

## 光谱仪器信噪比与建模波长数对光谱分析精度的影响

李刚<sup>1</sup>, 赵喆<sup>1</sup>, 王慧泉<sup>1</sup>, 林凌<sup>1</sup>, 吴晓荣<sup>2</sup>, 张宝菊<sup>2\*</sup>

1. 天津大学精密测试技术及仪器国家重点实验室, 天津 300072  
2. 天津师范大学物理与电子信息学院, 天津 300387

**摘要** 为了研究光谱建模分析中光谱仪器噪声和参与建模的波长数两个因素与定量分析精度之间的关系, 针对不同信噪比对模型精度的影响、参与建模的波长数与分析精度的关系、不同信噪比波段中多波长参与建模对分析精度的影响三个方面进行了理想样本建模分析。实验结果表明: 光谱测量仪器的噪声水平直接影响建模分析误差, 而使用多波长尤其是信噪比较好的波段的光谱参与线性建模, 近似满足波长数每增加4倍, 分析精度提高2倍的关系, 能够在一定程度上弥补信噪比的不足。研究为在光谱分析中充分利用和提高光谱仪器信噪比、合理选择参与建模波长数及波段, 从而提高光谱定量分析精度提供了实验基础和理论指导。

**关键词** 信噪比; 波长数; 光谱分析精度; 纯吸收; 偏最小二乘

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2012)08-2286-04

### 引言

对多组分混合样品中某种组分进行可见-近红外光谱分析时, 由于被测组分在所测波段吸收谱信号微弱, 光谱检测仪器的噪声会对被测组分的定量分析产生较大的影响<sup>[1-3]</sup>, 可能导致被测成分检测精度下降, 甚至不能满足检测要求。虽然光谱检测仪器的信噪比直接影响着组分浓度的定量分析精度, 但是由仪器信噪比所决定的检测精度可以通过后续的数据处理和建模分析得到提高<sup>[4-6]</sup>。

偏最小二乘回归建模方法是目前应用最广泛的光谱定量分析数学方法, 主要用来解决多元回归分析中变量的多重相关性和变量多于样本点等实际问题<sup>[7]</sup>, 通常采用多波长同时参与建模。根据过采样技术提高数据精度及信噪比的理论<sup>[8-10]</sup>, 多波长建模能够带来浓度检测精度的提高。

从不同信噪比对模型精度的影响、参与建模的波长数与分析精度的关系、不同信噪比波段中多波长参与建模对分析精度的影响三方面进行了仿真实验研究, 对光谱仪器信噪比和参与建模波长数对所测样本的光谱定量分析精度的影响进行了分析。

### 1 光谱测量仪器信噪比对光谱建模分析精度的影响

#### 1.1 建立仿真实验模型

光谱测量仪器的信噪比是决定光谱法建模分析物质成分含量精度的一个重要影响因素。仅考虑吸收且不考虑散射, 依照朗伯-比尔定律, 在已知摩尔吸光系数曲线、各组分浓度和光程的情况下, 求得三组分被测样品吸收谱。再根据实验需要在吸收谱上叠加一定均方值高斯白噪声, 以模拟不同信噪比情况下采集的光谱数据。仿真实验中, 被测液体三种组分的可见近红外波段摩尔吸光系数曲线如图1所示。

其中, 前两个组分为背景组分, 二者吸收谱曲线平缓, 由正弦曲线得到; 第三组分为被测组分, 共6个高斯吸收峰, 第 $N$ 个( $N=1, 2, 3, 4$ )吸收峰处的吸光系数为第 $N+2$ 个吸收峰处的10倍。

各样本光谱透射光程为10 mm。实验中三种组分的浓度范围依次为 $36\sim 55 \text{ mol} \cdot \text{L}^{-1}$ ,  $0\sim 0.05 \text{ mol} \cdot \text{L}^{-1}$ 和 $0\sim 0.01 \text{ mol} \cdot \text{L}^{-1}$ 。依据朗伯-比尔定律得到不同浓度的三种组分混合的吸收谱。在各样本的吸收谱上叠加的仿真仪器噪声的高斯白噪声以被测目标组分最大浓度含量的1%所造成的吸收

收稿日期: 2011-11-05, 修订日期: 2012-02-20

基金项目: 国家自然科学基金项目(30973964), 天津市应用基础及前沿技术研究计划项目(11JCZDJC17100), 天津市科技计划项目和科技型中小企业创新基金项目(10ZXCXSY10400)资助

作者简介: 李刚, 1959年生, 天津大学精密测试技术及仪器国家重点实验室教授 e-mail: ligang59@tju.edu.cn

\* 通讯联系人 e-mail: wdxyzbj@163.com

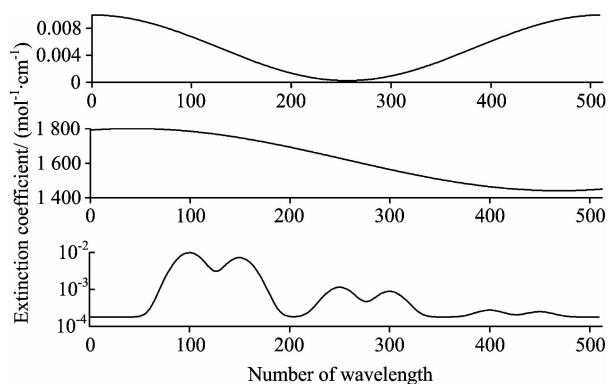


Fig. 1 Ideal curve of molar extinction coefficient of three components

谱最大变化幅值作为 1 个噪声单位。

共建立仿真样本 500 例, 其中校正集样本 400 例, 预测集样本 100 例, 三种组分的浓度值分别在两个样本集中均匀分布。各样本吸收谱向量包含 512 个波长。

### 1.2 不同信噪比对模型精度的影响

在各样本吸收谱上叠加不同噪声单位的噪声后, 由偏最小二乘回归建立第三组分的定量分析模型, 并对预测集进行预测, 以校正集的决定系数( $R_c^2$ )、标准偏差(RMSEC)、预测集决定系数( $R_p^2$ )、标准偏差(RMSEP)为评价指标。叠加不同均方值噪声时的建模和预测效果如表 1 所示, 其中误差倍数为各噪声条件下的误差相对于叠加 1% 噪声时误差的倍数。其中, 校正集和预测集标准偏差与叠加噪声的关系如图 2 所示。

Table 1 Modeling and prediction results with different noise units

叠加噪声的 均方值/%	校正集		预测集	
	$R_c^2$	RMSEC	$R_p^2$	RMSEP
1	1.00	$1.9 \times 10^{-5}$	1.00	$1.8 \times 10^{-5}$
5	1.00	$9.6 \times 10^{-5}$	1.00	$9.7 \times 10^{-5}$
10	0.99	$1.9 \times 10^{-4}$	1.00	$2.0 \times 10^{-4}$
20	0.98	$3.7 \times 10^{-4}$	0.99	$3.3 \times 10^{-4}$
50	0.91	$8.7 \times 10^{-4}$	0.90	$9.4 \times 10^{-4}$
100	0.78	$1.4 \times 10^{-3}$	0.48	$2.1 \times 10^{-3}$
150	0.72	$1.5 \times 10^{-3}$	0.29	$2.5 \times 10^{-3}$
200	0.62	$1.8 \times 10^{-3}$	0.12	$2.8 \times 10^{-3}$

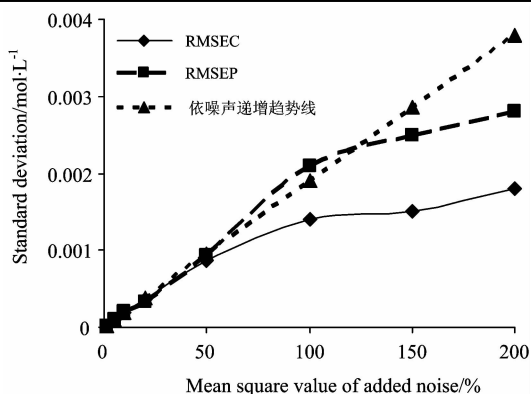


Fig. 2 Relationship between standard deviation and added noise

从图 2 中可以看出, 在仅改变所叠加噪声的均方值的情况下, 校正集和预测集的标准偏差与所叠加的噪声水平在数值上为近似的正比关系。实验结果表明, 光谱测量仪器的信噪比对建模分析误差起着决定性的作用, 使用信噪比较高的光谱仪器是提高物质成分定量分析精度的必要保证。

同时, 从实验结果中也能够看出, 当叠加噪声较大(超过 50%)时, 校正集和预测集的标准偏差并不是完全成比例的增加, 而是比噪声增长的缓慢。其原因可能是由于多波长线性建模具有平均效应, 对噪声有一定的抑制作用, 从而得到比仪器噪声水平所决定的分析精度更高的定量分析精度。因此, 除分析精度与信噪比相关外, 需要对组分定量分析精度是否还与多波长参与建模有关进行进一步实验和分析。

## 2 参与建模的波长数与分析精度的关系

在光谱法建模分析中, 通常会选取多个波长下的吸收谱值进行成分分析, 通过多波长的参与建模, 有可能在光谱仪器本身信噪比的基础上使组分浓度检测灵敏度得到提高。

### 2.1 在吸收峰处挑选波长

在叠加 10% 噪声且波长分辨率一定的情况下, 首先考虑从被测组分吸收峰处选择不同个数波长下的光谱参与建模。参与建模波长数为 1 时, 选用第 100 个波长; 波长数为 4 时, 选用吸收峰处的第 100, 150, 250, 300 个波长; 波长数为 16 时, 选用第 99~101, 149~151, 249~251, 299~301, 399~401, 450 个波长; 波长数为 64 时, 选用第 95~105, 145~155, 245~255, 295~305, 395~405, 446~454 个波长; 波长数为 256 时, 选用第 79~121, 129~171, 229~271, 279~321, 379~421, 430~470 个波长。不同波长数光谱数据得

Table 2 Modeling and prediction results of different wavelength numbers

参与建模 的波长数	校正集		预测集	
	$R_c^2$	RMSEC	$R_p^2$	RMSEP
1	0.00	$2.9 \times 10^{-3}$	0.00	$3.0 \times 10^{-3}$
4	0.84	$1.2 \times 10^{-3}$	0.86	$1.1 \times 10^{-3}$
16	0.94	$7.1 \times 10^{-4}$	0.94	$7.3 \times 10^{-4}$
64	0.98	$3.6 \times 10^{-4}$	0.98	$3.9 \times 10^{-4}$
256	0.99	$2.2 \times 10^{-4}$	0.99	$2.4 \times 10^{-4}$

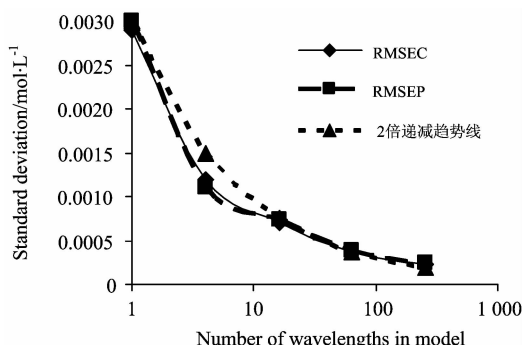


Fig. 3 Relationship between standard deviation and wavelength number

到的建模和预测结果如表 2 所示。其中,校正集和预测集标准偏差与参与建模的波长数的关系如图 3 所示。

从图 3 中实验结果可以看出,在改变参与建模的光谱数据波长数的过程中,建模分析精度具有随建模波长数的增加而提高的趋势,并且近似满足波长数每增加 4 倍,标准偏差减半,即建模精度提高 2 倍的趋势。

## 2.2 等间隔挑选波长

为证明这种定量分析精度随建模波长增加而增加的规律并不依赖于在吸收峰处选择波长,对等间隔选择参与建模波长进行进一步实验。参与建模波长数为 1 时,选用第 256 个波长;波长数为 4 时,选用吸收峰处的第  $102 \times N (N=1 \sim 4)$  个波长;波长数为 16 时,选用第  $30 \times N (N=1 \sim 16)$  个波长;波长数为 64 时,选用第  $8 \times N (N=1 \sim 64)$  个波长;波长数为 256 时,选用第  $2 \times N (N=1 \sim 256)$  个波长。噪声水平依然为 10% 且波长分辨率一定。不同波长数光谱数据得到的建模和预测结果如表 3 所示。其中,校正集和预测集标准偏差与参与建模的波长数的关系如图 4 所示。

Table 3 Modeling and prediction results of different wavelength number

参与建模的波长数	校正集		预测集	
	$R^2$	RMSEC	$R^2$	RMSEP
1	0.00	$2.9 \times 10^{-3}$	0.00	$3.0 \times 10^{-3}$
4	0.78	$1.4 \times 10^{-3}$	0.81	$1.3 \times 10^{-3}$
16	0.88	$9.9 \times 10^{-4}$	0.89	$9.7 \times 10^{-4}$
64	0.97	$5.0 \times 10^{-4}$	0.95	$6.1 \times 10^{-4}$
256	0.99	$2.7 \times 10^{-4}$	0.99	$2.8 \times 10^{-4}$

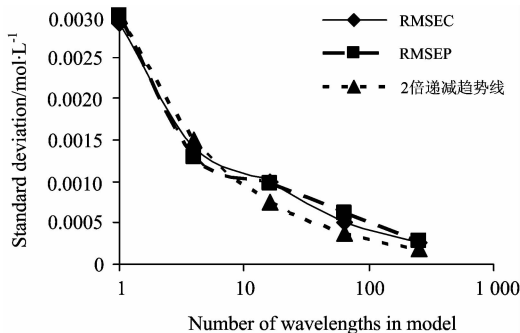


Fig. 4 Relationship between standard deviation and wavelength number

从图 4 中实验结果可以看出,在等间隔选择参与建模波长情况下,建模分析精度与建模波长数的关系依然近似满足波长数每增加 4 倍,建模精度提高 2 倍的趋势。

## 2.3 两种改变波长数方法的比较

进一步对比图 3 和图 4 能够看出,虽然采用在吸收峰处挑选波长和采用等间隔挑选波长两种改变参与建模波长数的方法的建模和预测结果存在一些共同点,即:首先,通过两种改变波长数方法所得到的标准偏差都近似满足波长数每增加 4 倍,模型标准偏差减小 2 倍的趋势。其次,随着波长数的增加,标准偏差的数值趋于平坦,这种变化趋势表明在信

噪比一定的情况下,一味通过增加波长数的方法并不能使模型精度得到无限度的提高。

两种改变波长数方法的建模和预测结果也存在一些不同点:图 3 中波长数较少的前半段的校正集和预测集标准偏差均低于趋势线,图 4 中前半段的校正集和预测集标准偏差与趋势线的拟合程度比图 3 高,这说明在波长数较少的情况下,尽量挑选在吸收峰处的波长用于模型的建立,有助于获得较高分析精度;图 3 中波长数较多的后半段的校正集和预测集标准偏差曲线与趋势线的重叠程度较好,图 4 中后半段的校正集和预测集标准偏差均高于趋势线,这说明在同等波长数参与建模的情况下,多利用吸收峰处的波长与等间隔增加波长相比,能够利用更多有用信息,从而得到更高的分析精度。

## 3 不同信噪比波段中多波长参与建模对分析精度的影响

依然在叠加 10% 的高斯噪声条件下,由第三组分的吸收谱曲线可知:第 1 和 2 吸收峰处,信号大于噪声 1 个量级;第 3 和 4 吸收峰处,信号与噪声同量级;第 5 和 6 吸收峰处,噪声大于信号 1 个量级。分别在不同信噪比的三个波段吸收峰及其两侧挑选 1, 4, 16, 64 个波长参与偏最小二乘回归模型的建立。由于实验结果中校正集和预测集的趋势相同,只列出校正集建模结果,如表 4 所示。其中,校正集标准偏差与参与建模的波长数的关系如图 5 所示。

Table 4 Modeling results of different wave bands

参与建模的波长数	第一波段		第二波段		第三波段	
	$R^2$	RMSEC	$R^2$	RMSEC	$R^2$	RMSEC
1	0.00	$2.9 \times 10^{-3}$	0.00	$2.9 \times 10^{-3}$	0.00	$2.9 \times 10^{-3}$
4	0.51	$2.0 \times 10^{-3}$	0.06	$2.8 \times 10^{-3}$	0.03	$2.9 \times 10^{-3}$
16	0.79	$1.3 \times 10^{-3}$	0.15	$2.7 \times 10^{-3}$	0.04	$2.8 \times 10^{-3}$
64	0.92	$8.2 \times 10^{-4}$	0.23	$2.5 \times 10^{-3}$	0.18	$2.6 \times 10^{-3}$

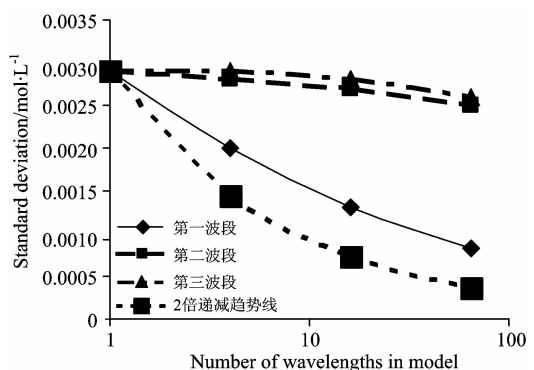


Fig. 5 Relationship between standard deviation and wavelength number

由图 5 中可以看出,当用于建模分析的光谱中的所有信号均小于仪器噪声时(第二、三波段),仅通过增加参与波长数,分析精度并没有明显提高,这可能是由于在增加波长数

而获得更多光谱信息的同时也引入了过大的噪声；当用于建模分析的光谱信号大于仪器噪声时(第一波段)，通过适当增加波长数，能够使分析精度得到明显提高。

## 4 结 论

利用三组分纯吸收光谱模型，使用偏最小二乘回归方法，对不同信噪比对组分浓度定量分析精度的影响、参与建模的波长数与分析精度的关系、不同信噪比波段中多波长参与建模对分析精度的影响进行定量分析。

实验结果表明：光谱测量仪器的信噪比对建模分析误差起着决定性的作用，使用信噪比较高的光谱仪器是提高物质成分定量分析精度的必要保证。使用多波长线性建模方法对噪声有一定的抑制作用。参与建模的光谱数据波长数与模型精度近似满足波长数每增加 4 倍，分析精度能够提高 2 倍；增加信噪比较高波段的光谱数据能够有效的提高定量分析精度。

研究为在光谱分析中提高和利用光谱仪器信噪比、合理选择参与建模波长数以及充分利用信噪比较高波段以提高光谱定量分析精度提供了实验基础和理论指导。

## References

- [1] Swayze G A, Clark R N, Goetz A F H, et al. *Journal of Geophysical Research*, 2003, 108(E9): 5105.
- [2] LI Qing-bo, XU Ke-xin, WANG Yan(李庆波, 徐可欣, 汪 曩). *Journal of Tianjin University(天津大学学报)*, 2003, 36(2): 139.
- [3] YU Lei, LIN Guan-yu, QU Yi, et al(于 磊, 林冠宇, 曲 艺, 等). *Journal of Optoelectronics · Laser(电子·激光)*, 2011, 22(1): 42.
- [4] Rocco Difoggio. *Applied Spectroscopy*, 2000, 54(3): 94A.
- [5] Schneider Hendrik, Reich Gabriele. *Analytical Chemistry*, 2011, 83(6): 2172.
- [6] Foca G, Ferrari C, Sinelli N, et al. *Analytical and Bioanalytical Chemistry*, 2011, 399(6): 1965.
- [7] Gowen A A, Downey G, Esquerre C, et al. *Journal of Chemometrics*, 2011, 25(7): 375.
- [8] ZHANG Xin-gang, WANG Ze-zhong(张新刚, 王泽忠). *Automation of Electric Power Systems(电力系统自动化)*, 2004, 28(14): 58.
- [9] LI Gang, ZHANG Li-jun, LIN Ling, et al(李 刚, 张丽君, 林 凌, 等). *Acta Electronica Sinica(电子学报)*, 2008, 36(7): 1465.
- [10] LI Gang, TANG Hong-ying, LIN Ling(李 刚, 汤宏颖, 林 凌). *Journal of Tianjin University(天津大学学报)*, 2010, 43(10): 901.

# The Effects of Signal to Noise Ratio of Instrument and Number of Wavelengths on the Accuracy of Spectral Analysis

LI Gang<sup>1</sup>, ZHAO Zhe<sup>1</sup>, WANG Hui-quan<sup>1</sup>, LIN Ling<sup>1</sup>, WU Xiao-rong<sup>2</sup>, ZHANG Bao-ju<sup>2\*</sup>

1. State Key Laboratory of Precision Measurement Technology and Instruments, Tianjin University, Tianjin 300072, China

2. College of Physics & Electronic Information, Tianjin Normal University, Tianjin 300387, China

**Abstract** In order to study the influence of the two factors, the signal to noise ratio (SNR) of instrument and number of wavelengths, on the accuracy of quantitative analysis in spectral modeling analysis, the influences of different SNRs, the number of wavelengths participating in partial least square regression, and multi-wavelength modeling in different SNR bands were taken into account. The relationships between these factors and quantitative analysis accuracy were explored by modeling analysis experiments. The results show that the analysis accuracy is directly related to the SNR of instrument. And increasing wavelengths, especial the ones with high SNR, reflects a phenomenon that four times of wavelength number will bring 2 times accuracy increasing, which can compensate for the lack of SNR. The research provides the experimental basis and theoretical guidance to improve the model's quality and reduce the prediction error by the utilization and improvement of the instrument's SNR and the reasonable choice of wavelengths and bands of modeling.

**Keywords** Signal to noise ratio; Number of wavelengths; Spectral analysis accuracy; Pure absorption; Partial least square

(Received Nov. 5, 2011; accepted Feb. 20, 2012)

\* Corresponding author