中国光学（中英文） Chinese Optics 核心期刊

基于多尺度小波变换的光谱数据预处理算法

钱方 许永博 赵伟

A spectrum signal pre-processing algorithm based on multi-scale wavelet transform

QIAN Fang, XU Yong-bo, ZHAO Wei

引用本文:
钱方, 许永博, 赵伟. 基于多尺度小波变换的光谱数据预处理算法[J]. 中国光学, 2025, 18(4): 794-802. doi: 10.37188/CO.2024-0230
QIAN Fang, XU Yong-bo, ZHAO Wei. A spectrum signal pre-processing algorithm based on multi-scale wavelet transform[J]. *Chinese Optics*, 2025, 18(4): 794-802. doi: 10.37188/CO.2024-0230

在线阅读 View online: https://doi.org/10.37188/CO.2024-0230

您可能感兴趣的其他文章

Articles you may be interested in

小波变换下基于分形编码和 Zernike 矩的织物图像检索算法
Fabric image retrieval algorithm based on fractal coding and Zernike moment under the wavelet transform
中国光学（中英文）. 2023, 16(3): 715 https://doi.org/10.37188/CO.EN-2022-0021

多尺度奇异值分解的偏振图像融合去雾算法与实验
Multi-scale singular value decomposition polarization image fusion defogging algorithm and experiment
中国光学（中英文）. 2021, 14(2): 298 https://doi.org/10.37188/CO.2020-0099

高超噪比宽带毫米波噪声信号光子学产生研究
Photonics generation of broadband millimeter wave noise signals with high excess noise ratios
中国光学（中英文）. 2022, 15(2): 251 https://doi.org/10.37188/CO.2021-0158

利用低信噪比小样本太赫兹光谱实现心肌淀粉样变检测
Detection of myocardial amyloidosis by a small number of terahertz spectra with low signal-to-noise ratio
中国光学（中英文）. 2022, 15(3): 443 https://doi.org/10.37188/CO.2021-0223

多尺度注意力融合的图像超分辨率重建
Image super-resolution reconstruction with multi-scale attention fusion
中国光学（中英文）. 2023, 16(5): 1034 https://doi.org/10.37188/CO.2023-0020

基于希尔伯特-黄变换的双马赫-曾德分布式光纤传感振动定位方法
A Hilbert-Huang transform method for vibration localization based on a dual Mach-Zehnder distributed optical fiber sensor
中国光学（中英文）. 2021, 14(6): 1410 https://doi.org/10.37188/CO.2021-0065

文章编号 2097-1842(2025)04-0794-09

基于多尺度小波变换的光谱数据预处理算法

钱 方[1*], 许永博[2], 赵 伟[3]
（1. 中国电子科技集团有限公司电子科学研究院, 北京 100041;
2. 苏州长光华芯光电技术股份有限公司, 江苏 苏州 215100;
3. 西昌卫星发射中心 四川 西昌 615000）

摘要: 光谱技术可以从大量原始信号中提取有用的特征信息, 直接用来分析和识别被观测样品的物质成分, 在生物医药、食品安全、军事侦察中具有极高的应用价值。由于预处理目的与效果不同, 目前存在多种光谱预处理方法。针对目前光谱预处理方法使用时存在的问题, 本文提出了一种基于多尺度小波变换的光谱数据预处理方法, 并利用仿真光谱和实测光谱对提出算法和设计软件的性能进行了测试。信噪比为 0.5 dB 的仿真信号经本文算法处理后, 信噪比可达 8.978 dB。仿真中加入 5 种不同类型的基线, 包括线型、高斯型、多项式型、e 指数型、sigmoidal 函数型, 使用本文算法进行基线估计, 估计值的均方根误差 RMSE 分别为 0.3759、0.2883、0.6631、0.3489、0.4520。使用共聚焦显微拉曼光谱仪测量了聚四氟乙烯光谱, 并用本文算法进行了预处理。结果表明该算法具有良好的可操作性, 能够有效去除噪声和校正基线, 并完整保留谱峰信息, 该算法为光谱数据预处理提供了新思路。

关 键 词: 光谱探测; 光谱信号处理; 小波变换; 去噪声; 基线校正
中图分类号: Tb96 **文献标志码:** A **doi:** 10.37188/CO.2024-0230 **CSTR:** 32171.14.CO.2024-0230

A spectrum signal pre-processing algorithm based on multi-scale wavelet transform

QIAN Fang[1*], XU Yong-bo[2], ZHAO Wei[3]
（1. *China Academy of Electronics and Information Technology*, *Beijing* 100041, *China*;
2. *Suzhou Everbright Photonics Co.*, *Ltd.*, *Suzhou* 215100, *China*;
3. *Xichang Satellite Launch Center*, *Xichang* 615000, *China*）
* *Corresponding author*, *E-mail*: *qfmail@sina.cn*

Abstract: Spectral technology can extract useful characteristic information from a large number of raw signals, which can be directly utilized for analyzing and identifying the material components of the observed samples. It has high application value in fields such as biomedicine, food safety and military reconnaissance. Due to the varying objectives and effects of the pretreatment, there are currently multiple spectral pre-processing methods available. We propose a spectrum signal pre-processing algorithm based on multi-scale wavelet transform, and the performance of the proposed algorithm and the designed softwere are evaluated through tests using both simulated and experimental spectra. The signal-to-noise ratio (SNR) of the simu-

收稿日期: 2024-12-26; **修订日期:** 2025-01-22
基金项目: 吉林省自然科学基金（No. 20210203174SF）
Supported by Natural Science Foundation of Jilin Province (No. 20210203174SF)

lated signal is 0.5 dB. After processing with the algorithm proposed in this paper, the SNR can reach to 8.978 dB. In the simulation, five different types of baselines are introduced, including linear, Gaussian, polynomial, exponential, and sigmoidal function types. Baseline estimation is performed using the algorithm proposed in this paper. The root mean square errors (RMSE) of the estimated values are 0.375 9, 0.288 3, 0.663 1, 0.348 9, 0.452 0, respectively. The spectrum of Polytetrafluoroethylene was measured using a confocal micro-Raman spectrometer and preprocessed with the algorithm proposed in this paper. The results demonstrate that the algorithm is capable of fast and accurate processing of the spectra. The algorithm can be used to reduce noise and correct baseline. This study put on a set of new ideas on spectrum signal processing.

Key words: spectral detecion; spectrum signal processing; wavelet transformation; de-noising; baseline correction

1　引　言

20 世纪 80 年代初, 光谱探测技术逐渐兴起, 90 年代后形成研发热潮, 由于其兼有探测和光谱分析的优点, 已经广泛应用于生物医药、食品安全、军事目标侦察等方面。光谱探测技术能够在不同谱段上对同一地物同步进行分谱成像, 获取随波长分布的光谱数据。光谱数据揭示了各种地物的光谱特性、存在状况以及物质成分, 极大提高了目标探测的准确性。随着光谱技术的发展, 越来越多的复杂体系或高维光谱数据被检测出, 但是测量过程中, 激光的发射噪声、探测器的散粒噪声、宇宙射线、荧光背景和环境中的黑体辐射等都会降低测试结果的准确性和稳定性。因此, 需要对光谱数据进行预处理以去除干扰信号的影响, 预处理方法主要用于调整数据变量之间的差异和关联关系, 消除光谱信号的干扰或冗杂信息, 也可以对有缺失、有错误的谱图数据进行处理[1-4]。

在处理不同性质的光谱数据时, 预处理和特征选择方法对光谱计量学多元模型的性能有很大的影响, 不同的预处理方法组合及其使用顺序都会对模型性能产生不同的影响[5]。预处理方法使用不当或过多, 会造成部分原始光谱数据的丢失或失真, 从而降低光谱分析精度, 使得采用预处理后数据的后续分析相比采用原始数据变得更差[6-8]。

Abraham Savizky 提出了一种经典的光谱预处理算法: Savitzky-Golay 滤波法。该方法通过一个移动窗口扫描数据, 对窗口内部的元素做多项式最小二乘拟合, 得到窗口中心位置处元素平滑后的值。后期, Schulze H. G.对这种方法进行了改进, 提出了循环三点零阶 Savitzky-Golay 滤波法, 即窗口宽度为 3, 多项式拟合阶数为 0。由于窗口宽度最小, 每次循环滤波所滤除的噪声量最少, 故可以缓慢地去除原始光谱中的噪声, 较好地保留光谱中的有用信息[9]。Donoho D L[10] 提出了软阈值(Soft thresholding)、硬阈值(Hard thresholding)去噪算法, 为去噪算法阈值选取提供理论分析。赵瑞珍提出了基于稀疏表示的小波去噪方法[11], 其具有较好的鲁棒性和有效性。刘帅奇提出了小波-Contourlet 变换与 Cycle Spinning 相结合的去噪算法[12], 在提高图像质量方面具有较为明显的优势, 为小波去噪算法在图像视觉增强方面的应用提供了思路。

Phillips G. R.等提出了缺点多项式滤波器法[13]、Liu H C 提出了 Hampel 滤波器法[14]、Yukiteru 等提出了循环消去法, 此类方法可以去除噪声引起的光谱尖峰, 但是它们一般都需要设定一些参数, 从而增加了计算冗余度, 不适合处理大量数据。

光谱预处理算法还包括基线校正, Gan F 提出的多项式拟合法(文中简称 PF)是最常用的全自动去基线方法。在该方法中, 选择合适的拟合阶数是能否准确去除基线的关键[15]。Schulze H G 提出的小窗口滑动法(文中简称 SWMA)采用宽度为 3 的滑动窗口对原始光谱进行平滑, 将原始光谱信号值大于平滑后光谱信号值的位置作为谱峰, 并将其去除, 重复这一过程并逐渐增大窗口宽度, 循环消去的谱峰面积一般会先增大后减小, 当谱峰的面积再次增大时, 说明谱峰已经基本去除, 此时停止循环, 即可得到估计出的基线[16]。自适应循环加权惩罚最小二乘法(文中简称 AIRPLS)是一种半自动基线去除方法, 计算中需要设定平滑控制参数和表征粗糙度的差分阶数等参

数[17]。这类半自动方法一般去基线效果较好，但是设置参数会增加计算的复杂度。

小波变换近年来已成功应用于许多领域的信号处理中。小波变换是一种时间-频率分析方法，具有多分辨率分析的特点，可以表征信号的局部特征，低频部分具有较高的频率分辨率和较低的时间分辨率，高频部分具有较高的时间分辨率和较低的频率分辨率，适于探测正常信号中出现的瞬态反常现象并展示其成分。同时，学者们致力于小波变换与其他算法联合使用的研究，有望在信号分析中得到更为广泛的应用[18]。

本文提出了一种基于多尺度小波变换的光谱预处理算法，包括基于小波动态阈值的光谱信号去噪方法（dynamic threshold de-noising algorithm, DTD）和基于小波特征点和线性分段插值的光谱信号去基线方法（baseline correction algorithm based on wavelet feature points and segment interpolation, WFPSI）。以上算法能够较好地去除干扰信号并保留谱峰信息，同时计算过程中无需设置任何参数，避免了人为设定参数带来的不确定性，计算结果准确、操作简便、易行。

2 多尺度小波光谱预处理算法分析

2.1 去噪算法

基于小波动态阈值的光谱信号去噪方法（DTD）原理为：对信号进行小波变换后，设置一个阈值，认为小于该阈值的小波系数是噪声所引起，将其归零，大于该阈值的小波系数是有用信号，将其保留或者做某种收缩，最后对处理过的小波系数做小波逆变换，重构出去除噪声的信号。

大量实验结果表明，5~6 级小波变换重构后更加接近原始信号，但其运算数据量过大，因此，一般认为在保证精确性的前提下，运算复杂度不高的 4 级小波变换比较适合实际应用。本文方法中，首先对原始信号进行 4 级小波变换，然后分别计算每层小波系数的去噪阈值，并分别对每层小波系数进行滤波，最后做小波逆变换，重构出光谱信号。该方法不需要人工设置任何参数，可实现光谱去噪的完全自动化。

常用的小波基函数包括：Harris、Daubechies、Biorthgonal、Coiflets 和 Symlets，考虑到光谱信号的采样频率较高、数据量较大，为减少计算冗余度，需要采取能够进行离散小波变换并具有正交性的小波基函数。多数光谱信号属于非线性信号，需要选择在时域和频域上快速衰减即具有支撑性的小波基函数。为了避免降噪结果出现失真现象，保证降噪结果与有用信号的高相似性，需要采取具有较高对称性的小波基函数。由于 Harris 小波的局部化分析能力较弱，Biorthgonal 小波不具有对称性，Symlets 小波在支撑长度和对称性方面优于 Daubechies 和 Coiflets 小波。综上所述，本文选取“Sym8”作为小波基函数。

如图 1 所示，对原始信号进行小波变换后，噪声通常包含在 cD1、cD2、cD3 和 cD4 中，可以通过设定阈值的方式对小波系数进行处理，达到去噪的目的。由于进行了 4 级小波变换，如果仅设置一个滤波阈值，在较低的尺度上，单一阈值会使一部分有用信号被去除，而在较高的尺度上，单一阈值又会导致一部分噪声被保留。因此，在所有尺度上都采用同一阈值是不合适的。本文算法对不同的分解尺度采用不同的阈值。

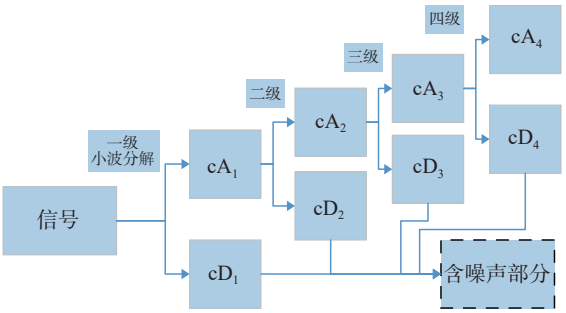

图 1 小波去噪方法原理图

Fig. 1 Principle diagram of wavelet de-noising

小波阈值一般采用经验值或通过大量试验得到，阈值计算方法如公式（1）所示：

$$T_l = \sigma_l \sqrt{2\ln(N_l)} \quad , \tag{1}$$

其中，T_l 为第 l 层系数阈值，N_l 是第 l 层小波系数的长度，σ_l 为子带内噪声标准差，按照鲁棒估计法估计小波系数的噪声标准差，如公式（2）所示：

$$\sigma_l = \frac{\mathrm{middle}(|W_l|)}{0.674\,5} \quad , \tag{2}$$

其中，W_l 是第 l 层的小波系数，middle 为取平均数。

确定阈值后，对各尺度的小波系数进行调整，计算方法见式（3）：

$$cD_l = \begin{cases} \mathrm{sign}(cD_l) \times (|cD_l| - T_l), & |cD_l| > T_l \\ 0, & |cD_l| \leqslant T_l \end{cases} , \tag{3}$$

其中, cD_l 是分解后第 l 层的小波细节系数, sign 是符号函数。确定以上参数后, 可以对光谱数据进行小波阈值去噪, 算法流程图如图 2 所示。

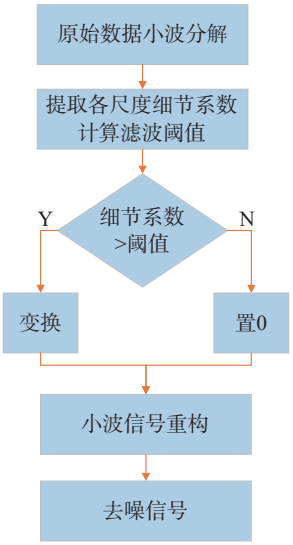

图 2　小波阈值去噪算法流程图

Fig. 2　Flow chart of wavelet de-noising

宇宙射线或者环境中物质发出的射线照射到探测器上, 会形成一系列非常尖锐的峰型, 其宽度通常不超过 2 cm^{-1}, 称为尖峰。光谱信号中尖峰的存在会引起基线校正误差, 应将其去除。

去尖峰算法流程图如图 3 所示, 光谱去噪后, 提取信号局部极值点, 确定谱峰可能出现的位置, 然后在可能的谱峰点处分析该峰的半峰全宽, 若它的值≤1 cm^{-1}, 则认为该处为尖峰, 最后对尖峰处及受尖峰影响的值重新进行拟合, 得到去尖峰的光谱。

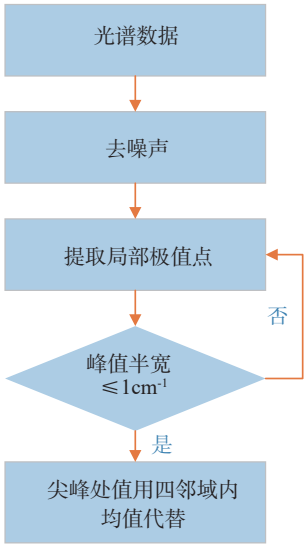

图 3　去尖峰算法流程图

Fig. 3　Flow chart of spike removal algorithm

2.2　去基线算法

光谱中的基线主要由样品及基底的荧光或磷光背景、样品及其周围环境的黑体辐射形成, 一般表现为缓慢变化的曲线。基线的存在会影响后续光谱的归一化及匹配等操作, 需提前将其去除。

基于小波特征点和线性分段插值的光谱信号去基线算法(WFPSI)原理为: 通过小波变换确定光谱曲线特征点, 然后在选定的特征点间做曲线拟合, 计算基线, 用原始谱线与估计的基线相减, 得到去基线光谱, 算法流程图如图 4 所示。

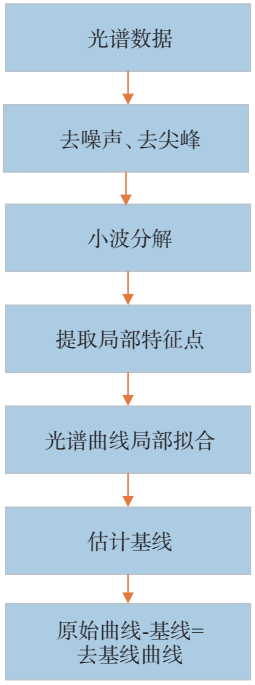

图 4　基线校正算法流程图

Fig. 4　Flow chart of baseline correction

算法首先对光谱信号进行连续小波变换, 通过对比原始信号和小波信号确定局部特征点, 见图 5(彩图见期刊电子版)。

特征点属性如表 1 所示。谱峰、波谷、起点和终点在小波变换信号中均为局部极小值; 谱峰在光谱信号中为局部极大值, 且它的一阶导数通过零点, 用“∨”表示; 波谷在光谱信号中为局部极小值, 且它的一阶导数通过零点, 用“∨”表示; 起点和终点在光谱信号中为局部极小值, 且它们的一阶导数不通过零点, 用“×”表示。通过比较这些特征点的属性, 可以识别出各光谱特征点。分别在各个谱峰的起点和终点之间做线性分段插值拟合。考虑到插值算法的合理性和冗余性, 分段插值步长取 3 个数据点。通过上述分析可以估

计出整个光谱的基线，将原始光谱曲线与估计的基线相减，可以得到基线校正后的光谱。

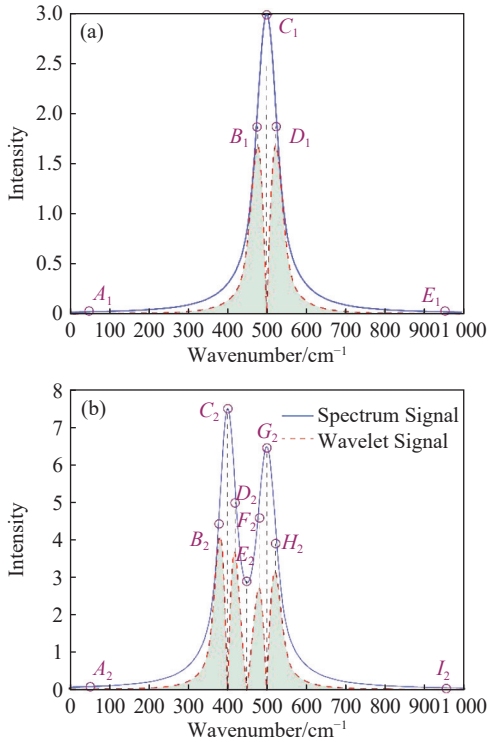

图 5 光谱信号与小波信号对比图。(a) 单谱峰；(b) 多谱峰

Fig. 5 Comparison between spectral signal and wavelet signal. (a) Single spectrum peak; (b) mult-spectrum peaks

表 1 光谱特征点属性

Tab. 1 Attribute of spectral characteristic

谱峰类型	单峰	多峰	光谱值	小波变换值	一阶导数是否通过零点
起始点	A_1	A_2	最小值	最小值	×
左拐点	B_1	B_2, F_2	上升沿	最大值	×
谱峰	C_1	C_2, G_2	最大值	最小值	√
右拐点	D_1	D_2, H_2	下降沿	最大值	×
波谷	--	E_2	最小值	最小值	√
结束点	E_1	I_2	最小值	最小值	×

3 实验结果与分析

3.1 仿真数据分析

3.1.1 去噪算法对比

实验中仿真了一种包含 3 个单峰和 3 个重叠峰的理想光谱，其中谱峰位置、高度和半峰全宽的设置如表 2 所示，表中半峰全宽，指高度为谱峰高度一半处的谱带全宽，仿真光谱如图 6 所示。

表 2 仿真光谱信号

Tab. 2 Simulated spectral signal

谱峰位置(cm⁻¹)	645	1090	1535	2130	2200	2270
谱峰高度	3	6	9	6	6	6
半峰全宽	25	30	35	40	45	50

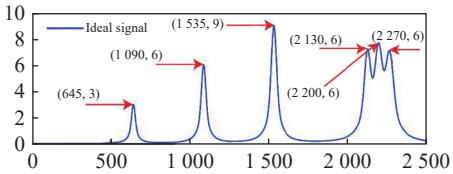

图 6 仿真光谱信号去噪效果对比图

Fig. 6 Comparison of denoising effects of simulated spectral signals

图 7 是仿真数据的去噪效果图，图中原始信号信噪比为 0.5 dB，采用 Savitzky-Golay 滤波法、软阈值法（Soft thresholding）、硬阈值法（Hard thresholding）、本文 DTD 算法 4 种方法去噪。去噪前后信噪比见表 3。

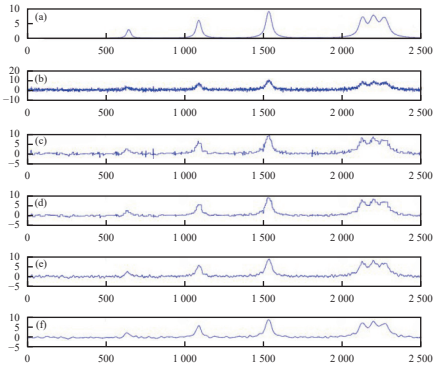

图 7 仿真数据去噪效果对比图。(a) 原始信号；(b) 加噪声信号；(c) 硬阈值去噪；(d) 软阈值去噪；(e) Savitzky-Golay 去噪；(f) 本文算法去噪

Fig. 7 Contrast map of simulated de-noising spectral signal. (a) Initial signal; (b) signal with noise; (c) hard thresholding denoising; (d) soft thresholding denoising; (e) Savitzky-Golay denoising; (f) DTD denoising

表 3 不同方法去噪后信号的信噪比

Tab. 3 SNR of signals after denoising using different methods

	原始信号	硬阈值算法	软阈值算法	Savitzky-Golay 滤波算法	本文算法 DTD
SNR(dB)	0.5	8.0621	8.3231	8.4106	8.9775

3.1.2　去基线算法对比

在仿真光谱中加入 5 种不同类型的基线, 包括线型、高斯函数型、多项式型、e 指数型和 sigmoidal 函数型, 如图 8 所示。使用多项式拟合法(PF)、小窗口滑动法(SWMA)、自适应循环加权惩罚最小二乘法(AIRPLS)和本文提出的算法(WFPSI)对仿真数据进行处理, 得到去基线的光谱数据。文中计算了仿真基线和估计基线之间的均方根误差(Root mean square error, RMSE), 以比较不同算法的去基线效果。

$$RMSE = \sqrt{\sum_{n=1}^{L}[BL_{\text{ideal}}(n) - BL_{\text{est}}(n)]^2 / L} \quad , \quad (4)$$

式中 L 是光谱信号长度, BL_{ideal} 是仿真基线, B_{Lest} 是算法估计的基线, n=1, 2, …, L。

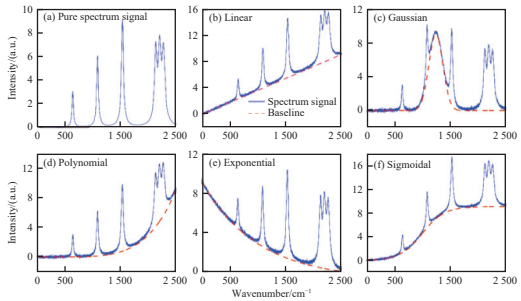

图 8　加入基线的光谱图。(a) 原始信号; (b) 加入线型基线; (c) 加入高斯型基线; (d) 加入多项式型基线; (e) 加入 e 指数型基线; (f) 加入 sigmoidal 函数型基线

Fig. 8　Map of spectral signal with baseline. (a) Initial signal; (b) signal with linear baseline; (c) signal with Gaussian baseline; (d) signal with polynomial baseline; (e) signal with exponential baseline; (f) signal with sigmoidal baseline

为了比较所提算法的性能, 仿真实验中分别测试不同基线类型和不同噪声水平下算法的基线校正效果, 即分别计算不同算法在不同基线、不同信噪比水平下的 RMSE 的值, 其中信噪比系数为: 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 和 2, 4, 6, 8, 10。表 4 是不同算法的平均 RMSE 值。

从表中数据可知, 本文算法方法的准确率高于 PF 和 SWMA 法, 在高斯函数型、e 指数型和 sigmoidal 函数型基线的校正中, WFPSI 的准确率高于 AIRPLS 方法, 在线型和多项式型基线的校正中, 本文算法 WFPSI 的准确率略低于 AIRPLS 方法, 但本文算法在使用过程中不需要设置参数, 可以实现基线自动校正, 更适用于处理大量数据。

表 4　不同算法的去基线性能比较

Tab. 4　Comparison of baseline removal performance among different algorithms

基线类型	PF算法	AIRPLS算法	SWMA算法	本文算法 WFPSI
线型	1.2629	0.3141	0.4751	0.3759
高斯函数型	1.0733	0.3290	0.4881	0.2883
多项式型	1.2986	0.6501	0.7076	0.6631
e指数型	1.2755	0.3953	0.4729	0.3489
sigmoidal函数型	1.2562	1.2178	1.2330	0.4520

3.2　实测数据分析

实验中使用的拉曼光谱仪为 Renishaw 公司生产的 inVia 型共聚焦显微拉曼光谱仪, 如图 9 所示。仪器的分辨率<1 cm^{-1}, 激光激发波长为 488 nm, 最大输出功率约为 200 mw, 测试光谱范围设置为 100~3 200 cm^{-1}, 测试聚四氟乙烯的拉曼光谱。

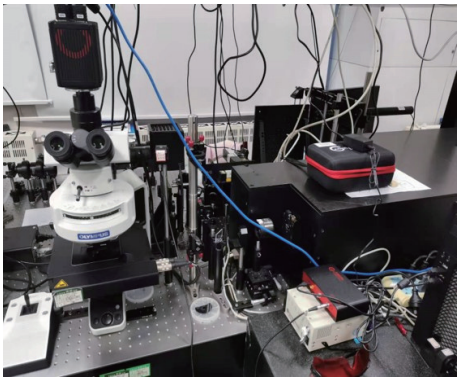

图 9　inVia 型共聚焦显微拉曼光谱仪

Fig. 9　inVia confocal micro Raman spectrometer

实验中采用 488 nm 激光作为激发光源, 准直激光经过柱透镜与汇交透镜形成线光斑, 然后通过物镜共轭成像产生激发线光斑, 成像至样品上。样品被激发产生的信号光沿原光路返回。在形成的二维光谱图像中, 垂直维度反映了样品沿激发线光斑的空间位置信息, 水平维度反映了不同波长的光谱信息。光谱成像原理图和测试样品载玻片如图 10—11 所示。

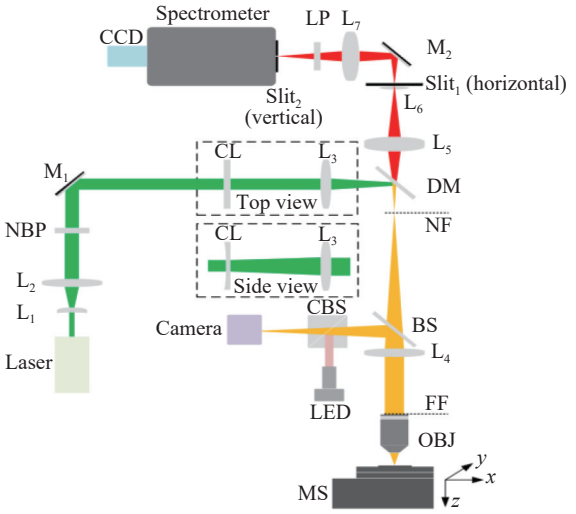

图 10 光谱成像原理图

Fig. 10 Principle diagram of spectral imaging

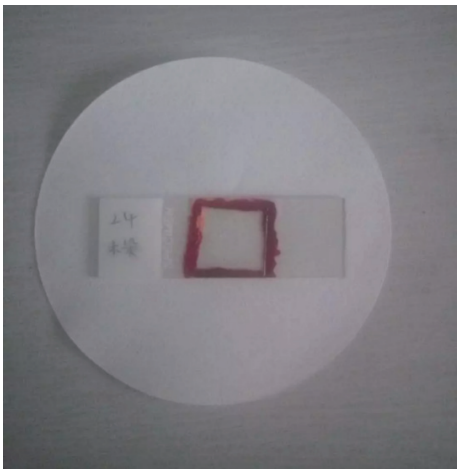

图 11 测试样品载玻片

Fig. 11 Sample slide

3.2.1 去噪声算法对比

使用硬阈值法、软阈值法、Savitzky-Golay、本文 DTD4 种算法对实测光谱信号进行处理，对比去噪效果。

如图 12 所示，光谱曲线后侧的方框内是放大的局部细节图，可以看出 Savitzky-Golay 和本文 DTD 算法的去噪效果较好，去噪后光谱曲线较为平滑，但是 Savitzky-Golay 算法去噪后谱峰出现了失真，半峰全宽出现了明显改变，而且 Savitzky-Golay 算法需要设定多项式拟合阶数和窗口大小，会导致测试结果出现不确定性。本文提出的 DTD 算法去噪效果较好，并完整地保留了谱峰信息，同时计算过程中不需要设定任何参数，可以实现光谱的自动去噪，简单易用。

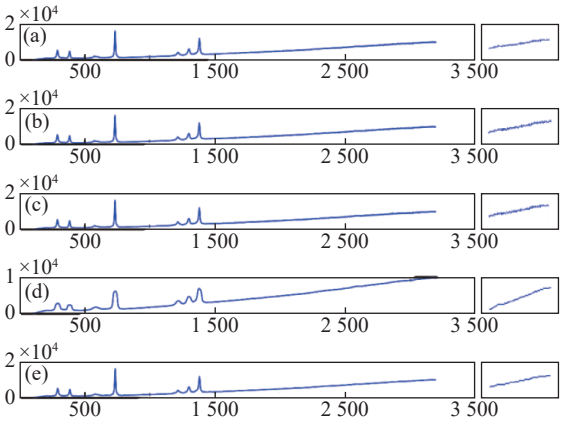

图 12 实测数据去噪对比图。(a) 原始信号；(b) 硬阈值去噪；(c) 软阈值去噪；(d)Savitzky-Golay 去噪；(e) 本文算法去噪

Fig. 12 Comparison of denoising results for measured data. (a) Initial signal; (b) hard thresholding denoising; (c) soft thresholding denoising; (d) Savitzky-Golay denoising; (e) DTD denoising

3.2.2 去基线算法对比

使用共聚焦显微拉曼光谱仪测量了聚四氟乙烯光谱，使用本文提出的 WFPSI 算法对采集光谱进行基线校正，计算结果如图 13（彩图见期刊电子版）所示。

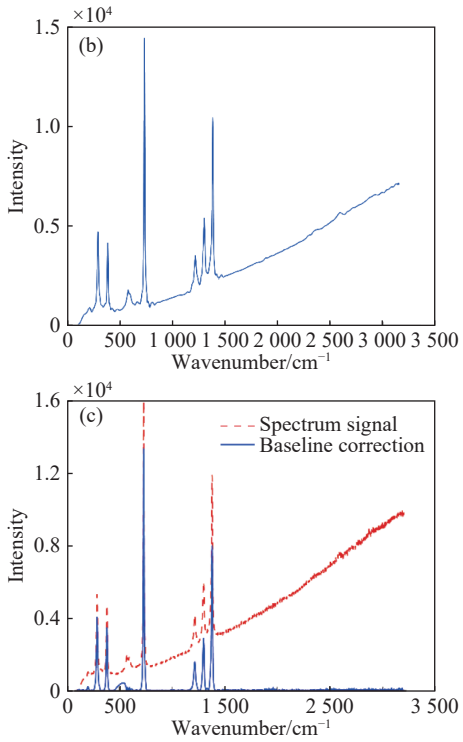

图 13 实验结果图。(a) 原始信号；(b) WFPSI 算法基线校正信号

Fig. 13 Measured spectral signal. (a) Initial signal; (b) corrected signal by WFPSI algorithm

为了比较不同算法对实测数据的基线校正效果, 计算了校正基线前后数据经主成分分析 (Principal components analysis, PCA) 后“凸包”面积的变化情况, 见表 5。“凸包”的致密性可以表征降维效果, 光谱数据基线校正后凸包面积减小百分比越高, 说明光谱数据分析和归类的准确性越高。

表 5　光谱数据基线校正后凸包面积减小百分比

Tab. 5　Percentage reduction in area of convex hull

算法	凸包面积减小百分比
PF算法	54.69%
AIRPLS算法	64.76%
SWMA算法	58.51%
WFPSI算法	66.21%

4　结　论

光谱数据预处理是建立精确的光谱计量学模型的一个重要环节, 本文主要围绕光谱预处理方法总结了一些去噪声、去基线算法, 提出了基于小波动态阈值的光谱信号去噪方法(DTD)和基于小波特征点和线性分段插值的光谱信号去基线方法(WFPSI), 通过仿真数据和实测数据验证表明, DTD 算法可以实现自动去噪功能, 并完整地保留谱峰信息。实验中初始仿真信号信噪比为 0.5 dB, 经本文算法处理后, 信噪比可达 8.978 dB, 结果优于经典滤波算法。仿真信号中加入 5 种不同类型的基线, 包括线型、高斯型、多项式型、e 指数型、sigmoidal 函数型, 使用本文算法进行基线估计, 估计值的均方根误差 RMSE 分别为 0.3759、0.2883、0.6631、0.3489、0.4520, 效果较好。此外, 本文提出的 WFPSI 算法计算过程中无需设置任何参数, 实现了光谱数据基线的自动校正。实验中还使用共聚焦显微拉曼光谱仪测量了聚四氟乙烯光谱, 使用本文算法进行了预处理, 通过主成分分析后“凸包”面积的变化分析了算法优劣,

经本文算法校正后, 光谱数据基线校正后凸包面积减小百分比为 66.21%, 优于经典算法。但是由于实际探测物质的多样性, 不同地物的光谱数据存在较大差异, 单一的数据预处理方法因其自身的局限性往往很难达到理想的处理效果, 在今后的研究中应测量大量样本, 设计、完善和验证算法。

参考文献:

[1] 孙嘉豪, 张伟, 施鉴芩, 等. 光谱数据预处理策略选择及应用[J]. 计量学报, 2023, 44(8): 1284-1292.
SUN J H, ZHANG W, SHI J Q, *et al*.. Selection and application of spectral data preprocessing strategy[J]. *Acta Metrologica Sinica*, 2023, 44(8): 1284-1292. (in Chinese).

[2] 周粲入, 王哲涛, 杨思危, 等. 化学计量学和深度学习方法在拉曼光谱处理方面的应用研究进展[J]. 分析化学, 2023, 51(8): 1232-1242.
ZHOU C R, WANG ZH T, YANG S W, *et al*.. Application progress of chemometrics and deep learning methods in Raman spectroscopy signal processing[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(8): 1232-1242. (in Chinese).

[3] 张雪容, 梁维新, 杨玉敏, 等. 顶空固相萃取-纸基表面增强拉曼光谱法快速测定水中痕量汞[J]. 分析化学, 2023, 51(9): 1536-1544.
ZHANG X R, LIANG W X, YANG Y M, *et al*.. Determination of trace mercury in water by headspace solid phase extraction combining paper-based surface enhanced Raman spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(9): 1536-1544. (in Chinese).

[4] 刘成员, 于江玉, 李奉翠, 等. 拉曼光谱测试技术在可充电铝离子电池储能机理的研究进展[J]. 应用化学, 2023, 40(10): 1347-1358.
LIU CH Y, YU J Y, LI F C, *et al*.. Research progress of Raman spectroscopy technique in energy storage mechanism of rechargeable aluminum-ion batteries[J]. *Chinese Journal of Applied Chemistry*, 2023, 40(10): 1347-1358. (in Chinese).

[5] 李艳坤, 董汝南, 张进, 等. 光谱数据解析中的变量筛选方法[J]. 光谱学与光谱分析, 2021, 41(11): 3331-3338.
LI Y K, DONG R N, ZHANG J, *et al*.. Variable selection methods in spectral data analysis[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3331-3338. (in Chinese).

[6] ESTEVES C S M, DE REDROJO E M M, MANJÓN J L G, *et al*.. Combining FTIR-ATR and OPLS-DA methods for

magic mushrooms discrimination[J]. *Forensic Chemistry*, 2022, 29: 100421.

[7] CAMPOS M P, REIS M S. Data preprocessing for multiblock modelling-a systematization with new methods[J]. *Chemometrics and Intelligent Laboratory Systems*, 2020, 199: 103959.

[8] YANG W Y, XIONG Y R, XU ZH ZH, *et al*.. Piecewise preprocessing of near-infrared spectra for improving prediction ability of a PLS model[J]. *Infrared Physics & Technology*, 2022, 126: 104359.

[9] SCHULZE H G, FOIST R B, IVANOV A, *et al*.. Fully automated high-performance signal-to-noise ratio enhancement based on an iterative three-point zero-order Savitzky–Golay filter[J]. *Applied Spectroscopy*, 2008, 62(10): 1160-1166.

[10] DONOHO D L. De-noising by soft-thresholding[J]. *IEEE Transactions on Information Theory*, 1995, 41(3): 613-627.

[11] ZHAO R ZH, LIU X Y, LI C C, *et al*.. Wavelet denoising via sparse representation[J]. *Science in China Series F: Information Sciences*, 2009, 52(8): 1371-1377.

[12] 刘帅奇, 胡绍海, 肖扬. 基于小波-Contourlet 变换与 Cycle Spinning 相结合的 SAR 图像去噪[J]. 信号处理, 2011, 27(6): 837-842.
LIU SH Q, HU SH H, XIAO Y. SAR image de-noised based on wavelet-Contourlet transform with Cycle Spinning[J]. *Signal Processing*, 2011, 27(6): 837-842. (in Chinese).

[13] PHILLIPS G R, HARRIS J M. Polynomial filters for data sets with outlying or missing observations: application to charge-coupled-device-detected Raman spectra contaminated by cosmic rays[J]. *Analytical Chemistry*, 1990, 62(21): 2351-2357.

[14] LIU H C, SHAH S, Jiang W. On-line outlier detection and data cleaning[J]. *Computers & Chemical Engineering*, 2004, 28(9): 1635-1647.

[15] GAN F, RUAN G H, MO J Y. Baseline correction by improved iterative polynomial fitting with automatic threshold[J]. *Chemometrics and Intelligent Laboratory Systems*, 2006, 82(1-2): 59-65.

[16] SCHULZE H G, FOIST R B, OKUDA K, *et al*.. A small-window moving average-based fully automated baseline estimation method for Raman spectra[J]. *Applied Spectroscopy*, 2012, 66(7): 757-764.

[17] ZHANG ZH M, CHEN SH, LIANG Y Z. Baseline correction using adaptive iteratively reweighted penalized least squares[J]. *Analyst*, 2010, 135(5): 1138-1146.

[18] ALI F, KABIR M, ARIF M. DBPPred-PDSD: Machine learning approach for prediction of DNA-binding proteins using discrete wavelet transform and optimized integrated features space[J]. *Chemometrics and Intelligent Laboratory Systems*, 2018, 182: 21-30.

作者简介：

钱　方(1987—)，女，吉林长春人，博士，高级工程师，2015 年于中国科学院长春光学精密机械与物理研究所获得博士学位，主要从事激光辐照效应研究。E-mail: qfmail@sina.cn

许永博(1998—)，男，山东临沂人，硕士，工程师，2024 年于中国科学院长春光学精密机械与物理研究所获得硕士学位，主要从事激光辐照效应研究。E-mail: xyb12172022@163.com