and accuracy of the analysis. Therefore, eliminating the interference of background noise to improve the accuracy of analysis has attracted great attention. In recent years, scholars at home and abroad have proposed many related methods based on chemometric methods. This article focuses on spectral preprocessing, variable optimization and modeling analysis. On the one hand, starting from traditional chemometric methods, we summarize and analyze the application and respective characteristics of these methods in the near-infrared spectroscopy quantitative analysis of complex solutions such as human blood, and provide a reference for research on improving the accuracy of quantitative spectral analysis.

Keywords Near-infrared spectroscopy; Background interference; Variable optimization; Model; Accuracy

(Received Jul. 7, 2020; accepted Nov. 2, 2020)

* Corresponding author