

## 激光与光电子学进展

光谱处理技术对提升光纤诱导荧光光谱检测  
低浓度 NaCl 气溶胶能力的实验研究职辰好<sup>1</sup>, 高亚敏<sup>1</sup>, 向钰言<sup>1</sup>, 张智<sup>2</sup>, 尚滨鹏<sup>2</sup>, 郭杰伟<sup>2</sup>, 谢茂强<sup>1\*</sup>, 刘明铭<sup>1\*\*</sup>, 张楠<sup>2</sup>, 孙陆<sup>2</sup>, 刘伟伟<sup>2</sup><sup>1</sup>南开大学软件学院, 天津 300350;<sup>2</sup>南开大学现代光学研究所, 天津 300350

**摘要** 大气污染对人类的生产生活有极大影响,气溶胶作为污染物的重要部分,不容忽视。提高对大气气溶胶浓度检测的精确性,尤其是低浓度气溶胶,具有十分重要的意义。本文基于光纤诱导荧光光谱技术,对 NaCl 气溶胶数据进行预处理,并结合偏最小二乘法建立预测模型,探索不同预处理方法对模型检测精度的影响。讨论如何科学合理地选择预处理方法,按照预处理方法效果分为散射校正、平滑去噪、基线校正 3 个方面,并提出波峰显著度算法。通过无预处理、单一预处理以及组合预处理进行最优预处理方法的选择,并分析其建模精度的影响。实验结果表明,应用多个预处理方法的组合,与无预处理相比,均方根误差降低至 0.03,预测相对误差减少 60%;与直接观察光谱信号选择预处理方法相比,根据光谱信噪比的提升及预测组分的建模效果可以更为准确地选择最佳预处理方法。该研究为开展低浓度大气污染物的分析研究提供了一定的参考。

**关键词** 大气光学与海洋光学; 光纤诱导荧光光谱; 光谱处理; 低浓度氯化钠气溶胶检测

中图分类号 O433.4

文献标志码 A

DOI: 10.3788/LOP212962

Experimental Study of Spectral Processing Technology for Improving  
Detection Ability of Low-Concentration NaCl Aerosol by  
Filament-Induced Fluorescence SpectrumZhi Chenhao<sup>1</sup>, Gao Yamin<sup>1</sup>, Xiang Yuyan<sup>1</sup>, Zhang Zhi<sup>2</sup>, Shang Binpeng<sup>2</sup>, Guo Jiewei<sup>2</sup>,  
Xie Maoqiang<sup>1\*</sup>, Liu Mingming<sup>1\*\*</sup>, Zhang Nan<sup>2</sup>, Sun Lu<sup>2</sup>, Liu Weiwei<sup>2</sup><sup>1</sup>College of Software, Nankai University, Tianjin 300350, China;<sup>2</sup>Institute of Modern Optics, Nankai University, Tianjin 300350, China

**Abstract** Air pollution greatly impacts the productivity and life of people. Aerosol is an important pollutant and cannot be ignored. It is significant to improve the detection accuracy of atmospheric aerosol concentrations, especially low-concentration aerosols. In this study, filament-induced fluorescence spectrum (FIFS) of NaCl aerosol is preprocessed and combined with partial least squares (PLSR) to establish a prediction model and explore the impact of different preprocessing methods on the detection accuracy of the model. To choose the preprocessing method, this study divides the preprocessing methods into three aspects according to their effects: scattering correction, smoothing and denoising, and baseline correction, and the significance of peak algorithm is proposed. The optimal preprocessing method is selected after comparing no preprocessing, single preprocessing, and combined preprocessing and analyzing the influence of different preprocessing methods on the accuracy of FIFS spectral prediction model. The experimental results show that the combined preprocessing of multiple methods reduces the root mean square error to 0.03 compared with no preprocessing, and the relative prediction error is reduced by 60%. Compared with the direct observation of spectral signal selection preprocessing method, the best preprocessing method can be selected more accurately according to the improvement of spectral signal-to-noise ratio and the modeling effect of predicted components. The present study provides a reference for the analysis and research of low-concentration air pollutants.

收稿日期: 2021-10-15; 修回日期: 2021-11-18; 录用日期: 2021-11-29; 网络首发日期: 2021-12-10

基金项目: 国家重点研发计划(2018YFB0504400)

通信作者: \*xiemq@nankai.edu.cn; \*\*liumingming@nankai.edu.cn

**Key words** atmospheric optics and ocean optics; filament-induced fluorescence spectrum; spectral processing; low-concentration NaCl aerosol detection

## 1 引言

随着中国工业化的不断发展,煤炭燃烧、工业排放、机动车尾气等因素导致的大气污染给人类的健康及生态环境带来了极大的危害<sup>[1]</sup>。其中,大气气溶胶污染物是导致雾霾天气的重要因素,其成分复杂、形态多样。因此,对大气气溶胶污染物,尤其是低浓度大气气溶胶污染物的精确检测至关重要。

为了检测大气气溶胶成分及质量浓度,很多研究者对其光谱进行分析,并且对于不同的光谱采用了不同的处理手段。郭旺林等<sup>[2]</sup>通过建立重金属元素定标曲线,实现了对大气气溶胶中的重金属元素含量的检测;且进一步采用快速傅里叶变换、5点SG平滑法和小波阈值降噪法对气溶胶激光诱导击穿光谱(LIBS)数据进行降噪实验<sup>[3]</sup>,通过对比3种降噪算法的信噪比及平滑度值,表明小波阈值降噪法效果最优;第五鹏瑶等<sup>[4]</sup>对9种物质的光谱测试了不同的预处理方法组合,并使用偏小二乘法(PLSR)建立定量模型,实验结果表明,根据光谱信号特点选择的预处理组合很难使定量模型的预测效果达到最佳;张朱珊莹等<sup>[5]</sup>将预处理方法分为4类,选择每类预处理方法中的最优预处理方法进行组合,通过PLSR建模效果来选出最优预处理方法。但不同的数据集其复杂性存在较大差异,且对于大气气溶胶污染物的相关研究较少,尤其是低浓度大气气溶胶污染物。

传统的LIBS会产生较强的连续背景噪声辐射,同时光谱仪在信号处理中带来光/电信号以及数/模信号转换的附加噪声,会影响分析元素时的光谱强度<sup>[6]</sup>,难以获取包含复杂成分和相似化学结构的谱线。在光丝飞秒激光技术下<sup>[7-9]</sup>,物质被高度激发和解离,形成分子碎片的“指纹”荧光,可用于识别大气中的各种物质,包括气体、气溶胶等<sup>[10]</sup>。使用高强度激光光丝诱导

荧光光谱(FIFS)可以实现大气污染远程、多组分、快速测量,是未来实现光丝激光雷达的关键技术<sup>[11-13]</sup>。本文以此为背景,在实验室中使用低浓度氯化钠气溶胶和云室模拟大气中的气溶胶,使用高功率超快激光激发出光丝并获取其FIFS数据。为降低各类噪声对信号的影响,提升多组分检测和浓度预测的精度。本文面向低浓度NaCl气溶胶的FIFS数据处理方法展开实验研究,并对比了中值滤波(MF)、SG(Savitzky-Golay)平滑、快速傅里叶变换(FFT)滤波、多元散射校正(MSC)、标准正态变量变换(SNV)和一阶导(FD)、二阶导(SD)、离散小波变换(DWT)、波峰显著度(SOP)等预处理方法及其组合,获得光谱数据的预处理结果,并使用PLSR<sup>[14]</sup>建立定量模型,进行浓度预测。实验表明,合适的预处理方法能在很大程度上提高低浓度NaCl气溶胶的信噪比及模型预测的准确性。

## 2 实验部分

### 2.1 光谱采集与实验设备

光丝激发与光谱采集实验装置系统如图1所示,飞秒脉冲经过由凹透镜L1( $f=-10$  cm)、凸透镜L2( $f=50$  cm)和800 nm介质膜反射镜M1组成的望远镜聚焦系统。通过控制L1与L2之间的相对距离,可以在实验室允许范围内( $<15$  m)精确地控制飞秒激光的成丝位置。事实上该系统装置巧妙地将发射系统与荧光信号收集系统相结合,用于会聚成丝的凸透镜L2同时又可以对成丝处产生的光丝诱导背向荧光信号进行收集。信号采集装置主要由光纤束(纤芯直径0.3 mm)、光谱仪(Zolix 300 $\lambda$ )、Istar-sCMOS相机(Istar-sCMOS-18H-13),以及设备对应的计算机处理软件组成,对光丝电离云室中NaCl气溶胶产生的背向荧光信号进行光谱采集。

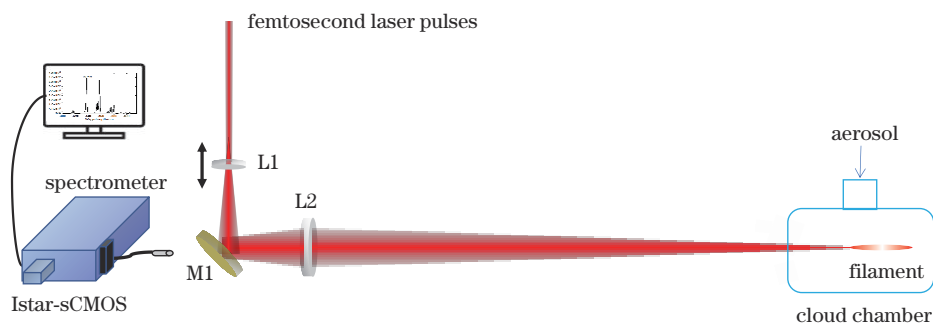


图1 FIFS采集光谱装置图

Fig. 1 Schematic diagram of filament-induced fluorescence spectrum (FIFS) collection spectrum device

实验中激光脉冲的波长为800 nm,能量设定为6 mJ,脉冲宽度为50 fs,工作频率为500 Hz。使用的光谱仪波长范围为180~920 nm,光谱分辨率为0.1 nm。

图2为气溶胶发生器实物图,是由北京慧荣合公司生产的6孔型号的气溶胶发生器,平均粒径为4  $\mu$ m。云室由透明的亚克力板材料构成,如图3所示。



图 2 气溶胶发生器实物图

Fig. 2 Photo of aerosol generator

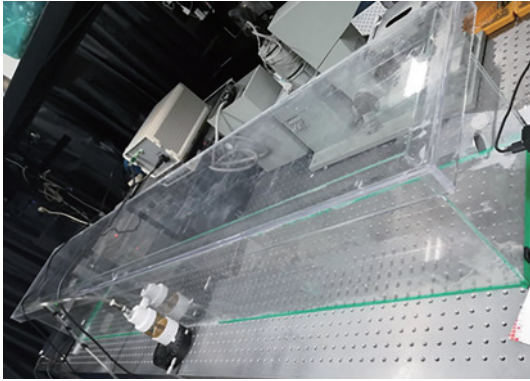


图 3 云室实物图

Fig. 3 Photo of cloud chamber

## 2.2 实验样品与数据

实验中的 NaCl 气溶胶由纯度为 99% 的 NaCl 固体与蒸馏水,按照一定的质量分数配比所配置,配置的 NaCl 气溶胶质量分数分别为  $150 \times 10^{-6}$ 、 $300 \times 10^{-6}$ 、 $1500 \times 10^{-6}$ 、 $3000 \times 10^{-6}$ ,对每种质量分数的 NaCl 测得 10 条光谱数据,共 40 条样本光谱。同时测得 10 条无 NaCl 溶质的光谱作为背景光谱。

在采集到的样品中,由于环境温度湿度、仪器误差、样品处理等原因,不可避免地会存在异常样本,而异常样本对模型的预测能力有很大的影响,必须从其样品集中剔除<sup>[15]</sup>。本文采用马氏距离剔除异常样本,然后采用基于联合  $x$ - $y$  距离的样本集划分 (SPXY)<sup>[16]</sup> 算法对样品集进行划分。SPXY 算法通过计算样品间的欧氏距离,来保证样品的空间分布均匀,同时将光谱与气溶胶的含量考虑在内,能有效地覆盖多维向量空间,提高模型的稳定性。实验训练集与预测集的比例为 3:1。

## 2.3 光谱预处理方法

在光谱数据采集的过程中,由于受到仪器、环境、测量条件等因素的影响,光谱中除有效信息外,还存在大量的噪声和冗余信息。而噪声的干扰,往往会造成谱峰偏移、基线漂移、信号毛刺等问题,对光谱分析有很大影响。为了获得更准确而有效的数据,需对光谱

数据进行预处理。目前光谱预处理的方法存在很多种,不同光谱预处理方式建立的模型也存在着不同的差异<sup>[17]</sup>。每种预处理方法都有特定的作用与效果,简单地将预处理的方法叠加起来,并不一定能取得预期的效果。此外,还需考虑算法叠加时的相互影响。因此,找出效果最佳的处理方法,对低浓度大气气溶胶光谱的分析至关重要。根据噪声的不同,本文主要从平滑去噪<sup>[18-19]</sup>、散射校正<sup>[20-21]</sup>、基线校正<sup>[22-23]</sup>3 个方面,比较不同处理方法对光谱处理的影响<sup>[24-25]</sup>。

本文采用了 3 种平滑去噪方式:MF、SG 和 FFT;两种常见的散射校正算法:MSC 和 SNV;3 种基线校正方法:FD、SD 和 DWT;以及本文提出的 SOP 算法。表 1 所示为本文预处理方法组合。

表 1 光谱的预处理方法组合

Table 1 Combinations of spectral preprocessing methods

No.	Preprocessing method
1	Original
2	SG
3	FFT
4	MSC
5	SD
6	DWT
7	FFT-SOP
8	SG-FD
9	SG-FFT-SD
10	MF-SG-FFT
11	MF-SG-FFT-DWT
12	MF-SG-FFT-SOP
13	MSC-MF-SG-FFT
14	MSC-FFT-MF-SG
15	MSC-MF-SG-FFT-DWT
16	MSC-MF-SG-FFT-SOP
17	SNV-SG-FFT-FD
18	SNV-MF-SG-FFT-DWT

## 2.4 评价指标

### 2.4.1 去噪性能指标

去噪性能采用信噪比 (SNR)<sup>[26]</sup> 与均方根误差 (RMSE)<sup>[26]</sup> 作为评价指标。SNR 越大,则表明噪声处理得越好。RMSE 越小,则认为重构的信号越接近。其分别为

$$P_{\text{SNR}} = 10 \times \lg \left\{ \frac{\sum_{i=1}^L f^2(i)}{\sum_{i=1}^L [s(i) - f(i)]^2} \right\}, \quad (1)$$

$$P_{\text{RMSE}} = \sqrt{\frac{\sum [s(i) - f(i)]^2}{L}}, \quad (2)$$

式中: $L$  为光谱样本维度; $s(i)$  和  $f(i)$  分别为原始光谱信号和不同方法处理后的光谱信号数据。

### 2.4.2 预测模型性能指标

浓度预测采用决定系数  $R^2$ 、预测均方根误差 (RMSEP) 以及相对分析误差 (RPD) 来衡量模型性能<sup>[27]</sup>。

决定系数  $R^2$  用于评价模型对样本值的拟合程度, 其值越接近 1 说明模型的拟合程度越好, 其计算公式为

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (\hat{y}_i - \bar{y})^2}, \quad (3)$$

式中:  $y_i$  和  $\hat{y}_i$  分别为真实浓度与预测浓度值;  $\bar{y}$  为真实浓度值的平均值。

RMSEP 是回归分析中最常用的分析指标, 代表着预测的误差大小, 其值越低则表明模型预测效果越好, 其公式为

$$P_{\text{RMSEP}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (4)$$

式中,  $n$  为样本个数。

RPD 是基于决定系数的一种分析指标, RPD 小于 1.4 时认为模型不可靠, RPD 大于 1.4 时认为模型已经比较可靠, 当其大于 2 时, 则认为模型已经具有较高的可靠性, 其计算公式为

$$P_{\text{RPD}} = \frac{1}{\sqrt{1 - R^2}}. \quad (5)$$

## 3 结果与讨论

用 2.3 节中表 1 列出的预处理方法对 2.2 节中的数据进行处理, 然后对预处理后的数据进行 PLSR 定量分析。将实验分为两部分: 首先用无预处理、单一预处理及组合预处理的方法对数据进行处理, 并对比不同预处理方法处理后的 SNR 及 RMSE, 选出预处理后 SNR 高且 RMSE 小的处理方法; 之后, 使用 PLSR 对不同预处理方法处理后的数据建立预测模型, 并对比不同预处理方法的  $R^2$ 、RPD 及 RMSEP, 选出预测效果好的处理方法, 结合预处理及预测模型结果, 选择最优预处理方法。

### 3.1 光谱的预处理

#### 3.1.1 光谱特征分析

FIFS 信号在采集时会受到激光器激光能量波动、光谱仪分辨率差异、外部环境差异以及样品不均匀等因素影响, 为了尽可能地消除或减弱这些扰动噪声, 使得特征突出, 更好地对光谱特别是低浓度光谱信号进行分析, 需对 FIFS 数据进行光谱预处理。图 4 给出了 4 种质量分数的 NaCl 气溶胶 FIFS 实测数据。从中可以看到, NaCl 特征不突出, 噪声影响严重, 特征信号峰淹没在噪声中, 不利于光谱信号的分析。

质量分数分别为  $150 \times 10^{-6}$  和  $300 \times 10^{-6}$  的 NaCl 气溶胶在 8 种预处理方法处理后的效果, 如图 5 所示。

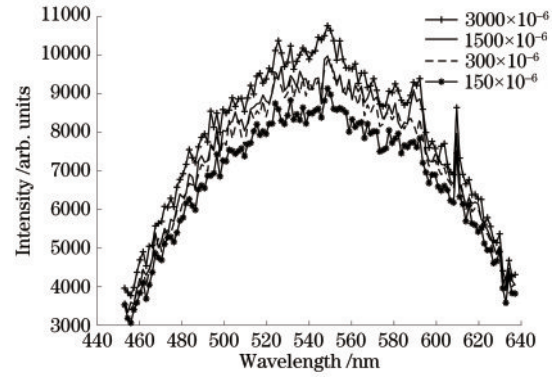


图 4 NaCl 气溶胶的 FIFS

Fig. 4 FIFS of NaCl aerosol

图 5(c) 和 5(d) 是分别经过 SNV 和 MSC 处理的效果, 可以发现经过散射校正后, 保留了原始光谱的特征峰并使得光谱重叠度增加, 减弱散射对原始光谱的影响。在原始光谱信号中, 存在脉冲噪声的干扰, 图 5(e) 为进行 MF 处理后的效果, 其将每一光谱点的强度值设置为该点某邻域窗口内的所有特征点强度值的中值, 使得周围的特征信号强度值接近真实值, 从而既保留原始光谱的信息又去除了脉冲噪声的影响。SG 卷积平滑通过设定窗口内的数据进行加权滤波, 消除信号抖动噪声, 且在滤波平滑的同时, 有效地保留原有信号的特征波动信息。如图 5(a) 所示, 经过 SG 平滑处理, 可以看到平滑后的光谱的随机噪声明显减弱, 但许多高频随机噪声、基线偏移等噪声信息依然严重影响着光谱。图 5(b) 为经过 FFT 滤波处理后的效果, 在频域滤除信号的高频噪声, 逆变换后去除光谱的高频噪声, 提高了光谱的 SNR, 这暗示着, 其可以降低内外部环境因子的干扰, 将原本淹没在原光谱数据中的一些微弱的特征光谱信息挖掘出来。

导数光谱通过对原始光谱求导, 从光谱中分离出特征峰, 精确地定位特征峰的位置, 提高光谱的检测灵敏度, 且有助于消除或降低背景干扰。对于斜线背景, 经 FD 处理后, 其 FD 为常数, 从而可消除线性背景。对于 SD 则可扣除曲线背景。从图 5(f) 和 5(g) 可以看出, 经 FD 和 SD 处理后的光谱, 消除了基线和背景的干扰, 但同时增大了散射和光谱差异的影响。光谱信号经小波分解后, 基线干扰集中在低频段的小波系数中, 而光谱信号分布在较高频段的小波系数上, 用低频段小波系数置零的办法可以实现基线与信号的分离, 采用迭代 DWT 去除基线偏移的效果如图 5(h) 所示, 与原始信号相比, 其较好地去除了低频背景信号, 更好地提取特征峰。

经过预处理后消除了一部分噪声的干扰, 但依旧存在伪峰。为了消除伪峰, 对于经过预处理后的光谱进行寻峰处理。对于物质, 其采集的光谱必含有代表其存在的特征峰, 特征峰应在每次测量时均出现。在同浓度的一组数据中, 这里将波峰出现的频率称为

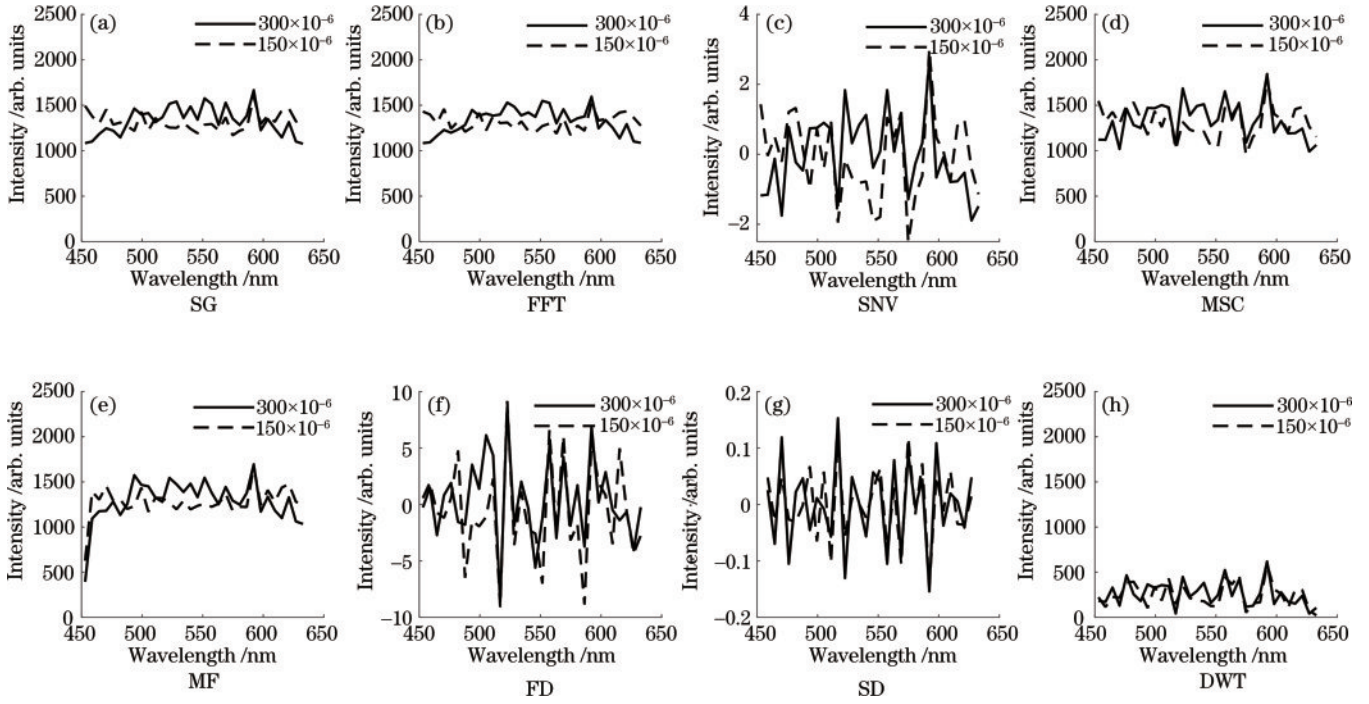


图 5 单一预处理方法光谱图。(a)SG;(b)FFT;(c)SNV;(d)MSC;(e)MF;(f)FD;(g)SD;(h)DWT

Fig. 5 Spectra of single preprocessing methods. (a) SG; (b) FFT; (c) SNV; (d) MSC; (e) MF; (f) FD; (g) SD; (h) DWT

SOP, 计算公式为

$$S_{\text{sop}} = \frac{m}{N}, \quad (6)$$

式中： $m$  为某一波峰位置出现的次数； $N$  为样本总个数。

当波峰出现的频率大于某一阈值时，认为其为特征峰，这里将这一阈值称为 SOP 阈值。对于同一组数据来说，当光谱波峰出现次数的比例超过 SOP 阈值时，认为其为特征谱峰，并且特征峰处应是固定存在且变化不大的光谱数据，因此对寻峰结果进行 SOP 筛选，认为 SOP 高的波峰为物质的特征峰。以进行“MSC-MF-SG-FFT”处理、质量分数为  $150 \times 10^{-6}$  的 NaCl 气溶胶光谱为例，列出其前 15 个波峰 SOP 如表 2 所示。

从表 2 可以看出，590 nm 峰为 SOP 最高的特征峰，本实验中设 SOP 的阈值为 0.75。图 6 为 SOP 处理前的光谱，可以看出存在较多伪峰，经 SOP 处理后，再将筛选出来的特征峰进行高斯拟合，得到最终处理效果如图 7 所示，

经处理后，使低浓度 NaCl 气溶胶的 FIFS 的特征峰更显著，较大程度上减少了噪声的影响，提高了 SNR。

### 3.1.2 参数优化

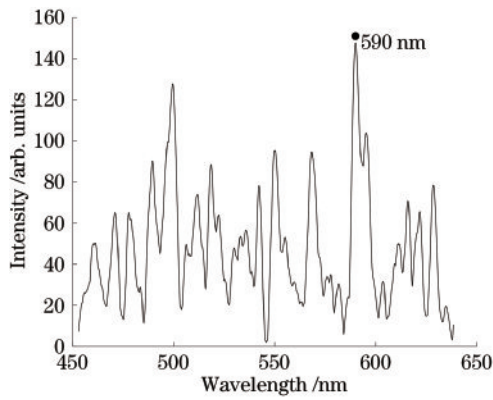
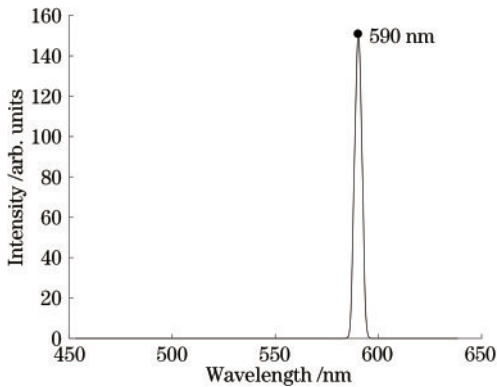
本实验中首先通过累加平均技术来降低光谱信号的波动影响，之后采用 MF、SG、FFT、MSC、SNV、FD、SD、DWT 以及其组合的方法进行预处理，除 MSC、SNV 之外，其余预处理方法需进行参数优化，

表 2 波峰及 SOP

Table 2 Peak positions and significance of peak (SOP)

No.	Peak position /nm	SOP
1	590	0.8
2	499	0.7
3	542	0.7
4	549	0.7
5	518	0.6
6	480	0.5
7	492	0.5
8	511	0.5
9	551	0.5
10	569	0.5
11	496	0.3
12	594	0.3
13	616	0.3
14	620	0.3
15	628	0.3

MF、SG、FFT、FD、SD 窗口的优化方法是窗口数在 3~89 进行变化，间隔为 2，分别计算各自的 RMSEP 值，RMSEP 最小值对应的窗口即为最佳窗口。DWT 采用了 Haar、db2-db10、coif1-coif5、sym2-sym8 等 22 个小波函数；分解尺度在 1~15 进行变化，分别计算不同小波函数和分解尺度对应的 RMSEP 值。优化后的窗口参数分别是：MF 窗口为 20，SG 平滑的平滑点为 3，窗口大小为 61，FFT 窗口为 35，FD 和 SD 窗口分别为

图 6  $150 \times 10^{-6}$  NaCl 在 SOP 处理前的光谱Fig. 6 Spectra of  $150 \times 10^{-6}$  NaCl before processing of SOP图 7  $150 \times 10^{-6}$  NaCl 在 SOP 处理后的光谱Fig. 7 Spectra of  $150 \times 10^{-6}$  NaCl after processing of SOP

60 和 80, DWT 采用 db7 为小波函数, 分解尺度为 8。

### 3.2 去噪性能效果分析

合理的光谱预处理方法可以去除光谱中的无用信息, 保留特征信息, 提高模型预测能力。本实验对低浓度 NaCl 气溶胶 FIFS 分别采用表 3 所示的单个预处理方法以及不同组合预处理方法进行处理。

从表 3 可见, 对于原始光谱, 谱线稳定性较差, 且特征不明显, 不易识别。对 NaCl 光谱进行预处理后, “MSC-MF-SG-FFT-SOP” 方法得到的 SNR 最大, 表现出最强的去噪能力。且 “MSC-MF-SG-FFT-SOP” 方法的 RMSE 值最小, 为 0.03, 较无预处理时, 是其 1/5, 表明其能较好地保持光谱数据的相似性。其次, 在没有采用 SOP 时, “MSC-MF-SG-FFT” 方法的 SNR 及 RMSE 均差于进一步采用 SOP 方法处理的方案, 说明 SOP 对噪声去除效果具有提升作用。综上分析, “MSC-MF-SG-FFT-SOP” 方法在噪声去除和光谱细节保持上具有优越性, 效果最佳, 但是对于预处理的效果需要进一步的 PLSR 定量分析结果来验证。

### 3.3 浓度预测结果分析

对不同的预处理方法, 分别建立模型, 实现对 NaCl 气溶胶的浓度预测。基于 PLSR 对原始谱线和经过预处理后的数据建模, 得到的 NaCl 气溶胶含量的拟合结果散点图如图 8 所示。不同预处理方法对所建

表 3 组合预处理方法

Table 3 Combinations of preprocessing methods

Preprocessing method	SNR	RMSE
Original	25.369	0.163
SG	37.002	0.127
FFT	38.698	0.121
MSC	26.632	0.166
SD	-11.879	0.631
DWT	2.432	0.139
FFT-SOP	42.427	0.031
SG-FD	-7.410	0.408
SG-FFT-SD	-6.677	0.453
MF-SG-FFT	38.920	0.111
MF-SG-FFT-DWT	6.946	0.169
MF-SG-FFT-SOP	41.873	0.030
MSC-MF-SG-FFT	44.180	0.110
MSC-FFT-MF-SG	44.351	0.106
MSC-MF-SG-FFT-DWT	6.889	0.172
MSC-MF-SG-FFT-SOP	46.630	0.030
SNV-SG-FFT-FD	-6.691	0.451
SNV-MF-SG-FFT-DWT	5.673	0.167

立的 PLSR 定标模型结果有所差异, 其中 “MF-SG-FFT” 的拟合效果偏差较大, 相关程度较低。经组合 “MF-SG-FFT-SOP”、“MF-SG-FFT-DWT”、“MSC-MF-SG-FFT-SOP” 处理后的建模拟合效果较好。

不同预处理的模型预测结果如表 4 所示, 可以看

表 4 不同预处理方法 PLSR 建模结果

Table 4 Prediction results of PLSR models under different preprocessing methods

Pretreatment method	$R^2$	RPD	RMSEP
Original	0.540	1.474	0.259
SG	0.524	1.450	0.263
FFT	0.492	1.403	0.272
MSC	0.896	3.105	0.123
SD	0.854	2.622	0.145
DWT	0.947	4.353	0.088
FFT-SOP	0.879	2.873	0.133
SG-FD	0.843	2.520	0.151
SG-FFT-SD	0.794	2.205	0.173
MF-SG-FFT	0.466	1.369	0.279
MF-SG-FFT-DWT	0.923	3.602	0.106
MF-SG-FFT-SOP	0.936	3.963	0.096
MSC-MF-SG-FFT	0.847	2.556	0.149
MSC-FFT-MF-SG	0.847	2.558	0.149
MSC-MF-SG-FFT-DWT	0.901	3.179	0.120
MSC-MF-SG-FFT-SOP	0.955	4.712	0.081
SNV-SG-FFT-FD	0.649	1.689	0.226
SNV-MF-SG-FFT-DWT	0.821	2.363	0.161

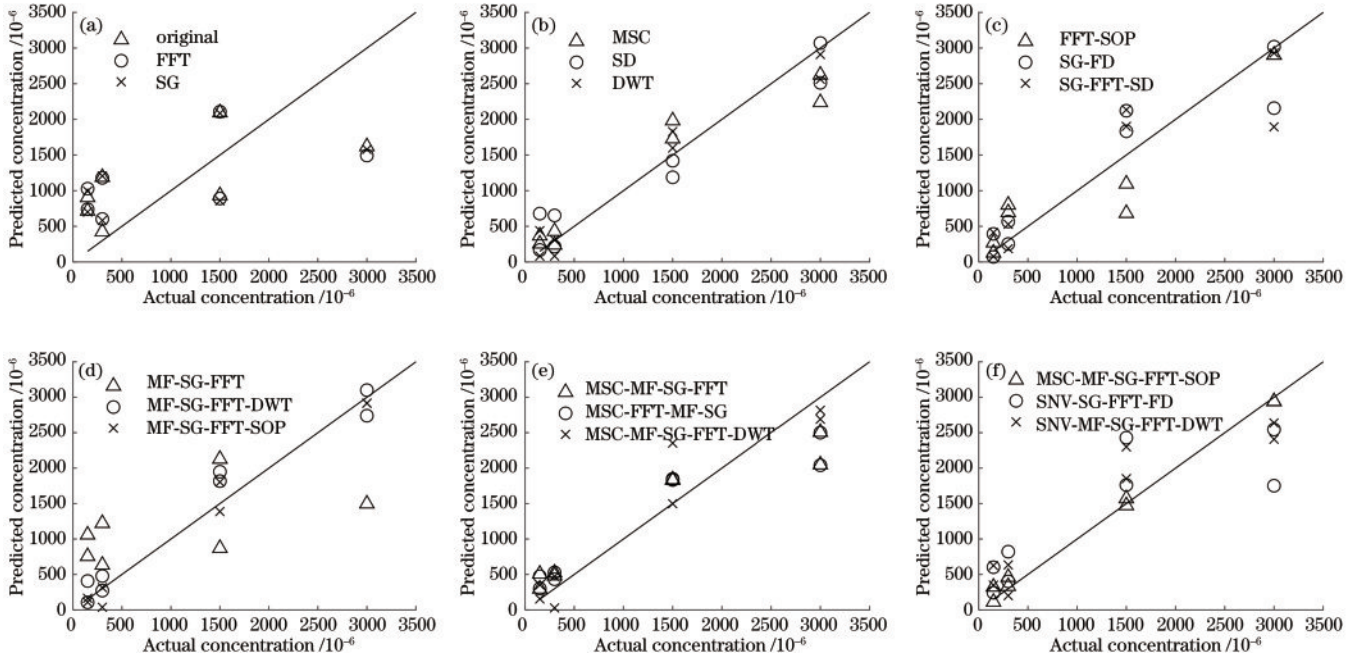


图 8 不同预处理方法 PLSR 预测值拟合散点图。(a) Original、FFT、SG; (b) MSC、SD、DWT; (c) FFT-SOP、SG-FD、SG-FFT-SD; (d) MF-SG-FFT、MF-SG-FFT-DWT、MF-SG-FFT-SOP; (e) MSC-MF-SG-FFT、MSC-FFT-MF-SG、MSC-MF-SG-FFT-DWT; (f) MSC-MF-SG-FFT-SOP、SNV-SG-FFT-FD、SNV-MF-SG-FFT-DWT

Fig. 8 Fitting scatter plots of PLSR prediction value for different preprocessing methods. (a) Original, FFT, SG; (b) MSC, SD, DWT; (c) FFT-SOP, SG-FD, SG-FFT-SD; (d) MF-SG-FFT, MF-SG-FFT-DWT, MF-SG-FFT-SOP; (e) MSC-MF-SG-FFT, MSC-FFT-MF-SG, MSC-MF-SG-FFT-DWT; (f) MSC-MF-SG-FFT-SOP, SNV-SG-FFT-FD, SNV-MF-SG-FFT-DWT

出,不同的预处理组合方法得到的结果不同。当使用“SNV-MF-SG-FFT-DWT”组合方法时, $R^2$ 为 0.821,RPD 达到 2.363,预测误差为 0.161,而当使用“MSC-MF-SG-FFT-DWT”时, $R^2$ 和 RPD 分别为 0.901 和 3.179,预测误差降至 0.120。因此可以看出,在使用多种预处理方法时,不同的预处理方法组合对预测结果会产生不同的影响。

表 4 结果显示,对于原始 NaCl 气溶胶光谱,谱线相对强度小,特征不突出,噪声影响严重,特征信号峰淹没在噪声中。其  $R^2$  仅为 0.540, RMSEP 为 0.259,RPD 为 1.474。预测结果最好的预处理方法组合是“MSC-MF-SG-FFT-SOP”,RPD 达到 4.712,PLSR 算法的定标曲线的  $R^2$  提高至 0.955, RMSEP 为 0.081,误差相较于原始数据方法降低了 60%。综合噪声去除结果分析,“MSC-MF-SG-FFT-SOP”方法在噪声去除及 PLSR 建模两方面的性能均表现较好,效果最佳。由此可见,采用组合预处理后的光谱的浓度预测效果最佳,光谱中提取的有效信息更多,可以更大程度地去除噪声,保留特征信息,以提高定量分析的精度。

## 4 结 论

采用光谱仪获取 NaCl 气溶胶光谱,并对比分析了 MF、SG、FFT、MSC、SNV、FD、SD、DWT、SOP 以及

其组合预处理方法对光谱噪声去除及 PLSR 建模的影响。以 40 个 NaCl 气溶胶样品为研究对象,从噪声处理结果可以看出,“MSC-MF-SG-FFT-SOP”组合处理后的 NaCl 气溶胶的特征峰突出明显,与原始光谱相比, RMSE 由 0.163 降至 0.030, SNR 由 25.369 提升至 46.630,在特征突出及性能保持上有较大的提升,有利于进一步的数据分析。使用 SPXY 划分数据集,然后对各种预处理方法分别建立浓度预测模型。浓度预测结果表明,当采用“MSC-MF-SG-FFT-SOP”组合时,预测结果表现最好, $R^2$ 、RPD 和 RMSEP 分别达到 0.955、4.712 和 0.081,准确程度得到较大的提升,能实现对低浓度 NaCl 气溶胶的预测,具有一定的实际意义。故对本实验的 NaCl 气溶胶光谱,“MSC-MF-SG-FFT-SOP”的预处理组合方法为其最优预处理方法。由此可见,不同预处理方法,使用顺序及组合的不同,会获得不同的预测结果,在对光谱数据进行预处理时,选择恰当的方法对去噪性能及定量分析的准确度具有较大的影响。

## 参 考 文 献

- [1] 朱首正,卜令兵,刘继桥,等.机载高光谱分辨率激光雷达探测大气气溶胶光学特性及污染研究[J].中国激光,2021,48(17):1710003.  
Zhu S Z, Bu L B, Liu J Q, et al. Study on airborne high spectral resolution lidar detecting optical properties and

- pollution of atmospheric aerosol[J]. Chinese Journal of Lasers, 2021, 48(17): 1710003.
- [2] 郭旺林, 邱荣, 王昶洋, 等. 激光诱导击穿光谱技术用于大气气溶胶中 Cr 元素的检测[J]. 量子电子学报, 2020, 37(6): 745-751.  
Guo W L, Qiu R, Wang C F, et al. Detection of chromium in atmospheric aerosol by laser induced breakdown spectroscopy[J]. Chinese Journal of Quantum Electronics, 2020, 37(6): 745-751.
- [3] 郭旺林. 基于气溶胶成分分析的 LIBS 系统集成及应用[D]. 绵阳: 西南科技大学, 2020.  
Guo W L. Integration and application of LIBS system based on aerosol composition analysis[D]. Mianyang: Southwest University of Science and Technology, 2020.
- [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 P Y, Bian X H, Wang Z F, et al. Study on the selection of spectral preprocessing methods[J]. Spectroscopy and Spectral Analysis, 2019, 39(9): 2800-2806.
- [5] 张朱珊莹, 顾瀚文, 谢凯文, 等. 基于近红外光谱的预处理及组合方法[J]. 激光与光电子学进展, 2021, 58(16): 1617001.  
Zhang Z S Y, Gu H W, Xie K W, et al. Pretreatment and combined method based on near infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1617001.
- [6] 陈添兵, 刘木华, 黄林, 等. 不同光谱预处理对激光诱导击穿光谱检测猪肉中铅含量的影响[J]. 分析化学, 2016, 44(7): 1029-1034.  
Chen T B, Liu M H, Huang L, et al. Effects of different pretreatment method on laser-induced breakdown spectroscopy measurement of Pb in pork[J]. Chinese Journal of Analytical Chemistry, 2016, 44(7): 1029-1034.
- [7] Tan D Z, Wang Z, Xu B B, et al. Photonic circuits written by femtosecond laser in glass: improved fabrication and recent progress in photonic devices[J]. Advanced Photonics, 2021, 3(2): 024002.
- [8] He J, He J, Xu X Z, et al. Single-mode helical Bragg grating waveguide created in a multimode coreless fiber by femtosecond laser direct writing[J]. Photonics Research, 2021, 9(10): 2052-2059.
- [9] Li Y F, Hu J, Liu W, et al. High period frequency LIPSS emerging on 304 stainless steel under the irradiation of femtosecond laser double-pulse trains[J]. Chinese Optics Letters, 2021, 19(12): 123801.
- [10] Yuan S, Chin S L, Zeng H P. Femtosecond filamentation induced fluorescence technique for atmospheric sensing[J]. Chinese Physics B, 2015, 24: 014208.
- [11] 应璐娜, 周卫东. 对比分析多种化学计量学方法在激光诱导击穿光谱土壤元素定量分析中的应用[J]. 光学学报, 2018, 38(12): 1214002.  
Ying L N, Zhou W D. Comparative analysis of multiple chemometrics methods in application of laser-induced breakdown spectroscopy for quantitative analysis of soil elements[J]. Acta Optica Sinica, 2018, 38(12): 1214002.
- [12] 曾和平, 牛盛. 一种飞秒等离子体光栅诱导击穿光谱检测的方法及装置: CN109884034A[P]. 2019-06-14.  
Zeng H P, Niu S. A method and device for detecting femtosecond plasma grating induced breakdown spectrum: CN109884034A[P]. 2019-06-14.
- [13] 张贵银, 张连水, 杨晓冬, 等. NO 分子  $A^2\Sigma \rightarrow X^2\Pi$  双光子激光感生荧光光谱[J]. 光学学报, 2003, 23(9): 1119-1122.  
Zhang G Y, Zhang L S, Yang X D, et al. Two-photon laser-induced fluorescence spectrum of NO ( $A^2\Sigma \rightarrow X^2\Pi$ ) transition[J]. Acta Optica Sinica, 2003, 23(9): 1119-1122.
- [14] 周昆鹏, 刘双硕, 崔健, 等. 基于荧光发射光谱的水质化学需氧量的检测[J]. 光谱学与光谱分析, 2020, 40(4): 1143-1148.  
Zhou K P, Liu S S, Cui J, et al. Detection of chemical oxygen demand (COD) of water quality based on fluorescence emission spectra[J]. Spectroscopy and Spectral Analysis, 2020, 40(4): 1143-1148.
- [15] 石鲁珍, 张景川, 王彦群, 等. 马氏距离与浓度残差剔除近红外异常样品研究[J]. 中国农机化学报, 2016, 37(6): 99-103.  
Shi L Z, Zhang J C, Wang Y Q, et al. Research for eliminating outlier samples in near-infrared spectroscopy by method of mahalanob distance-concentration residual[J]. Journal of Chinese Agricultural Mechanization, 2016, 37(6): 99-103.
- [16] 郑开逸, 封韬, 张文, 等. 加权 SPXYE(WSPXYE)算法及其在近红外光谱模型转移中的应用(英文)[J]. 光谱学与光谱分析, 2021, 41(3): 984-989.  
Zheng K Z, Feng T, Zhang W, et al. Weight SPXYE (WSPXYE) and its application to transfer set selection in near infrared spectra[J]. Spectroscopy and Spectral Analysis, 2021, 41(3): 984-989.
- [17] 徐杰, 陈颂伟, 谭煜亮, 等. 红外光谱预处理对粘纤/氨纶交织物定量分析模型的影响[J]. 化纤与纺织技术, 2020, 49(1): 49-52.  
Xu J, Chen S W, Tan Y L, et al. Influence of infrared spectrum pretreatment on the quantitative analysis model of viscose-spandex interweave fabric[J]. Chemical Fiber & Textile Technology, 2020, 49(1): 49-52.
- [18] 徐向君, 王宪双, 李昂泽, 等. 基于激光诱导击穿光谱的茶叶品种快速分类[J]. 中国激光, 2019, 46(3): 0311003.  
Xu X J, Wang X S, Li A Z, et al. Fast classification of tea varieties based on laser-induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2019, 46(3): 0311003.
- [19] 杨晖, 黄林, 陈添兵, 等. 光谱滤波法提高激光诱导击穿光谱对蔬菜中元素 Pb 的检测精度[J]. 分析化学, 2017, 45(8): 1123-1128.  
Yang H, Huang L, Chen T B, et al. Spectral filtering method for improvement of detection accuracy of lead in vegetables by laser induced breakdown spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2017, 45(8): 1123-1128.
- [20] 杨雪倩, 于慧春, 殷勇, 等. 拉曼光谱法检测玉米中黄曲霉毒素  $B_1$  和玉米赤霉烯酮[J]. 核农学报, 2021, 35(1): 159-166.  
Yang X Q, Yu H C, Yin Y, et al. Detection of aflatoxin  $B_1$  and Zearalenone in maize by Raman spectroscopy[J].



- Journal of Nuclear Agricultural Sciences, 2021, 35(1): 159-166.
- [21] 章琳颖, 黎静, 饶洪辉, 等. 基于 LIBS 的黄龙病脐橙元素检测与品质鉴别[J]. 激光与光电子学进展, 2020, 57(23): 233002.
- Zhang L Y, Li J, Rao H H, et al. LIBS-based element detection and quality identification of Huanglongbing navel oranges[J]. Laser & Optoelectronics Progress, 2020, 57(23): 233002.
- [22] 张根伟, 彭思龙, 郭腾霄, 等. 一种离子迁移谱谱图重构及特征峰提取算法[J]. 光谱学与光谱分析, 2020, 40(9): 2681-2685.
- Zhang G W, Peng S L, Guo T X, et al. Ion mobility spectrometry spectrum reconstruction and characteristic peaks extraction algorithm research[J]. Spectroscopy and Spectral Analysis, 2020, 40(9): 2681-2685.
- [23] 惠光艳, 孙来军, 王佳楠, 等. 可见-近红外光谱的小麦硬度预测模型预处理方法的研究[J]. 光谱学与光谱分析, 2016, 36(7): 2111-2116.
- Hui G Y, Sun L J, Wang J N, et al. Research on the pre-processing methods of wheat hardness prediction model based on visible-near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2016, 36(7): 2111-2116.
- [24] 王淑贤, 肖航, 杨振发, 等. 香精掺假普洱茶的近红外光谱检测[J]. 激光与光电子学进展, 2020, 57(23): 233005.
- Wang S X, Xiao H, Yang Z F, et al. Detection of flavor adulterated Pu'er tea by near-infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(23): 233005.
- [25] 刘燕德, 张雨, 徐海, 等. 基于近红外光谱检测不同产地石榴的糖度[J]. 激光与光电子学进展, 2020, 57(1): 013002.
- Liu Y D, Zhang Y, Xu H, et al. Detection of sugar content of pomegranates from different producing areas based on near-infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(1): 013002.
- [26] 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 F, Wang P, Zhang N C, et al. Improved filtering algorithm based on wavelet transform and its application in spectral denoising[J]. Foreign Electronic Measurement Technology, 2020, 39(8): 98-104.
- [27] 余其彪. 基于高光谱的沉积物特征组分预测及沉积环境指示研究[D]. 南京: 南京师范大学, 2014.
- Yu Q B. Research on prediction of sediment characteristic components and deposition environment indicator based on hyperspectral[D]. Nanjing: Nanjing Normal University, 2014.