

〈红外应用〉

基于天然气组分红外光谱图的数据预处理方法研究

康明¹, 韩森坪¹, 杨洪杰¹, 唐德东¹, 李妍君¹, 汪智琦²

(1. 重庆科技学院 电气工程学院 重庆 401331; 2. 中国石油化工股份有限公司荆门分公司 湖北 荆门 448000)

摘要: 利用红外光谱分析仪对天然气组分进行组分分析时所获得光谱信号往往会受杂散光、噪声、基线漂移等因素的干扰, 从而影响最终定量分析结果, 故需要在建模前对原始光谱进行预处理。为解决仪器测量光谱图的噪声干扰问题, 本文提出一种 Savitzky-Golay 平滑法结合 sym6 小波函数软阈值去噪法对光谱图进行预处理。将传统的预处理方法与 SG 平滑法结合小波函数法进行对比分析。结果表明, 以 SG 平滑法结合 sym6 小波函数软阈值去噪法对光谱图进行预处理, 其拟合优度数值最高为 0.98652, 残差平方和数值最低为 5.50694, 证明使用该方法后的函数分峰拟合效果最佳, 处理效果优于传统方法。

关键词: 天然气; 预处理; 红外光谱图; 去噪

中图分类号: TE81 文献标识码: A 文章编号: 1001-8891(2021)08-0804-05

Data Preprocessing Method for Infrared Spectra Analysis of Natural Gas Components

KANG Ming¹, HAN Senping¹, YANG Hongjie¹, TANG Dedong¹, LI Yanjun¹, WANG Zhiqi²

(1. College of Electrical Engineering, Chongqing University of Science and Technology, Chongqing 401331, China;

2. China Petroleum & Chemical Corporation Jingmen Branch, Jingmen 448000, China)

Abstract: When using infrared spectroscopy to analyze the components of natural gas, the obtained spectral signals often contain interference from stray light, noise, baseline drift, and other factors, which affects the resulting quantitative analysis. Therefore, it is necessary to preprocess the original spectrum before modeling. As a potential solution, an SG smoothing method combined with the soft threshold denoising method of the sym6 wavelet function was proposed to preprocess the spectrogram. The traditional preprocessing method and the proposed method are compared and analyzed. The results show that when the proposed method is used to preprocess the spectrogram, the highest goodness of fit value is 0.98652, and the lowest residual sum of squares value is 5.50694, which proves that the function peak fitting effect is the best after using this method, and the processing effect is better than that of the traditional method.

Key words: natural gas, pretreatment, infrared spectra, denoising

0 引言

采用红外光谱分析仪对天然气组分进行分析时^[1]。由于仪器或检测环境、光照强弱等因素影响, 仪器检测得到的光谱数据中掺杂有噪声^[1], 该类噪声会直接影响到后续模型的判别性能, 故有必要对光谱图进一步进行预处理^[2]。

天然气组分红外光谱图的特点是数据量大^[3], 低含量区域信噪比低, 有效波段少, 噪声信号夹杂程度高。目前已有的光谱预处理方法种类繁多, 如何选择合适的预处理方法是亟待解决的问题^[4-5]。研究表明, 一阶和二阶导数, 虽能够有效地分离重叠峰及去除基线和背景, 但同时将高频噪声成分放大, 应用场景因此受到限制。Savitzky-Golay (SG) 平

收稿日期: 2021-01-05. 修订日期: 2021-01-19.

作者简介: 康明 (1997-), 女, 硕士研究生, 主要从事天然气组分图谱解析研究。E-mail: 1551043341@qq.com.

基金项目: 重庆市技术创新与应用发展专项 (cstc2019jcsx-msxmX0054); 重庆市重庆科技学院研究生科技创新计划项目 (YKJCX1920405)。

滑算法可以消除部分频率较高的噪声,提高图谱信噪比,但其性能与窗口大小相关,不合适的窗口大小反而影响数据质量,平滑过多或者过少均会使数据失真,导致精度降低^[6]。中值滤波器计算量小,运算速度快,在滤除脉冲噪声和扫描噪声方面表现了良好的性能^[7],但中值滤波技术在处理混有高斯白噪声的图像时效果不理想,可能会出现边缘位移的现象,因此适用于干扰较少的情况。小波分析是针对傅里叶变换(fast Fourier transform, FFT)的不足而发展出来的^[8-9]。因傅里叶变换在时域中没有分辨能力,小波变换则克服了傅里叶变换和 Gabor 变换的一些缺陷^[10],它可以根据窗口宽度进行时频调节,以此提高分辨率,比傅里叶变换更为宽广和有效,小波变换则具有优良的局部化视频分析特性。

综合上述多种预处理方法各自性能,并结合天然气多组分的实际情况,本文提出了 Savitzky-Golay 平滑法结合 sym6 小波函数软阈值去噪法对光谱图进行预处理,具体验证过程将在谱图处理进行详细阐述。

1 SG 平滑+sym6 小波变换原理和方法

1.1 SG 平滑法

SG 平滑法是把奇数个 $NSP=2m+1$ 光谱点看作一个窗口^[11],采用多项式法对窗口内的光谱数据做最小二乘拟合,利用得到的多项式系数计算出窗口中心点的各阶导数值和平滑数据值。在去掉窗口内最前端的数据后添加窗口最末端相邻的光谱数据,使得平滑窗口在整个图谱内移动,得到平滑后经不同导数分析后的图谱。设光谱波长为 λ ,经过多项式拟合进行平滑后的平均值为:

$$y_{\lambda} = \bar{y}_{\lambda} = \frac{1}{K} \sum_{-w}^{+w} y_{\lambda} k_i \quad (1)$$

式中: k_i 为平滑系数; K 为归一化因子, $K = \sum_{-w}^{+w} k_i$, 每一测量值乘以平滑系数 k_i 的目的是尽可能减小平滑对有用信息的影响。

1.2 小波变换法

连续小波^[10]变换其定义如下:

$$W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (2)$$

式中: 信号 $f(t)$ 的小波变换是一个二元函数; b 为时间中心参数; a 为尺度参数, a 表示以 $t=b$ 为中心的波动范围,可以改变频谱结构和窗口的形状,所以小波变换也是时频分析函数。小波变换具有优良的局部化视频分析特性,可以改变频谱结构和窗

口形状(窗口大小是不变的),对分析突变信号和奇异信号非常有效^[12]。

1.3 滑动窗口和小波基的选择

SG 平滑的性能与窗口大小相关,不合适的窗口大小反而影响数据质量,平滑过多或者过少均会使数据失真^[13],经过多次实验验证,当 $N=20$ 的卷积平滑后的混合光谱图处理效果相当,可以进行一定程度的去噪。

在利用小波去噪的过程中,能否选择合适的阈值函数和阈值直接影响着图像的去噪效果,此外,小波基的选择也是一个不可忽略的问题,因选择的小波基不同,产生的去噪效果也不同。常用的小波函数有 Haar、Symlets 及 Daubechies^[8]等,其中 Daubechies 基是具有紧支集的光滑正交小波基,其光滑性由支集的长度 N 来确定, Symlet 小波函数是 Ingrid Daubechies 提出的近似对称的小波函数,它是对 db 函数的一种改进^[14]。Symlet 小波系通常表示为 $\text{sym}N(N=2,3,\dots,8)$ ^[15]。sym N 小波的支撑范围为 $2N-1$,消失矩为 N ,同时也具备较好的正则性。该小波与 db N 小波相比,在连续性、支集长度、滤波器长度等方面与 db N 小波一致,但 sym N 小波具有更好的对称性,即一定程度上能够减少对信号进行分析和重构时的相位失真,经过多次实验验证,小波基 sym6 处理图谱效果较好。

本文实验选择将 sym6 小波与 db 小波进行对比分析,对于分解层数的确定,本次实验采用了 3 层和 4 层两种进行对比实验。阈值量化选择软阈值去噪,随后将处理后的低频系数和高频系数进行重构并输出谱图。

2 天然气组分光谱图预处理

2.1 实验系统搭建

实验原理基于朗伯比尔定律:物质对某一波长光吸收的强弱与吸光物质的浓度及其气体吸收池厚度间的关系可定量描述,如公式(3)所示:

$$A(\nu) = S(T) P \Gamma(\nu) c L \quad (3)$$

式中:影响吸光度 $A(\nu)$ 主要有温度 T 、压强 P 、气体浓度 c 以及光程 L ,其中气体浓度和光程是实验装置的固有特性,无法改变。而对于温度以及压强则需要按照实际情况进行控制和记录,保证实验结果的准确性。

整个系统主要由气体吸收池、温度加热控制仪、进出气阀门、傅里叶红外光谱仪、PC 机等组成。采样气体或背景气体由调节阀通过管路送入气体吸收池,同时利用温度加热控制仪进行保温操作,再利

用傅里叶红外光谱仪对其进行测量，最后结果则输入 PC 机进行相应数据处理，废气则由旁路排出。实验基本结构图如图 1 所示。

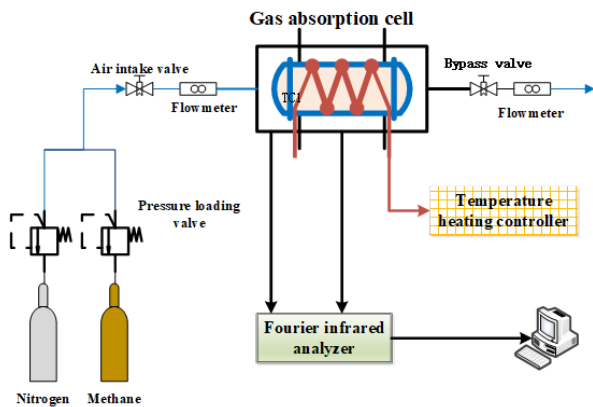


图 1 标准库建立实验结构图

Fig.1 Experiment structure of the standard library

此次实验的样品气体为标准气体；主要包括甲烷、乙烷和丙烷 3 种天然气的主要成分，4 种气体均为 2 L 纯度为 99.9% 的标准气体，此类气体主要用作标准数据库建立；同时也配置了其他相应浓度的标准气体，用作定量分析。而对于混合气体，则是按照天然气的含量配置了甲烷 90%、乙烷 5%、丙烷 1% 的定量等多个配比的测试气体。

对于气体吸收池的选择也是需要格外考虑的，根据公式(3)可知，吸收池的光程越长，吸光度数值越高，变化趋势越明显则越有利于测量。故此次测量采用的气体吸收池光程达到 50 cm。本次实验采用的是多功能温度加热控制仪，其主要作用是保证气体吸收池内的温度的恒定，避免温度波动对检测结果造成的影响。连接气体通路管线则选用聚酯型 PU 管，它有耐高压、高温且不吸附气体等优点。同时气路进出口需要配备减压阀，保证其压力和流量可记录。

2.2 实验数据

本文采用的数据由傅里叶红外光谱仪测得，测量时设置波段为中红外波段，扫描范围为：4000~400 cm⁻¹，分辨率为 8 cm⁻¹，扫描次数为 16 次。

由于实验过程严重存在各类噪声以及干扰气体，为保证光谱数据的准确性，本文分别采用 SG +db3 和 SG+sym6 两种预处理方法对光谱图进行处理。

2.3 天然气组分预处理

利用 SG (N=20) +db3 方法对含噪光谱去除后的效果如图 2 所示。从图 2 可以看出混合光谱图经过 3 层小波分解滤波以后，图谱变得平滑，且去除了高频噪声信号，相较于单纯的 SG 平滑以及中

值滤波来说，在实现了谱图平滑去噪的同时又很好保留了低频信号。

利用 SG (N=20) +sym6 方法，对含噪光谱去除后的效果如图 3 所示。使用的 sym6 小波函数处理效果明显好于 db3 小波函数，其原因主要在于相比于 db 小波，sym 小波具有更好的对称性型，能够减少重构时的位移，同时在使用 sym6 小波分解层数设置为 4 层，相对于 3 层的 db 小波来说去除更为彻底，基线保留也越多。

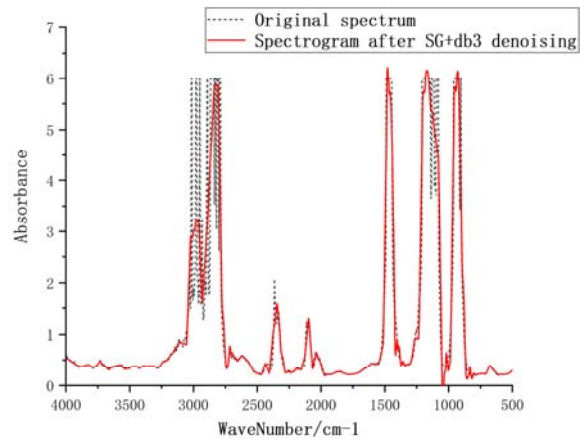


图 2 混合光谱经 SG+db3 小波去噪后的光谱图

Fig.2 The mixed spectrum is denoised by SG+db3 wavelet

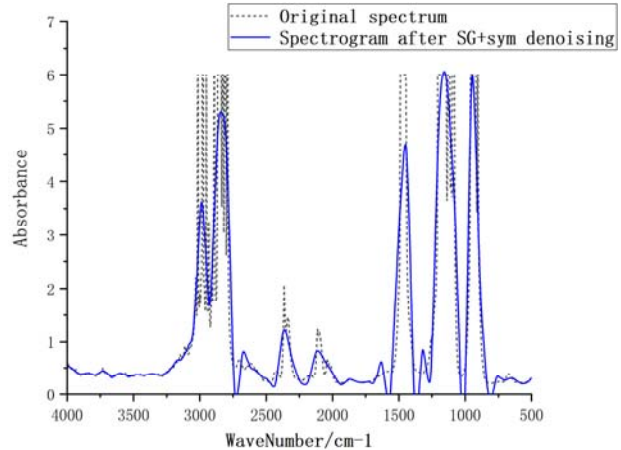


图 3 经 SG+sym 小波去噪后混合气体光谱图

Fig.3 Spectrum of mixed gas after denoising by SG+sym wavelet

3 预处理性能评价

对于光谱图预处理性能好坏的判断标准，主要步骤是利用各类预处理方法将原始光谱图理后采用高斯函数进行曲线拟合，利用函数拟合的各项性能评价指标进行评判。本次仍以混合气体光谱图中的波段 3300~2700 cm⁻¹ 重叠峰为例进行实验。

实验分别利用 SG 平滑处理过后的光谱图、中值滤波处理后的光谱图、以及 SG 平滑结合 db3 小

波去噪和 SG 平滑结合 sym6 小波去噪后的图谱分别进行高斯函数拟合, 查看其拟合情况。

图 4~图 7 中采用高斯拟合时峰位个数选择两个, 同时将拟合后的各项评价指标记录如表 1 所示。

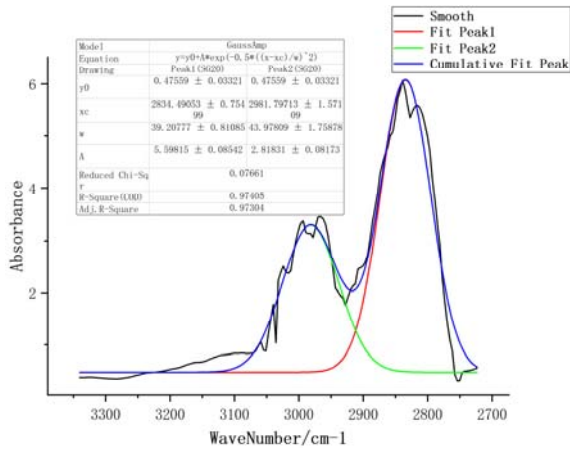


图 4 经 SG 平滑后的高斯拟合图
Fig.4 Gaussian fit after SG smoothing

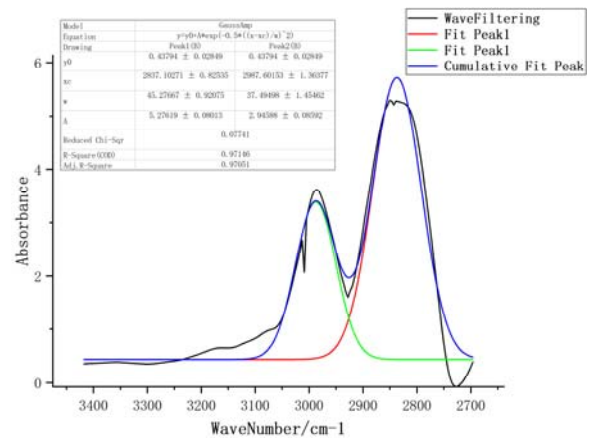


图 5 经滤波后的高斯拟合图
Fig.5 Gaussian fit after wave filtering

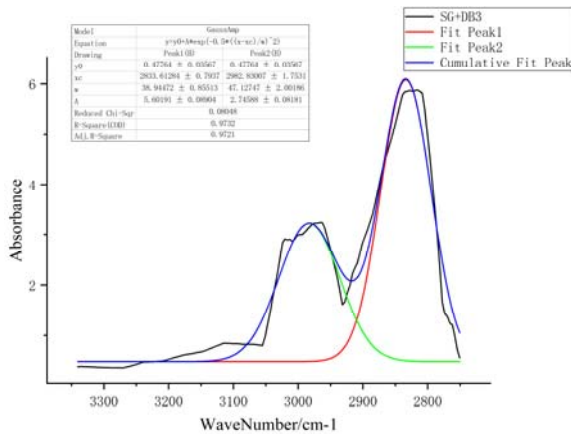


图 6 经 SG+db3 小波去噪后的高斯拟合图
Fig.6 Gaussian fit after SG+db3 wavelet denoising

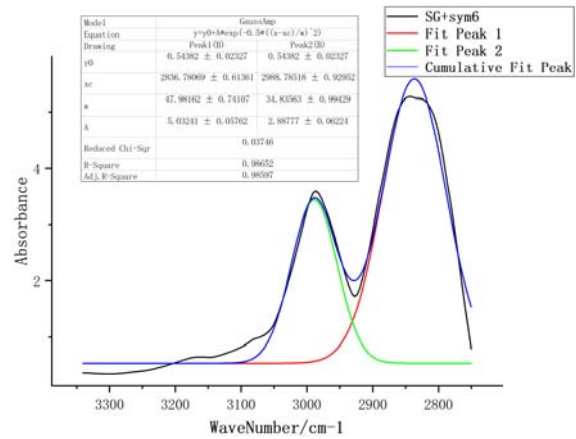


图 7 经 SG+sym6 小波去噪后的高斯拟合图
Fig.7 Gaussian fit after SG+sym6 wavelet denoising

表 1 SG 平滑+sym6 小波变换与传统方法性能指标对比

Table 1 Performance index comparison of SG smoothing + sym6 wavelet transform and traditional methods

Pretreatment method	Fitting performance		Various fitting parameters				
	R-Square	SSE	x_c	w	FWHM	A	Area
SG smoothing method ($N=20$)	0.97405	11.79833	2834.49055	39.20784	92.32725	5.59823	550.1827
			2981.79713	43.97818	103.56054	2.81837	310.68118
Median filter	0.97146	14.0104	2837.10274	45.2766707	106.618	5.27621	598.804
			2987.60154	37.49498	88.29445	2.94599	296.87252
SG($N=20$)+db3 3-layer decomposition wavelet	0.9732	11.82987	2833.61284	38.94472	91.7078	5.60191	546.85842
			2982.83007	47.12747	110.97672	2.74588	324.37366
SG($N=20$)+sym6 4-layer decomposition wavelet	0.98652	5.50694	2836.78074	47.98166	112.9881	5.03243	605.2586
			2988.78528	34.83569	82.03177	2.8878	252.1598

预处理性能好坏主要考察高斯函数拟合后的拟合优度 (R-Square) 以及残差平方和 (sum of squares due to error, SSE), 根据表 1 可知, 当采用 SG 平滑 ($N=20$) 结合 sym6 小波基函数, 在经过 4 层分解后软阈值去噪情况下, 其拟合优度数值最高为 0.98652, 残差平方和数值最低为 5.50694, 证明使用该方法后的函数分峰拟合效果最佳, 故本实验结果验证选择 SG 平滑与 sym6 小波软阈值去噪相结合的方法处理效果优于传统方法, 可将此方法应用于天然气组分检测领域。

4 结论

针对仪器测量光谱图的噪声干扰问题, 本文以天然气多组分检测为基础, 提出一种以 SG 平滑法结合 sym6 小波函数软阈值去噪法对光谱图进行预处理, 实验证明选择 SG 平滑法结合 sym6 小波软阈值去噪法对谱图进行去噪处理效果优于传统方法, 其拟合优度数值最高为 0.98652。采用该方法进行图谱分析有助于提高检测天然气各个组分含量的精度, 在新型天然气热值分析仪的研制上有广泛的应用前景, 能够促进石油天然气行业的发展。

参考文献:

- [1] JIAO Y, LI Z, CHEN X, et al. Preprocessing methods for near-infrared spectrum calibration[J]. *Journal of Chemometrics*, 2020, **34**(11): e3306.
- [2] YANG J, DU L, GONG W, et al. Analyzing the performance of the first-derivative fluorescence spectrum for estimating leaf nitrogen concentration[J]. *Optics Express*, 2019, **27**(4): 3978.
- [3] LI Y, HUANG Y, XIA J, et al. Quantitative analysis of honey adulteration by spectrum analysis combined with several high-level data fusion strategies[J]. *Vibrational Spectroscopy*, 2020, **108**: 103060.
- [4] 第五鹏瑶, 卞希慧, 王姿方, 等. 光谱预处理方法选择研究[J]. *光谱学与光谱分析*, 2019, **39**(9): 2800-2806.
DIWU Pengyao, BIAN Xi Hui, WANG Zifang, et al. Selection of spectral pretreatment methods[J]. *Spectroscopy and Spectral Analysis*, 2019, **39** (9): 2800-2806
- [5] Bastiaansen W, Klein S, Steegers mil heunissen R P, et al. VP24.02: towards a novel human embryonic brain atlas: fully automated preprocessing of prenatal ultrasound using artificial intelligence[J]. *Ultrasound in Obstetrics and Gynecology*, 2020, **56**(S1): 159-159.
- [6] HUANG X, HUANG C, ZHAI G, et al. Data processing method of multibeam bathymetry based on sparse weighted LS-SVM machine algorithm[J]. *IEEE Journal of Oceanic Engineering*, 2019, **45**(4): 1538-1551.
- [7] DING Y, ZHANG W, ZHAO X, et al. A hybrid random forest method fusing wavelet transform and variable importance for the quantitative analysis of K in potassic salt or using laser-induced breakdown spectroscopy[J]. *Journal of Analytical Atomic Spectrometry*, 2020, **35**(6): 1131-1138.
- [8] ZHANG J, WEN H, TANG L. Improved smoothing frequency shifting and filtering algorithm for harmonic analysis with systematic error compensation[J]. *IEEE Transactions on Industrial Electronics*, 2019, **66**(12): 9500-9509.
- [9] 杨帆, 王鹏, 张宁超, 等. 一种基于小波变换的改进滤波算法及其在光谱去噪方面的应用[J]. *国外电子测量技术*, 2020, **39**(8): 98-104.
YANG Fan, WANG Peng, ZHANG Ningchao, et al. An improved filtering algorithm based on wavelet transform and its application in spectral denoising[J]. *Foreign Electronic Measurement Technology*, 2020, **39** (8): 98-104.
- [10] Mahajan G R, Das B, Gaikwad B, et al. Monitoring properties of the salt-affected soils by multivariate analysis of the visible and near-infrared hyperspectral data[J]. *Catena*, 2021, **198**: 105041.
- [11] 侯培国, 李宁, 常江, 等. SG 平滑和 IBPLS 联合优化水中油分析方法的研究[J]. *光谱学与光谱分析*, 2015, **35**(6): 1529-1533.
HOU Peiguo, LI Ning, CHANG Jiang, et al. Research on SG smoothing and IBPLS joint optimization of oil-in-water analysis method[J]. *Spectroscopy and Spectral Analysis*, 2015, **35**(6): 1529-1533.
- [12] Zikiou N, Lahdir M, Helbert D. Hyperspectral image classification using graph-based wavelet transform[J]. *International Journal of Remote Sensing*, 2020, **41**(7): 2624-2643.
- [13] Bunaciu A A, Aboul-Enein H Y. Adulterated drug analysis using FTIR spectroscopy[J]. *Applied Spectroscopy Reviews*, 2021, **56**(5): 423-437.
- [14] LIU F, ZHANG Y, Yildirim T, et al. Signal denoising optimization based on a Hilbert-Huang transform-triple adaptable wavelet packet transform algorithm[J]. *EPL (Europhysics Letters)*, 2019, **124**(5): 54002.
- [15] ZHANG M, LU C, LIU C. Improved double-threshold denoising method based on the wavelet transform[J]. *OSA Continuum*, 2019, **2**(8): 2328-2342.