

研究论文

正交信号校正正常成人血清<sup>1</sup>H NMR谱的代谢组分析中的滤噪作用评价

冒海蕾<sup>1,3</sup>, 徐旻<sup>2</sup>, 王斌<sup>3</sup>, 王惠民<sup>3</sup>, 邓小明<sup>\*1</sup>, 林东海<sup>\*2</sup>

(<sup>1</sup>第二军医大学长海医院 上海 200433)

(<sup>2</sup>中国科学院上海药物研究所 上海 201203)

(<sup>3</sup>南通大学附属医院 南通 226001)

收稿日期 2006-4-3 修回日期 2006-6-19 网络版发布日期 2007-1-12 接受日期 2006-9-6

**摘要** 观察、比较正交信号校正(OSC)滤噪前后,用不同的模式识别方法对正常成人血清代谢组<sup>1</sup>H NMR谱进行分析的效果,以探讨NMR代谢组学技术应用于临床研究和疾病早期诊断的可行性. 78例正常成人在采血前按常规要求禁食8 h,记录血清一维600 MHz氢谱后,分别采用主成分分析(PCA)、偏最小二乘法-判别分析(PLS-DA)以及簇类的独立软模式法(SIMCA)对氢谱进行模式识别分析. 结果表明:虽然采血前并无其它诸如饮食、生活方式、生理周期等方面的严格限制,采用OSC滤噪后,PLS-DA能够完全区分不同性别的血清氢谱,其判别能力优于PCA和SIMCA. 而且采用OSC滤噪与文献报道的未经OSC处理的PLS-DA法获得的与性别分类有关的主要NMR积分区段基本相同. 从OSC去除不同数目的隐变量后所致的PLS-DA模型的性能改变可见: OSC去除两个隐变量时,前两个隐变量的特征值明显比后面的大; 剩余残差为20.82%,即去除了79.18%的X变量中与反应变量Y不相关的系统变异. 此时PLS-DA计算所得的隐变量个数为1; 而不使用OSC或用OSC去除一个隐变量时,PLS-DA所得的隐变量个数分别为3和2. 作为PLS-DA模型质量的评价指标, R2X表示PLS-DA模型计算所获得的隐变量反映自变量X的变异的百分比, R2Y则表示隐变量反映因变量Y的变异的百分比, Q2 (cum)为交叉验证后PLS-DA模型所获隐变量能够预测X和Y变异的累计百分比. R2X在OSC去除两个隐变量时达到最低值,表明此时PLS-DA计算模型包含的系统变异最少; R2Y与Q2 (cum)都达到80%以上并趋于稳定,说明OSC去除两个隐变量时PLS-DA模型的质量优良. 显然, OSC可去除饮食、环境等因素的影响,降低临床样本的不均一性,这对于NMR代谢组学技术应用于临床研究至关重要. OSC滤噪去除的隐变量个数应根据剩余残差、去除隐变量的特征值大小、PLS-DA模型计算所得的隐变量个数和反映模型质量的相关指标加以判断.

**关键词** [NMR](#) [代谢组学](#) [模式识别](#) [正交信号校正](#) [血清](#)

分类号

## Evaluation of Filtering Effects of Orthogonal Signal Correction on Metabonomic Analysis of Healthy Human Serum <sup>1</sup>H NMR Spectra

MAO Hai-Lei<sup>1,3</sup>, XU Min<sup>2</sup>, WANG Bin<sup>3</sup>, WANG Hui-Min<sup>3</sup>, DENG Xiao-Ming<sup>\*1</sup>, LIN Dong-Hai<sup>\*2</sup>

(<sup>1</sup> Changhai Hospital, Secondary Military Medical University, Shanghai 200433)

(<sup>2</sup> Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203)

(<sup>3</sup> Affiliated Hospital, Nantong University, Nantong 226001)

**Abstract** Three different pattern recognition methods before and after orthogonal signal correction (OSC) were employed to perform the metabonomic analysis of <sup>1</sup>H NMR spectra recorded on healthy human sera, in order to explore the potential of applying <sup>1</sup>H NMR-based metabonomics to clinical research. At first, 78 healthy human sera were collected after a routine fasting for 8 h, and the corresponding 1D <sup>1</sup>H NMR spectra were recorded on a Varian Unity INOVA-600 spectrometer, and then three pattern recognition analyses, PCA (principal component analysis), PLS-DA (partial least squares-discriminant analysis), and SIMCA (soft independent modeling of class analogy), were performed, respectively. In spite of no specific sample-collecting restriction on foods, life styles, and physiological cycles, the PLS-DA method after OSC is able to distinguish the NMR metabonomic profiles of male sera from those of female sera, more perfectly than both the PCA and SIMCA. Furthermore, the major NMR integral regions relevant to gender classification from PLS-DA after OSC were identical with those from PLS-DA without OSC filter in the literature. In the figure of displaying the variation of PLS-DA model before OSC and after removing different OSC latent variables (LVs), the eigenvalues of the first and second OSC-removed LVs were much greater than others. After removing two LVs by OSC, the remaining sum of square (RSS) in the X block was 20.82%, that is, 79.18% information unrelated to Y was removed from the PLS-DA model. Meanwhile, the LV number of PLS-DA model attained to one; while the LV number was two for the model with the first LV being removed by OSC, and three for the model without OSC. R2X, R2Y, and Q2 (cum) are usually used to evaluate the quality of PLS-DA model. R2X and R2Y are the fraction of the sum of square of the entire X's and Y's explained by the current LV of PLS-DA, and represent the variance of X and Y variables, respectively; while Q2 is cross validated R2. Q2 (cum) reflects the cumulative cross-validated percent of the total variation of the X's and Y's that can be predicted by the current LV of PLS-DA model. In our study, after OSC filtering the first two LVs, R2X reached the

扩展功能

### 本文信息

▶ [Supporting info](#)

▶ [PDF\(439KB\)](#)

▶ [\[HTML全文\]\(46KB\)](#)

▶ [参考文献](#)

### 服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

### 相关信息

▶ [本刊中 包含“NMR”的 相关文章](#)

▶ 本文作者相关文章

· [冒海蕾](#)

· [徐旻](#)

· [王斌](#)

· [王惠民](#)

· [邓小明](#)

· [林东海](#)

minimum, suggesting that the least systematic variance should be present in the current PLS-DA model. Meanwhile, both  $R^2Y$  and  $Q^2$  (cum) were always higher than 80%, indicative of the good quality of the PLS-DA model. Obviously, OSC is capable of eliminating the influence of dietary and environmental factors, and decreasing the heterogeneity of samples, which is fairly useful and important for clinical investigations. Additionally, the appropriate number of OSC-removed LVs should be determined on the basis of RSS in the  $X$  block, eigenvalue of OSC-removed latent variables, LV number and the qualitative indicators of the PLS-DA model.

**Key words** [NMR](#) [metabonomics](#) [pattern recognition](#) [orthogonal signal correction](#) [sera](#)

DOI:

---

通讯作者 林东海 [dhlin@mail.shcnc.ac.cn](mailto:dhlin@mail.shcnc.ac.cn)