

## 激光与光电子学进展

基于 SPA-SVR 的紫外光谱水质污染物含量解耦  
预测方法姜吉光<sup>1\*\*</sup>, 石磊<sup>1</sup>, 苏成志<sup>2\*</sup>, 常川<sup>1</sup>, 李笑天<sup>1</sup>, 侯小龙<sup>1</sup>, 田爱鑫<sup>1</sup><sup>1</sup>长春理工大学机电工程学院, 吉林 长春 130022;<sup>2</sup>长春理工大学人工智能研究院, 吉林 长春 130022

**摘要** 快速、精准地实现水体中多种污染物的耦合干扰解析及含量检测对野外水质实时监测具有重要意义。针对紫外光谱法同步检测化学需氧量(COD)和浊度时存在特征耦合及谱峰重叠干扰,进而严重影响检测精度的问题,提出了一种连续投影算法结合支持向量回归的水质污染物含量解耦预测方法。采用连续投影算法对水质样本的紫外吸收光谱特征波长进行筛选,消除无关冗余数据以提高模型迭代速率和精度。基于多分类支持向量机思想对支持向量回归算法进行多回归拟合改进,实现 COD 和浊度的紫外光谱耦合解析和含量的同步预测。通过实际水样检测验证,结果表明:耦合解析前的预测均方根误差改进率达到 76%,最大相对误差均降低至 4% 以内,优于同类方法的检测精度,该研究对紫外光谱法水质多耦合参数检测应用具有参考价值。

**关键词** 光谱学; 紫外光谱法; 化学需氧量; 浊度; 耦合预测; 支持向量回归

中图分类号 X853

文献标志码 A

DOI: 10.3788/LOP220700

Decoupled Prediction Method for Water Pollutant Concentration Based on  
SPA-SVR Using Ultraviolet SpectroscopyJiang Jiguang<sup>1\*\*</sup>, Shi Lei<sup>1</sup>, Su Chengzhi<sup>2\*</sup>, Chang Chuan<sup>1</sup>, Li Xiaotian<sup>1</sup>, Hou Xiaolong<sup>1</sup>,  
Tian Aixin<sup>1</sup><sup>1</sup>College of Mechanical and Electrical Engineering, Changchun University of Science and Technology,  
Changchun 130022, Jilin, China;<sup>2</sup>Artificial Intelligence Research Institute, Changchun University of Science and Technology,  
Changchun 130022, Jilin, China

**Abstract** The rapid and accurate coupling interference analysis and concentration detection of the multiple pollutants in complex water bodies are significantly important for the *in-situ* real-time monitoring of field water quality. To address the problems of characteristic coupling and the interference of spectral peaks in the synchronous detection of chemical oxygen demand (COD) and turbidity using ultraviolet spectroscopy, which significantly affect the detection accuracy, a decoupling method for predicting water pollutant concentration was developed in this study based on the continuous projection algorithm combined with support vector regression. The continuous projection algorithm was used to screen the characteristic wavelengths of the ultraviolet absorption spectra of water quality samples and eliminate irrelevant redundant numbers, to improve the iteration rate and accuracy of the model. Based on the concept of the multi-classification support vector machine, the support vector regression algorithm was improved via multi-regression fitting, and the ultraviolet coupling analysis of COD and turbidity, as well as the simultaneous prediction of concentration, was realized. The test results for actual water samples reveal that the maximum relative errors are reduced to less than 4%, and the improvement rate of the root mean square error of predictions before the coupling analysis reaches 76%. Thus, the proposed method offers a better detection accuracy, as compared with similar methods. Notably, this work is expected to serve as a reference for the application of ultraviolet spectroscopy in water-quality multi-coupling parameter detection.

收稿日期: 2022-02-09; 修回日期: 2022-02-10; 录用日期: 2022-02-16; 网络首发日期: 2022-02-26

基金项目: 国防基础科研计划(JCKY2019411B001)、吉林省科技发展计划项目(20210201041GX)、吉林省教育厅基金项目(JJ-KH20181124KJ)

通信作者: \*Chengzhi\_su@126.com; \*\*jiangjiguang1980@126.com

**Key words** spectroscopy; ultraviolet spectrum; chemical oxygen demand; turbidity; coupling prediction; support vector regression

## 1 引言

化学需氧量(Chemical oxygen demand, COD)和浊度(Turbidity, TURB)分别代表水体中还原性有机物和悬浮物的含量,是水质污染的重要监测指标<sup>[1]</sup>。光谱分析方法响应灵敏、分析速度快、重复性好,其中紫外光谱法因其检测原理成熟,可以间接反映水中有机物、悬浮物等理化参数,从而受到广泛关注<sup>[2-3]</sup>。在实际检测中,浊度散射会对COD的紫外吸收光谱造成严重干扰<sup>[4]</sup>,从而影响光谱法检测精度;基于透射法的浊度检测主要利用近红外波段,二者检测波段不统一,且透射法对低浊度液体的测量响应精度低,受水中有机污染物对光的吸收作用影响严重<sup>[5]</sup>。因此,基于紫外光谱法实现COD和浊度的精确同步检测,必须解决光谱特征耦合及谱峰重叠干扰等问题。

当前光谱法检测主要基于朗伯-比尔定律,通过透射吸收光谱进行污染物含量分析及干扰补偿研究。例如,Alam等<sup>[6]</sup>采用线性回归算法自动识别COD最大敏感光谱区和波长,对废水中的COD实现快速检测;李文等<sup>[7]</sup>设计了一种光强可调的浊度传感器,通过改变光强或机械结构来改善透射法浊度检测在不同量程的分辨力。Hu等<sup>[8]</sup>、吴德操等<sup>[9]</sup>分别提出了基于归一化数值动态模拟和Mie散射理论的浊度补偿方法,提高了COD紫外光谱法检测的精度;张鹤强<sup>[10]</sup>通过多参数耦合交叉分析方法探究了水体中COD和浊度、溶解氧等参数的相关性及干扰机理。Agustsson等<sup>[11]</sup>基于

漫反射理论探究了紫外-可见光谱法检测COD和浊度的可行性。

在野外复杂的水质情况下,COD和浊度本身存在耦合影响关系,二者检测方法不统一,耦合干扰下基于透光率测量的光谱分析方法精度有待提高,难以实现基于单一谱源的精准同步检测。本文探究了COD与浊度在紫外吸收光谱检测中的耦合干扰关系,基于多分类支持向量机在污染物类型判别<sup>[12]</sup>的思想改进了支持向量机在多耦合参数回归分析的应用,提出了一种基于连续投影算法结合支持向量回归机(SPA-SVR)的水质耦合参数含量解耦的预测方法,通过溶液的紫外吸收光谱对其COD、浊度进行耦合干扰解析和含量检测。结果表明,该模型改善了COD和浊度的耦合干扰对紫外光谱法检测的影响,具有较好的预测精度,能够实现基于单种谱源对多个水质污染指标含量的分析检测。

## 2 原理和方法

使用光谱检测系统采集到样本的紫外吸收光谱原始数据后,为探究COD和浊度的耦合干扰关系,并通过训练得到效果良好的多参数光谱检测模型,以实现COD和浊度耦合解析及含量检测,需要通过连续投影算法(SPA)对光谱数据进行特征提取,消除大量无关冗余数据及偏移明显的干扰噪点,选取具有代表性的特征波长来表征原始光谱信息<sup>[13]</sup>,并基于支持向量机(SVR)算法对光谱特征数据集进行学习训练,检测模型的流程如图1所示。

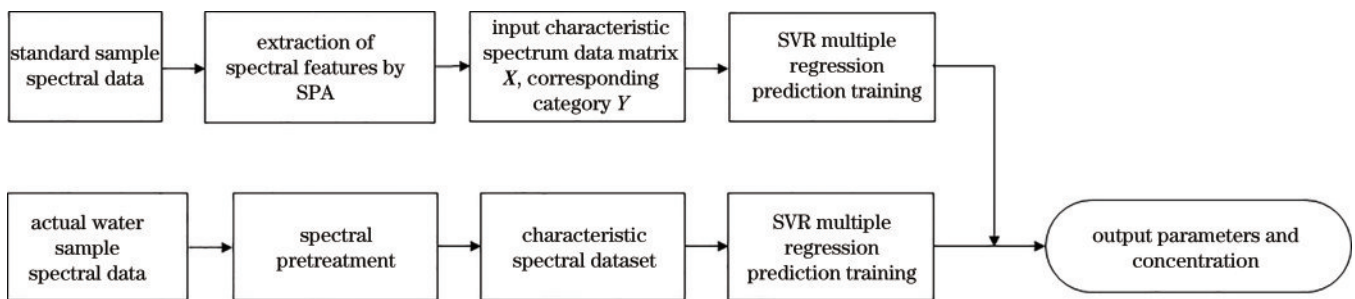


图1 耦合参数检测模型流程图

Fig. 1 Flow chart of coupling parameters detection model

### 2.1 基于SPA的光谱特征提取方法

根据朗伯-比尔定律,吸收光谱能够体现溶液所含成分及含量。但同时光谱数据中包含大量的冗余信息,训练数据的体量直接决定了预测模型的迭代速率以及预测效果的精确度,庞大的数据矩阵容易使得训练模型的稳定性降低,预测结果易陷入局部最优解。SPA在选择含有最少冗余信息及最小共线性的变量组合的同时保留了原始数据的绝大部分特征,被选取的特征波长数据集物理意义明确,具有很强的解释能

力<sup>[14-16]</sup>,可以有效地提高建模的速度和模型的稳定性,加快模型收敛到最优解。该算法的原理如下:

记初始迭代向量为 $\mathbf{X}_{k(0)}$ ,需要提取的变量个数为 $N$ ,光谱矩阵为 $J$ 列。

1) 迭代开始前,任选光谱矩阵的1列,用 $i$ 表示,即把原始光谱矩阵的第 $i$ 列 $\mathbf{x}_i$ 赋值给 $\mathbf{X}_{k(0)}$ ;

2) 把原始光谱矩阵中未选入的列向量位置集合记为 $s, s = \{j, 1 \leq j \leq m; j \notin [k(0), \dots, k(n-1)]\}$ ;

3) 计算当前多选变量  $\mathbf{x}_{k(n-1)}$  对剩余原始光谱数据列向量  $\mathbf{x}_i (i \in s)$  的投影:

$$\mathbf{P}_x = \mathbf{x}_i (\mathbf{x}_i^T \mathbf{x}_{k(n-1)}) \cdot \mathbf{x}_{k(n-1)} (\mathbf{x}_{k(n-1)}^T \mathbf{x}_{k(n-1)})^{-1}; \quad (1)$$

4) 提取最大投影波长变量序号:

$$k(n) = \arg [\max (\|\mathbf{P}_x\|)]; \quad (2)$$

5) 令  $\mathbf{x}_i = \mathbf{P}_x, i \in s$ ;

6)  $n = n + 1$ , 如果  $n < N$ , 回到步骤 2) 循环计算。

最后提取的特征波长变量为  $\{\mathbf{x}_{k(n)}, n = 0, \dots, N - 1\}$ 。对应每一个  $k(0)$  和  $N$ , 每循环一次算法, 对提取的特征波长处的光谱吸光度进行多元线性回归分析, 计算验证集的预测标准偏差, 其中最小预测标准偏差对应的  $k(0)$  和  $N$  就是最优的特征波长组合。

## 2.2 基于 SVR 的多参数回归分析算法

支持向量回归对非线性问题具有良好的拟合性能, 基于结构风险最小化原则使得其具有良好的泛化能力。基于多分类支持向量机 (SVM) 的思想, 通过标准 SVR 的计算流程有序地建立多个拟合超平面对不同类型的标签进行特征分类回归, 可以同时实现多个参数的拟合预测。因其具有稀疏性的特点, 多个耦合参数可通过引入适当的不敏感损失函数和惩罚因子来解决其耦合干扰的影响, 从而实现多参数耦合解析和回归分析。参数优化是 SVR 建模中一个重要的问题, 惩罚因子  $C$  与核参数  $g$  的选取直接影响 SVR 的预测精度和拟合效果。本文中的  $C$ 、 $g$  提前由粒子群算法选定最优值  $C = 14$ 、 $g = 0.29$ 。

非线性 SVR 的基本思想是通过事先确定的非线性映射将输入向量映射到一个高维特征空间 (Hilbert 空间) 中, 然后在此高维空间中进行线性回归, 从而得到在原空间非线性回归的效果。

训练样本  $(\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \dots, n, \mathbf{x}_i \in \mathbf{R}^n$  为输入量,  $\mathbf{y}_i \in \mathbf{R}^m$  为输出量, 其回归函数为

$$f(\mathbf{x}) = \boldsymbol{\omega} \cdot \boldsymbol{\phi}(\mathbf{x}) + b, \quad (3)$$

式中:  $\boldsymbol{\phi}(\mathbf{x})$  为非线性映射函数, 可以将输入向量  $\mathbf{x}$  映射到高维特征空间;  $\boldsymbol{\omega}$  为系数向量;  $b$  为阈值。

标准支持向量回归机采用  $\epsilon$ -不敏感损失函数, 且通过结构风险最小化原则得到 SVR 的目标函数和约束条件为

$$\begin{aligned} R(\boldsymbol{\omega}, \xi_i, \xi_i^*) &= \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \\ \text{s. t. } \begin{cases} \mathbf{y}_i - f(\mathbf{x}_i) \leq \epsilon + \xi_i, i = 1, 2, \dots, n \\ f(\mathbf{x}_i) - \mathbf{y}_i \leq \epsilon + \xi_i^*, i = 1, 2, \dots, n \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

式中:  $\xi_i, \xi_i^*$  为松弛因子。

此时回归问题就转化为关于待求参数  $\boldsymbol{\omega}$ 、 $b$  的凸二次规划问题, 引入拉格朗日函数得到式 (4) 的对偶形式:

$$\max(\alpha, \alpha^*) = \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \cdot$$

$$[\boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)] + \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{y}_i - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \epsilon,$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C], i = 1, 2, \dots, n \end{cases}, \quad (5)$$

式中:  $\boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)$  为高维空间的点积运算, 令  $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)$ , 称  $K(\mathbf{x}_i, \mathbf{x}_j)$  为核函数, 通过引入核函数解决维数问题, 可在不知映射函数的情况下实现回归估计。核函数的类型有多种, 将高斯径向基函数 (RBF) 作为预测模型的核函数。

将核函数代入式 (5) 求解, 得到非线性拟合函数表达式为

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b, \quad (6)$$

式中:  $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / 2\sigma^2)$ ,  $\sigma$  为标准差。

对于  $K$  分类问题, 多分类支持向量机有  $K$  组参数  $\{(\boldsymbol{\omega}_1, b_1), (\boldsymbol{\omega}_2, b_2), \dots, (\boldsymbol{\omega}_K, b_K)\}$ , 希望模型对属于正确标记的结果以 1 的间隔高于其他类的结果:

$$\begin{aligned} \min_{\boldsymbol{\omega}, b} \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K \max \{0, [\boldsymbol{\omega}_k^T \boldsymbol{\phi}(\mathbf{x}) + b_k] - \\ [\boldsymbol{\omega}_k^T \boldsymbol{\phi}(\mathbf{x}_i) + b_k] + 1\} + \frac{\lambda}{2} \sum_{k=1}^K \boldsymbol{\omega}_k^T \boldsymbol{\omega}_k \end{aligned} \quad (7)$$

式中:  $\lambda = C/m$ 。

基于这种思想, 多分类支持向量回归算法依次建立多个回归拟合超平面, 对每个参数进行分类回归拟合。对于无法直接通过  $\epsilon$ -不敏感损失函数进行分类评价的特殊点, 引入权重向量  $\boldsymbol{w}$  的 L1 范数代替其最大间隔项, 将未知测试样本的元素近似表示为所有训练样本的稀疏线性组合, 权重向量  $\boldsymbol{w}$  被当作稀疏表示系数, 从而将  $\epsilon$ -SVR 模型变为多分类回归模型:

$$\min \|\boldsymbol{w}\|_1 + C \sum_{i=1}^m (\xi_i + \xi_i^*),$$

$$\text{s. t. } \begin{cases} \mathbf{y}_i - \mathbf{x}_i \boldsymbol{w} - b \leq \epsilon + \xi_i \\ \mathbf{x}_i \boldsymbol{w} + b - \mathbf{y}_i \leq \epsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \dots, m \end{cases}. \quad (8)$$

## 3 数据分析与处理

### 3.1 仪器与样品

实验所用光谱采集系统由 ATP2000P 光谱仪 (OPTOSKY, 厦门)、ATG1020 脉冲氙灯、中控器、比色皿、光纤衰减器及显示器等组成。采用 10 mm 石英比色皿为样品池, 通过光纤衰减器对脉冲光源的输出光功率进行调节, 以保证吸收光谱的稳定性, 避免光饱和和失真和波动噪声。光谱仪测试波长范围为 180~1000 nm, 分辨率为 2.3 nm, 信噪比 > 500:1。光谱数



据在 Optosky Spectra V2.99 光谱分析软件显示。采集系统如图 2 所示。

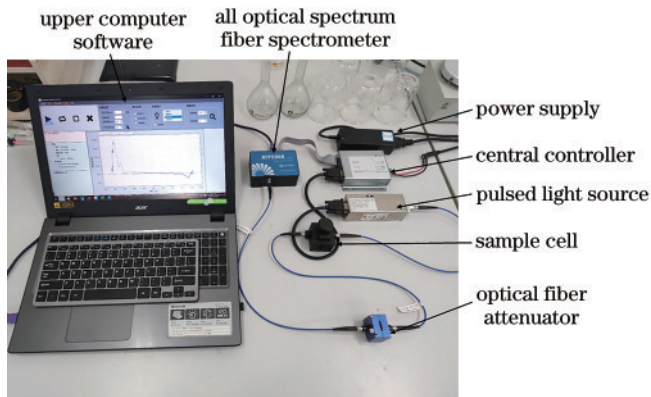
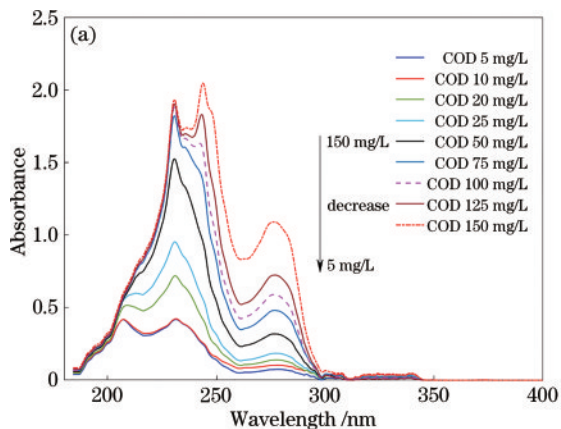


图 2 紫外光谱检测系统

Fig. 2 Ultraviolet spectrum detection system

实验样品采用国家有色金属及电子材料分析测试中心提供的以邻苯二甲酸氢钾为溶质,质量浓度为 500 mg/L 的 COD 标准溶液;杭州齐威仪器有限公司



提供 400 NTU 的福尔马肼浊度标准溶液。分别使用去离子水按比例稀释为质量浓度范围在 5~150 mg/L 间的梯度质量浓度 COD 标准溶液和 5~100 NTU 间的梯度浊度标准溶液,并按正交分析实验配置 COD-浊度交叉混合溶液作为训练样本。采集紫外吸收光谱数据后由 Matlab 软件进行模型训练和耦合分析。

### 3.2 光谱数据采集

在水体环境中,污染物对紫外光的吸收遵循朗伯-比尔定律,吸光度会随着特征污染物的含量增大而增大,如果吸光度与含量之间存在较强的正相关,就可以根据紫外光谱的吸光度确定水质特征污染物的含量<sup>[17]</sup>。

#### 3.2.1 COD 标准光谱

为提取 COD 的紫外吸收光谱特征,分别采集质量浓度为 5、10、20、25、50、75、100、125、150 mg/L 的 COD 标准溶液紫外吸收光谱,共 9 个样本,有效光谱范围为 180~400 nm,积分时间 9 ms,采样间隔为 500 ms,平均扫描次数为 20 次。每个样品采集 5 次光谱数据后取平均值,光谱图如图 3(a)所示。

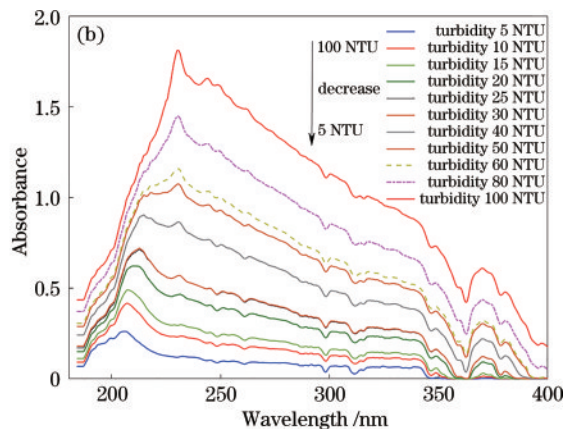


图 3 标准溶液紫外吸收光谱。(a)COD;(b)浊度

Fig. 3 UV absorption spectra of standard solution. (a) COD; (b) turbidity

#### 3.2.2 浊度标准光谱

为提取浊度的标准溶液紫外吸收光谱特征,分别采集浊度为 5、10、15、20、25、30、40、50、60、80、100 NTU 的福尔马肼浊度溶液紫外吸收光谱,样本共 11 个,采集方法与 COD 标准样品相同。采集到的紫外吸收光谱如图 3(b)所示。

#### 3.2.3 混合样本光谱

为探究 COD 与浊度的耦合干扰影响,配置 COD 质量浓度范围为 0~150 mg/L 而浊度范围为 0~60 NTU 的梯度浓度和浊度混合溶液样本共计 90 个,采集混合样本的紫外吸收光谱。分别以质量浓度为 50 mg/L 的 COD 与不同浓度和浊度的交叉混合溶液、浊度为 40 NTU 与不同质量浓度 COD 的交叉混合溶液为例,光谱图如图 4 所示。

通过对比各标准溶液样本光谱谱线可以看出, COD 和浊度的主要吸收特征集中在 180~400 nm 范

围。COD 特征峰范围在 180~300 nm 之间,随着质量浓度增大紫外光谱吸光度逐渐增大,两个特征吸收峰明显向上抬升;浊度光谱吸收特征峰范围在 180~400 nm 之间,随着浊度增大吸光度整体增大。混合样本中,浊度干扰增强使得 COD 吸收谱线向上抬升,主吸收峰向右偏移, COD 干扰增强使得浊度吸收峰向上明显抬升,在 300~400 nm 段受干扰影响较小。

### 3.3 光谱特征数据预处理

通过 SPA 对纯 COD 溶液、纯浊度溶液以及混合溶液共计 110 个实验样本的原始光谱数据进行特征波长选取,建立特征光谱数据集。SPA 能够筛选具有最小共线性的特征波长向量组合,使得选出的特征数据集里含有最少的冗余信息且能够代表完整的原始光谱数据特征,反映水质污染物含量与吸光度的相关性。选取的具有最小共线性的纯溶液特征波长向量组合如图 5 所示。

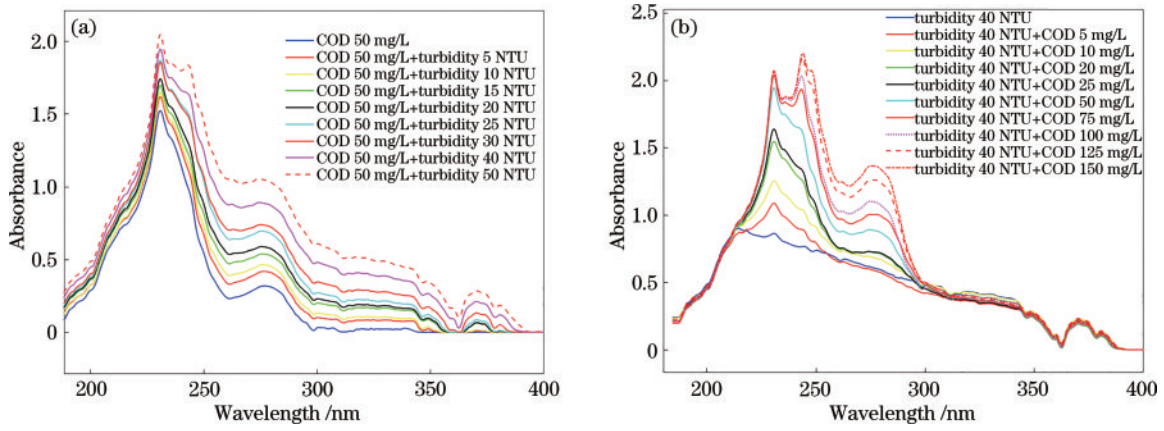


图 4 混合溶液紫外吸收光谱。(a) COD+浊度干扰;(b) 浊度+COD 干扰

Fig. 4 UV absorption spectra of mixed solution. (a) COD+turbidity interference; (b) turbidity+COD interference

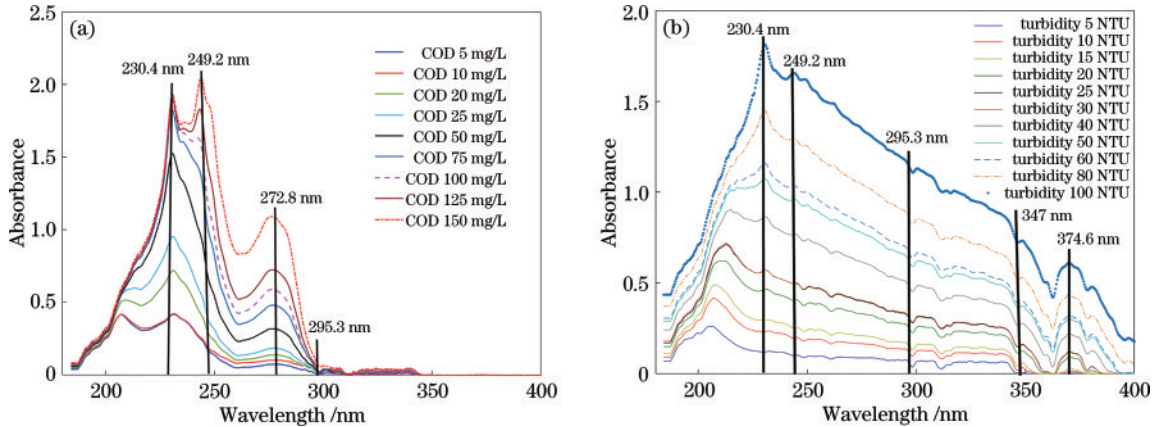


图 5 特征波长选取。(a) COD;(b)浊度

Fig. 5 Selection of characteristic wavelength. (a) COD; (b) turbidity

通过 SPA 进行光谱数据预处理选取 COD 光谱特征波长分别为 230.4、249.2、272.8、295.3 nm；浊度光谱特征波长为 230.4、249.2、295.3、347、374.6 nm。为了确保 SPA 算法所选取的特征波长的可靠性，分析了选取的各特征波长处样品含量与吸光度的线性相关性，如图 6 所示， $R^2$  均大于 0.98，由此可知，通过 SPA 选取的特征波长物理意义明确，具有很强的解释能力。(SSE 为残差平方和；RMSE 为均方根误差)将各标准溶液及混合样本的特征波长光谱数据建立特征数据集对 SVR 模型进行训练。部分特征数据如表 1 所示。

### 3.4 分析与讨论

通过对 COD、浊度纯溶液及交叉混合溶液样本的光谱曲线进行对比，可以看出 COD 和浊度之间的耦合干扰关系对特征吸收光谱存在明显的干扰规律。为进一步探究其耦合干扰关系对紫外光谱法检测的影响程度，选取 COD 和浊度特征相关性最强的波长分别为 272.8 nm 和 249.2 nm，建立吸光度与含量三维特征点图，如图 7 所示。并基于朗伯-比尔定律单波长紫外光谱法检测原理对耦合干扰样本进行线性拟合预测，得到的预测结果如图 8 所示。

表 1 紫外光谱特征数据集

Table 1 UV spectral characteristic dataset

Sample	COD /mg·L <sup>-1</sup>	Turbidity / NTU	Absorbance						
			$\lambda_{230.4}$	$\lambda_{249.2}$	$\lambda_{272.8}$	$\lambda_{295.3}$	...	$\lambda_{374.6}$	
1	5	0	0.415	0.147	0.069	0.020	...	0.001	
2	0	5	0.120	0.092	0.089	0.076	...	0.003	
3	5	5	0.400	0.211	0.139	0.090	...	0.004	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
108	125	50	2.108	2.047	1.384	0.710	...	0.252	
109	150	50	2.122	2.159	1.503	0.748	...	0.265	
110	150	60	2.157	2.394	1.712	0.763	...	0.384	



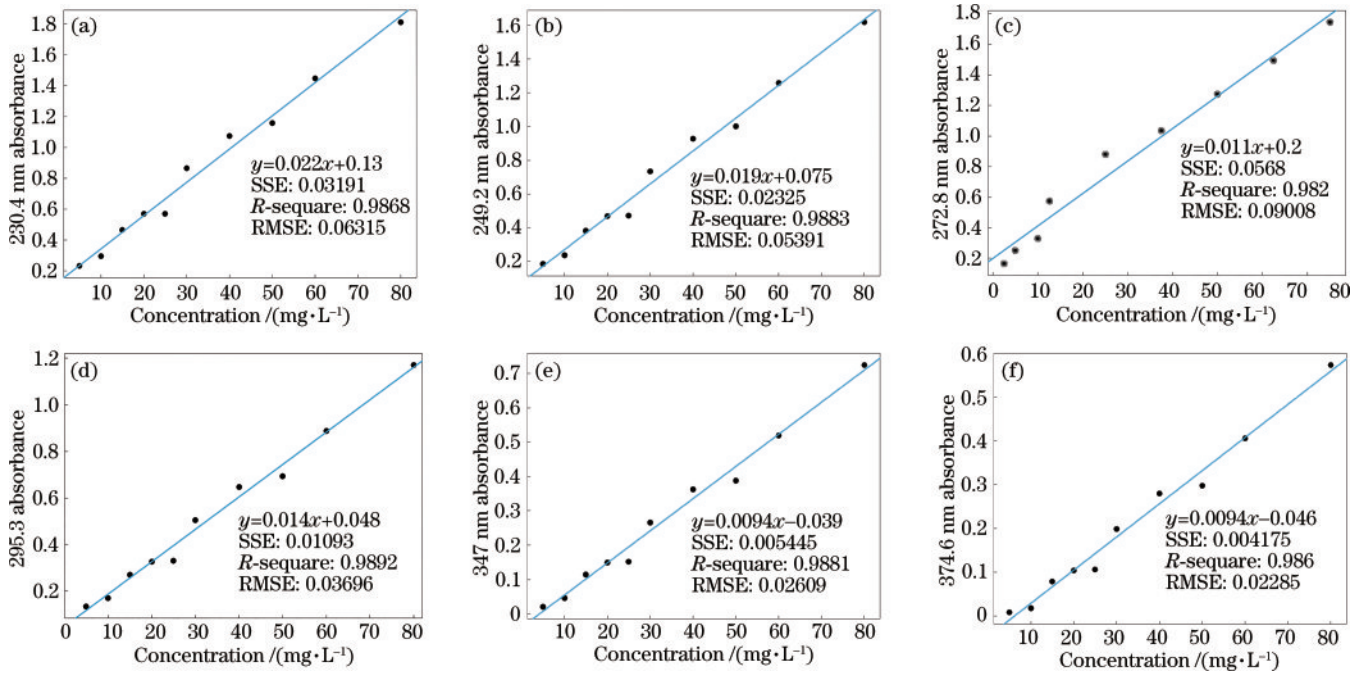


图 6 特征波长相关性分析。(a) 230.4 nm; (b) 249.2 nm; (c) 272.8 nm; (d) 295.3 nm; (e) 347 nm; (f) 374.6 nm  
 Fig. 6 Selection of characteristic wavelength of COD and turbidity. (a) 230.4 nm; (b) 249.2 nm; (c) 272.8 nm; (d) 295.3 nm; (e) 347 nm; (f) 374.6 nm

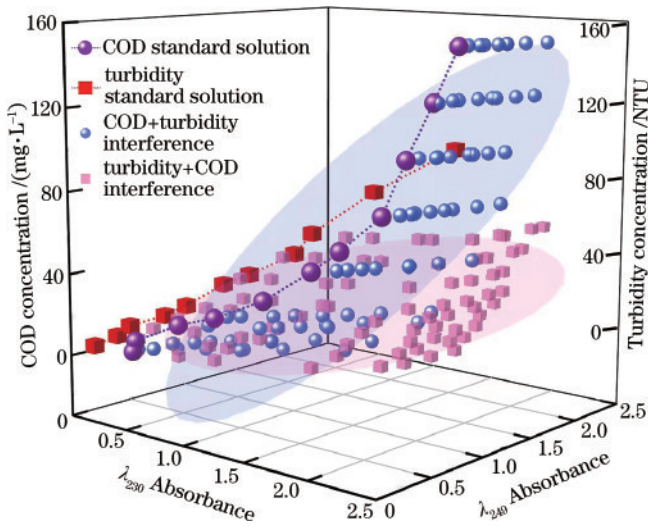


图 7 紫外光谱三维特征点图  
 Fig. 7 Three dimensional characteristic point diagram of ultraviolet spectrum

通过耦合干扰预测结果可以看出,浊度散射对 COD 的光谱法检测干扰明显,最大预测偏差超过 150 mg/L,最小预测误差为 8.4%,且其散射效应对不同质量浓度的 COD 标准溶液干扰程度相似,说明不同量程下的 COD 检测均有浊度补偿的必要。COD 对紫外光谱的吸收同样对浊度检测产生明显干扰,最大预测偏差超过 100 NTU,最小预测误差为 16.3%。COD 对于低浊度溶液耦合干扰程度较大,对高浊度溶液的干扰逐渐减小,这也符合透射法对高浊度溶液检测精度高、对低浊度溶液分辨率低的规律。

通过 SPA 提取特征波长后,光谱数据得到简化,排除了无关的冗余变量,极大地减少了干扰信息。通过粒子群优化算法对 SVR 进行参数优化,得到  $C=14g=0.29$ 。将特征光谱数据集矩阵随机排列后作为输入并代入 SVR 算法中进行训练,学习 COD、浊度与紫外光谱吸收度之间的映射特征。其中,训练集与预测集比例为 9:2,得到耦合解析后水质 COD 和浊度的预测结果,如图 9 所示。

通过耦合解析前后水质 COD 和浊度的预测结果可以得出,相较于传统线性回归拟合模型预测,SPA-SVR 模型的回归效果更好,预测均方根误差改进率达到 76%,经过耦合解析后模型预测误差均控制在 4% 范围以内。为了进一步验证多参数检测模型的精确性,分别采集长春某地区河流、湖泊及校园观景池等多处实际水样作为验证样本,采集光谱数据对模型进行检测验证。实际水样中 COD 含量和浊度值经天津某环境研究院采用重铬酸钾法及浊度计法测定标准值作为对照。验证结果如表 2 所示。

经实际水样检测验证,对比 SPA-SVR 模型的解耦预测结果与国标方法的检测数据,其预测误差均小于 4%,证明该方法预测精度更好,优于同类检测方法。根据预测结果计算出各模型的决定系数  $R^2$ 、均方根误差以及相对标准偏差作为模型的评价指标。实际水样的预测性能指标如表 3 所示。由此可知,基于多分类思想的支持向量回归模型可以有效减小耦合干扰影响,实现水样中 COD 含量和浊度的精确同步预测,且对低量程浊度溶液同样具有较高的检测精度,有效实现了 COD 和浊度紫外光谱法的同步检测。

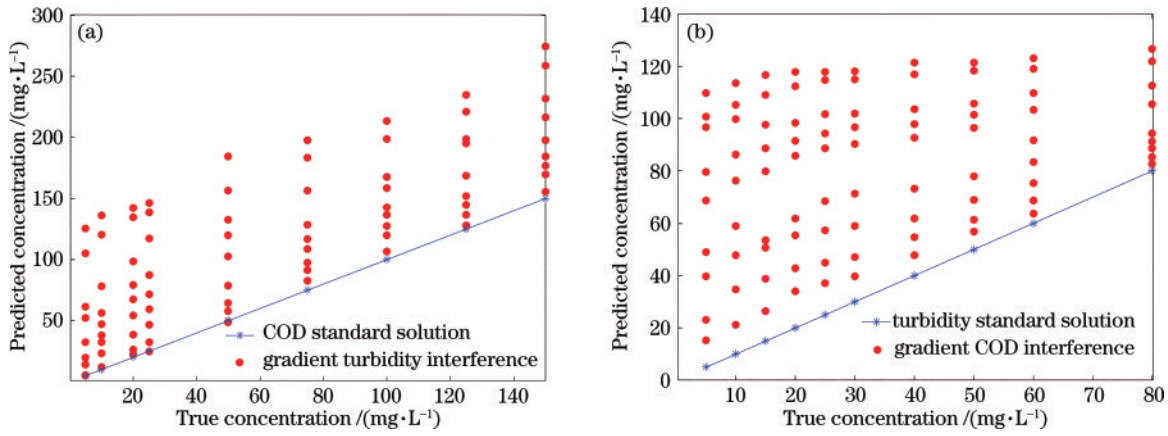


图 8 耦合干扰结果对比。(a) COD+浊度干扰;(b) 浊度+COD 干扰

Fig. 8 Comparison of coupling interference results. (a) COD+turbidity interference; (b) turbidity+COD interference

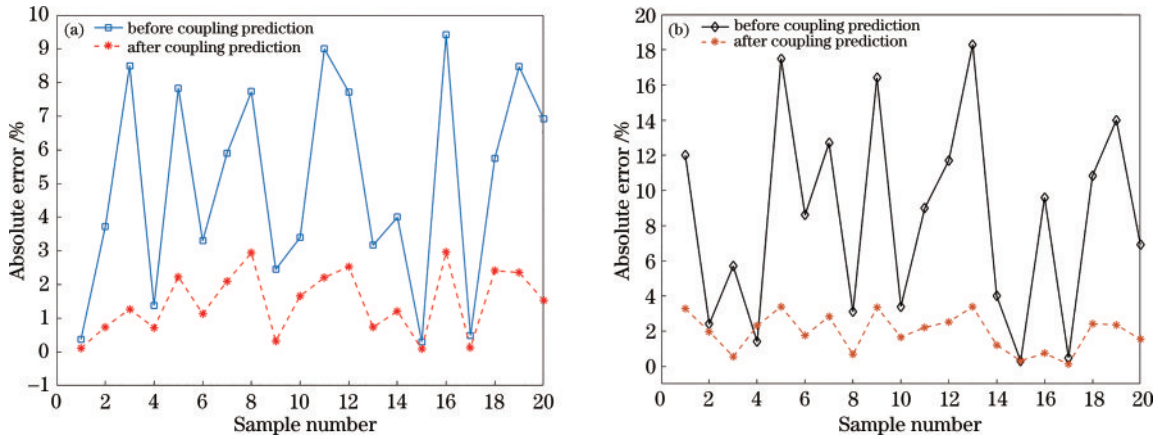


图 9 耦合解析预测结果对比。(a) COD;(b)浊度

Fig. 9 Comparison of coupling analytical prediction results. (a) COD; (b) turbidity

表 2 实际水样模型验证结果

Table 2 Verification results of actual water sample model

Sample	COD / (mg·L <sup>-1</sup> )		Prediction error COD / %	Turbidity / NTU		Prediction error Turbidity / %
	True value	Predicted value		True value	Predicted value	
1	9.60	9.46	1.46	0.40	0.39	2.50
2	17.50	17.72	1.26	3.00	3.05	1.60
3	10.60	10.24	3.40	0.40	0.41	2.50
4	16.40	16.42	0.12	0.50	0.51	4.00
5	25.30	24.84	1.82	12.00	11.79	1.75
6	8.50	8.62	1.41	6.30	6.42	1.90
7	21.30	21.33	0.14	21.20	20.87	1.56
8	72.80	71.97	1.14	34.80	35.20	1.15
9	43.70	44.60	2.02	19.60	19.57	0.15

表 3 SVR 模型预测性能指标

Table 3 SVR model prediction performance index

Prediction index	R <sup>2</sup>	RMSE / (mg·L <sup>-1</sup> )	RSD / %
COD	0.998	0.343	1.50
Turbidity		0.125	2.43

## 4 结 论

针对野外水中 COD 和浊度实时监测时,紫外光谱

法的检测精度受两者耦合干扰影响严重的问题,开展了基于 SPA-SVR 的水质耦合参数光谱检测研究,对紫外光谱法检测 COD 和浊度时的特征耦合和谱峰重叠干扰进行补偿。

通过实验获取了 COD 标准溶液、浊度标准溶液及混合溶液的紫外吸收光谱,对比分析了两者之间的耦合干扰关系,通过 SPA 提取了能够代表所有原始光谱数据特征的有效特征波长,分别为  $\lambda_{230.4}$ 、 $\lambda_{249.2}$ 、 $\lambda_{273.8}$ 、 $\lambda_{295.3}$ 、 $\lambda_{347}$ 、 $\lambda_{374.6}$ , 建立光谱特征数据矩阵对 SVR 模型进

行训练。训练结果表明,SVR算法与传统线性回归拟合预测模型相比,其对耦合解析前的预测均方根误差改进率达到76%,预测误差均保持在4%范围内。通过实际水样进行预测验证,其预测结果满足国标要求,优于同类检测方法。基于多分类SVM思想改进的多回归支持向量机一定程度上改善了COD和浊度的耦合干扰影响,从而提高多参数检测模型的准确率。该方法可适用于其他可溶性水质污染物的光谱特征分类和含量预测,对基于紫外光谱法的水质多耦合参数检测应用具有参考价值。

### 参 考 文 献

- [1] 周昆鹏,白旭芳,毕卫红. 荧光光谱法检测水质COD时温度、浊度、pH的影响分析[J]. 光谱学与光谱分析, 2019, 39(4): 1097-1102.  
Zhou K P, Bai X F, Bi W H. The temperature, turbidity and pH impact analysis of water COD detected by fluorescence spectroscopy[J]. Spectroscopy and Spectral Analysis, 2019, 39(4): 1097-1102.
- [2] 周思寒,胡新宇,汤斌,等. 一种基于EWMA-PCA的水质光谱数据标准化方法研究[J]. 光谱学与光谱分析, 2020, 40(11): 3443-3450.  
Zhou S H, Hu X Y, Tang B, et al. Application research of normalization algorithm based on EWMA-PCA in standardization of water quality spectral data[J]. Spectroscopy and Spectral Analysis, 2020, 40(11): 3443-3450.
- [3] 李庆波,何林倩,崔厚欣,等. 地表水亚硝酸盐氮浓度紫外可见光谱检测方法的基础研究[J]. 光谱学与光谱分析, 2020, 40(4): 1127-1131.  
Li Q B, He L Q, Cui H X, et al. Fundamental Study on the Detection of Nitrite Nitrogen Concentration in Surface Water by Ultraviolet Visible Spectroscopy, 2020, 40(4): 1127-1131.
- [4] 刘恩华,吴德操,王杰,等. 直接光谱法水体COD检测多场景优化方法[J]. 光学学报, 2021, 41(22): 2230001.  
Liu E H, Wu D C, Wang J, et al. Multi-scene optimization method of water COD measurement based on direct spectroscopy[J]. Acta Optica Sinica, 2021, 41(22): 2230001.
- [5] 罗继阳,魏彪,汤斌,等. 水质COD检测用光谱法的水体环境浊度解算方法[J]. 环境科学与技术, 2015, 38(S2): 246-250.  
Luo J Y, Wei B, Tang B, et al. A method to calculate the turbidity of water environment based on ultraviolet-visible spectrum for the measurement of COD in water[J]. Environmental Science & Technology, 2015, 38(S2): 246-250.
- [6] Alam T, Rezaia B, Bahreini B. A portable system for estimation of chemical oxygen demand in wastewater using ultraviolet-visible spectroscopy[C]//2015 IEEE SENSORS, November 1-4, 2016, Busan, Korea (South). New York: IEEE Press, 2016.
- [7] 李文,张志永,金旭,等. 基于光强可调的浊度智能检测传感器研究[J]. 仪表技术与传感器, 2020(2): 15-19.  
Li W, Zhang Z Y, Jin X, et al. Research on turbidity intelligent detection sensor based on adjustable light intensity [J]. Instrument Technique and Sensor, 2020(2): 15-19.
- [8] Hu Y T, Wen Y Z, Wang X P. Novel method of turbidity compensation for chemical oxygen demand measurements by using UV-vis spectrometry[J]. Sensors and Actuators B: Chemical, 2016, 227: 393-398.
- [9] 吴德操,魏彪,汤戈,等. 基于Mie散射的水体紫外-可见光谱浊度干扰补偿[J]. 光学学报, 2017, 37(2): 0230007.  
Wu D C, Wei B, Tang G, et al. Turbidity disturbance compensation for UV-VIS spectrum of waterbody based on Mie scattering[J]. Acta Optica Sinica, 2017, 37(2): 0230007.
- [10] 张鹤强. 基于多参数耦合交叉分析方法的COD测量技术研究[D]. 杭州: 浙江大学, 2020: 20-28.  
Zhang H Q. The research on COD measurement technologies based on multi-parameter cross-coupled analysis[D]. Hangzhou: Zhejiang University, 2020: 20-28.
- [11] Agustsson J, Akermann O, Barry D A, et al. Non-contact assessment of COD and turbidity concentrations in water using diffuse reflectance UV-Vis spectroscopy [J]. Environmental Science. Processes & Impacts, 2014, 16(8): 1897-1902.
- [12] 黄平捷,李宇涵,俞巧君,等. 基于SPA和多分类SVM的紫外-可见光谱饮用水有机污染物判别方法研究[J]. 光谱学与光谱分析, 2020, 40(7): 2267-2272.  
Huang P J, Li Y H, Yu Q J, et al. Classification of organic contaminants in water distribution systems developed by SPA and multi-classification SVM using UV-vis spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(7): 2267-2272.
- [13] 李伟晶. 基于紫外-可见光谱法的水质COD检测预测模型应用[J]. 东北水利水电, 2018, 36(1): 32-34, 44, 72.  
Li W J. Application of detection and prediction model of COD in water based on ultraviolet visible spectrometry [J]. Water Resources & Hydropower of Northeast China, 2018, 36(1): 32-34, 44, 72.
- [14] 李鑫,苏成志,于丹阳,等. 紫外光谱法检测COD时波长和低温影响的探究[J]. 光谱学与光谱分析, 2020, 40(8): 2403-2408.  
Li X, Su C Z, Yu D Y, et al. Study on the influence of wavelength and low temperature on COD detection by ultraviolet spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(8): 2403-2408.
- [15] 撒继铭,江河,谢凯文,等. 基于随机蛙跳算法的特征波长优选[J]. 光学学报, 2021, 41(15): 1530001.  
Sa J M, Jiang H, Xie K W, et al. Characteristic wavelength optimization based on random frog algorithm [J]. Acta Optica Sinica, 2021, 41(15): 1530001.
- [16] Yu Y S, Qu Y. Multi-component spectral detection based on neural network in water quality inspection[J]. Optik, 2020, 217: 164915.
- [17] 蔡嘉城,邓霄,阚崇轩,等. 基于紫外-可见光谱的水质COD在线检测系统设计[J]. 激光与光电子学进展, 2022, 59(1): 0130001.  
Cai J C, Deng X, Kan C X, et al. Design of water quality COD online detection system based on ultraviolet-visible spectroscopy[J]. Laser & Optoelectronics Progress, 2022, 59(1): 0130001.