

文章编号 1004-924X(2025)10-1627-11

基于三维荧光光谱的水体识别及 组分快速解析算法研究

姜赞成^{1,2}, 王瑞杰¹, 项晓亮¹, 叶彬强^{3,4*}, 冯 鹏^{1*}

(1. 重庆大学 光电技术及系统教育部重点实验室, 重庆 400044;

2. 四川碧朗科技有限公司, 四川 绵阳 621900;

3. 重庆理工大学 两江人工智能学院, 重庆 400054

4. 重庆大学 微电子与通信工程学院, 重庆 400044)

摘要:水环境污染监控和精细化管理需求日益增强,迫切需要对水中有机污染物进行快速、准确地检测和识别。三维荧光光谱技术可以提供丰富的污染物光谱信息,已成为当前水体污染物识别和溯源的研究热点。当前方法多集中于基于深度学习的光谱数据分析,对光谱数据量需求较大,难以在现场推广。本文利用三维荧光光谱数据(3D-EEM),提出一种基于二维 Gabor 小波结合支持向量机(SVM)的水体多分类识别和组分精确拟合的方法,有效提取三维荧光光谱的纹理特征和谱峰位置,提高了水样组分解析的效率。该方法采用空白扣除法和 Delaunay 三角形内插值法减少光谱数据中的背景噪声和散射干扰,通过扩展 Savitzky-Golay 平滑方法抑制光谱波动干扰。而后利用二维 Gabor 小波和荧光峰位提取法提取了 3D-EEM 数据的纹理特征信息和三维荧光峰位的全局信息。最后,基于 MSVC 和 CF_MSVM 构建了 EEM_MSVM 模型,以实现对不同水体的高准确度分类识别和组分预测。实际水样的实验结果表明,水体类别的分类准确度达到 97.6%,组分预测的均方根误差损失为 5.3,相关系数为 0.94,有效实现了对典型水体的准确分类和组分的解析。

关键词:三维荧光光谱;荧光组分图谱;水体识别;溯源追踪

中图分类号:X83;O433.5 文献标识码:A

doi:10.37188/OPE.20253310.1627

CSTR:32169.14.OPE.20253310.1627

Research on water identification and rapid analysis algorithm for components based on 3d fluorescence spectroscopy

JIANG Zancheng^{1,2}, WANG Ruijie¹, XU Xiaoliang¹, YE Binqiang^{3,4*}, FENG Peng^{1*}

(1. The Key Lab of Optoelectronic Technology and Systems, Ministry of Education,

Chongqing University, Chongqing 400044, China;

2. Sichuan Belam Technology Co., Ltd., Mianyang 621900, China;

3. College of Artificial Intelligence, Chongqing University of Technology, Chongqing 400054, China

收稿日期:2024-10-31;修订日期:2024-12-24.

基金项目:重庆市科委技术创新与应用发展专项(No. cstc2021jscx-gksbX0056);重庆市九龙坡区基础研究与成果转化类科技计划项目(No. 2022-02-003-Z);重庆市中小学创新人才培养工程项目(No. CY230903 & CY230104)

4. School of Microelectronics and Communication Engineering, Chongqing University,
Chongqing 400044, China;)

* Corresponding author, E-mail: coe-fp@cqu.edu.cn; ybq@cqu.edu.cn

Abstract: With the increasing severity of water environmental pollution, there is an urgent need for rapid and accurate detection and identification of organic pollutants in water. Three-dimensional fluorescence spectroscopy technology, which provides rich spectral information about pollutants, has become a hot topic in the research of pollutant identification and source tracing in water bodies. The current methods mainly focus on deep learning based spectral data analysis, which requires a large amount of spectral data and is difficult to promote on site. This paper utilized three-dimensional Excitation-Emission Matrix (3D-EEM) data and proposed a method for multi-classification identification and precise component fitting of water bodies based on a combination of two-dimensional Gabor wavelets and Support Vector Machine (SVM). This method effectively extracted texture features and peak positions of three-dimensional fluorescence spectra, which improved the efficiency of water sample component analysis. Here blank subtraction and Delaunay triangle interpolation were used to reduce background noise and scattering interference in spectral data, and spectral fluctuation interference was suppressed by extending the Savitzky-Golay smoothing approach. Subsequently, texture feature information of 3D-EEM data and global information of three-dimensional fluorescence peaks were extracted using two-dimensional Gabor wavelets and fluorescence peak extraction methods. Finally, an EEM_MSVM model based on MSVC and CF_MSVM was constructed to achieve high-accuracy classification identification and component prediction of water pollutants. Experimental results show that the classification accuracy for water body types is 97.6%. In terms of component prediction, the Root Mean Square Error (RMSE) loss is only 5.3, with a correlation coefficient of 0.94. This effectively achieves accurate classification of typical water bodies and analysis of their components.

Key words: three-dimensional fluorescence spectroscopy; fluorescence component spectra; water body identification; source tracing

1 引 言

水污染和水华现象是全球性的环境问题,严重威胁生态安全和人类健康^[1]。对水体污染物的检测和污染组分的识别解析是污染溯源、环境执法和污染治理的基础,具有重要意义^[2]。

三维荧光光谱技术作为一种具有高灵敏度和选择性的污染物光谱检测方法,在河流水质和污染监测^[3-6],海洋褐潮的污染识别^[7],工业废水中特定污染物^[8]和饮用水处理^[9-10]等领域具有广泛应用。三维激发-发射矩阵(Three-Dimensional Excitation-Emission Matrix, 3D-EEM)是三维荧光光谱法的核心测量数据,它能够提供关于化合物结构和环境性质的丰富信息^[11-15],然而污染物成分复杂,简单的数据处理手段往往难以获得精确的分析结果。随着人工智能的发展,很多研

究者将神经网络应用到三维荧光光谱的数据分析中^[16],侯茂泽^[17]基于 ConvNet 卷积神经网络构建水污染溯源模型并进行溯源,可成功识别实际废水的三维荧光光谱图。Xu 等人^[18]运用深度学习结合平行因子分析建立了荧光识别网络模型,可以预测荧光成分的数量和图谱。Ruan 等人^[19]则基于经典的 VGG 网络进行 3D-EEM 数据的分类和组分拟合。陈庆等人^[20]开展了基于 MobileNet 的水样分类研究,并拟合出荧光组分,获得较好效果。然而基于神经网络的分析方法往往依赖大量的数据,实际使用中各类型水样的采集较为麻烦,也使得该类方法的应用场景受到一定的限制,难以在现场推广。

为了用较少的数据实现水体污染物组分的快速检出,进行高效的水样识别和有机物分析,本文提出一种基于三维荧光光谱数据,利用二维 Gabor

小波结合支持向量机(Support Vector Machine, SVM)进行水样识别及水体组分图谱预测的方法。由于 3D-EEM 数据由激发数据 $\times$ 发射数据 $\times$ 强度数据组成,与灰度图像的数据结构相似,我们将 3D-EEM 数据中每个光谱点视为图像的一个像素。采用二维 Gabor 小波提取 3D-EEM 数据中的纹理信息,解析光谱数据中更为细微的变化和差异,增强对于污染物组分的区分能力。运用荧光峰提取法获取三维荧光峰位的全局信息。最后,基于多分类支持向量机模块(Multi-classification support vector, MSVC)MSVC 和多输出组分拟合模块(Component Fitting Multi-output support vector regression, CF_MSVC)构建了水样识别和组分拟合模型,其中 MSVC 模块负责对水样进行分类,CF_MSVC 模块负责对不同水样荧光组分图进行拟合,从而实现水体污染物的高效分类和组分解析,有利于追踪溯源和样品的在线监测与分析。

2 材料与方法

2.1 水溶液样本采集

本文从南京和重庆两个城市三个区域搜集了三类水样样本,时间跨度 1 年时间。为充分表征水样的代表性,每个采样区域包含 6 个不同的子区域,每个子区域内采样点数大于 30。水样主要包括:①收集于南京市江宁区的城镇污水管网,300 个生活污水(Domestic Wastewater, DW)样本;②收集于南京市江宁区雨水管网,150 个雨水(Rain, RA)样本;③收集于重庆市沙坪坝区民主湖,225 个地表水(Surface Water, SW)样本;④第四类水样样本来自于 drEEM 官方文档中的“Pure Tutorial”数据^[21],包含 60 个三维荧光数据样本(PU)。由于数据量相对较少,通过增加随机波动变量对样本数据进行扩增。包括如下几种方法:向光谱数据中加入随机噪声;改变光谱数据的斜率来模拟光谱受混合有机物波动的变化;乘性变换模拟在不同浓度下荧光信号的强度变化等。经过数据扩增后的样本详细信息如表 1 所示。采集的样本存储于棕色瓶中,冷藏于 4℃ 的冰箱中保存,并在 12 h 内保持低温完成三维荧光光谱测量。

通常而言,本文中所采集的四种不同水体在污染物种类方面具有显著差异:市政生活污水(DW)往往含有较高浓度的有机污染物,雨水

表 1 水溶液样本

Tab. 1 Water solution sample

Types	Label	Number
Domestic Wastewater	DW	1 200
Rain	RA	600
Surface Water	SW	900
Pure Fluorescent Groups	PU	240
Total		2 940

(RA)中往往包含大气沉降或地表径流带来的腐殖质物质,地表水(SW)则包含多种生态来源的有机物质,纯荧光基团(PU)代表了主要与蛋白质有关的有机物。因此,各种水样的荧光光谱特征谱峰、强度等信息截然不同,这也为后续水体识别奠定了基础。

2.2 荧光光谱测量

进行三维荧光光谱测量前,先使用无菌针筒与 0.45 μm 微孔过滤膜对样本进行过滤。使用上海棱光技术有限公司的 F98 荧光光谱仪进行测量,激发波长 Ex 设置为 250~600 nm,步长为 5 nm;发射波长 Em 为 250~650 nm,步长为 1 nm。激发和发射光的狭缝宽度为 1 nm,扫描速度为 12 000 nm/min,增益(PMT)电压设置为 750 V,样本放置于 1 cm $\times$ 1 cm $\times$ 4.5 cm 石英荧光比色皿中,空白水样为超纯水。用荧光光谱仪扫描每一个样本得到对应的 3D-EEM 数据,每个激发-发射矩阵由 71 个激发波长和 401 个发射波长构成,构成 71 $\times$ 401 数据矩阵。

2.3 数据预处理

实际测量得到的三维荧光光谱信号通常会包含大量瑞利散射和拉曼散射^[22-23]等背景噪声,这些噪声强度较大,往往会掩盖待测物质的固有荧光信号,降低检测结果的准确性和可靠性,在正式解算之前需要扣除。本文首先采用空白扣除法,通过减去参比溶液扣除背景噪声和拉曼散射。为了消除样品浓度和仪器因素对光谱强度的影响,进行了拉曼归一化,拉曼归一化公式如下。然后利用 Delaunay 三角形内插法进一步抑制瑞利散射和拉曼散射^[24]:

$$A_{\gamma p} = \int_{\lambda_{em}^1}^{\lambda_{em}^2} I_{\lambda_{em}} d\lambda_{em}, \quad (1)$$

$$F_{\lambda_{ex}, \lambda_{em}}(RU) = \frac{I_{\lambda_{ex}, \lambda_{em}}(AU)}{A_{\gamma p}}, \quad (2)$$

其中: I_λ 是发射波长 λ 处拉曼峰的光谱校正强度, $A_{\gamma p}$ 是激发荧光强度的积分值; λ_{em}^1 和 λ_{em}^2 分别表示发射光谱的波长上下限: 371~428 nm, $\lambda_{ex} = 350$ nm 表示激发荧光的波长是 350 nm, F 是拉曼归一化后的荧光强度, AU 表示任意单位, RU 表示拉曼单位。

除了上述背景噪声的干扰外, 荧光信号还常常伴随着各种随机噪声。这类噪声通常源于探测器的灵敏度、光源的不稳定性等因素, 本文将 Savitzky-Golay (S-G) 滤波方法扩展应用在三维荧光光谱数据中, 以实现随机噪声的抑制^[25]。

Savitzky-Golay 滤波可表示为:

$$Z_{x,y} = \sum_{r=0}^n \sum_{s=0}^{n-r} a_r x^r y^s, \quad (3)$$

其中: x 和 y 分别为横纵坐标, $Z_{x,y}$ 为 (x,y) 处的函

数值, a 为拟合系数。本文采用二维 S-G 平滑对三维荧光光谱进行降噪处理, 多项式参数 n 设为 3, 窗口大小 n, p 均设为 11, 数据预处理前后的效果对比展示在 3.1 节中。

2.4 基于 EEM_MSVM 快速检测模型的构建

本文基于三维荧光光谱数据构建了 EEM_MSVM 模型, 该模型利用二维 Gabor 小波结合支持向量机 (SVM) 处理三维荧光光谱数据 (3D-EEM)。首先对采集水样的 3D-EEM 进行去散射和去噪的数据预处理, 将预处理后的数据通过 PARAFAC 分析, 得出荧光组分数 F_{num} 和荧光组分图谱 Y_{map} 。将荧光组分数 F_{num} 、荧光组分图谱 Y_{map} 和水样类别 W_{cls} 作为前期训练标签, 进行水样分类模型 MSVC 和水样荧光组分拟合模型 CF_MSVM 的训练。EEM_MSVM 模型构建的整体流程图如图 1 所示。

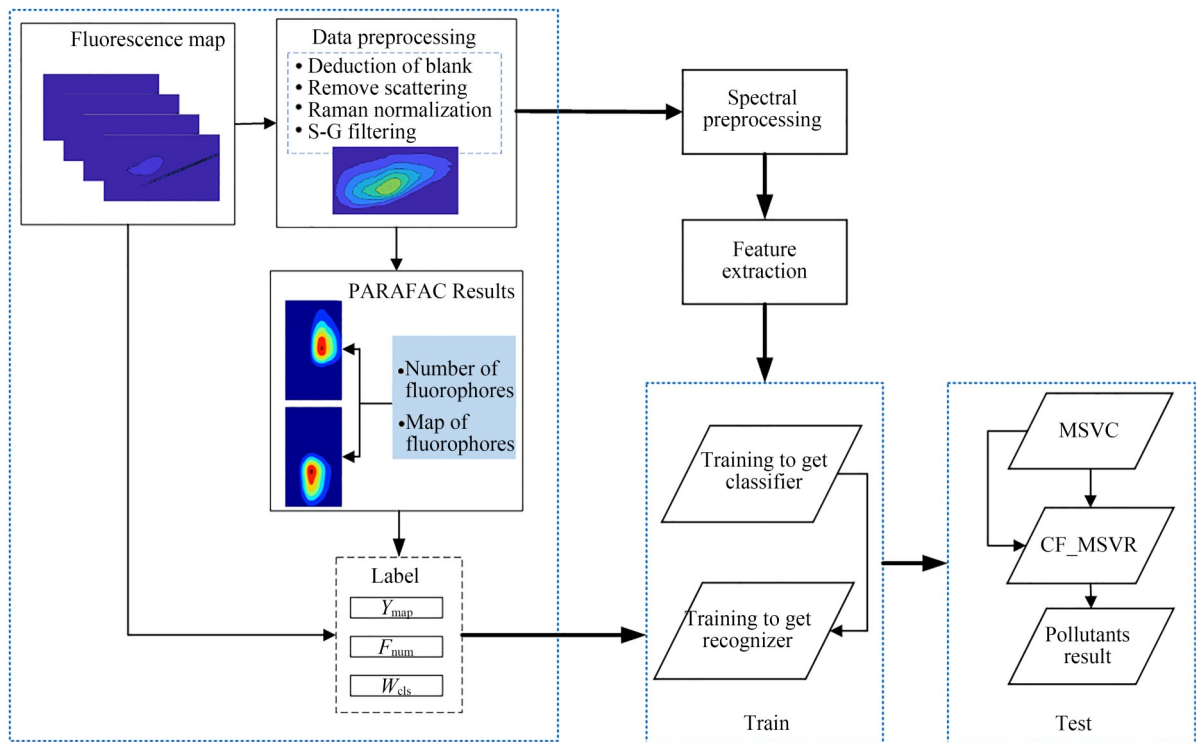


图1 水体有机污染物识别整体流程图

Fig. 1 Water body organic pollutant identification overall process diagram

将 EEM_MSVM 模型训练完成后, 新采集的水样数据经过数据预处理, 利用二维 Gabor 小波提取 3D-EEMs 数据中的纹理信息, 解析光谱数据中更为细微的变化。运用荧光峰提取法获取三维荧光峰位的全局信息。通过上述处理,

EEM_MSVM 模型得到了局部纹理特征和全局峰值特征, 这些特征被合并到一个统一的特征向量中 χ 中, 以便于后续模型使用。在模型的预测阶段, 训练完成的 EEM_MSVM 模型能够迅速且有效地应用其已学习的参数到新数据上, 进行

水样的分类和组分的解析。预测过程比训练阶段的计算量少很多,使得模型能够快速响应新的水样数据,提高了分析的效率,有利于在线分析场景中的应用。最后,将构建好的特征向量 χ

输入到 MSVC 进行水样的分类,并利用 CF_MSVC 对不同水样组分进行拟合,可以得到水样的类别和组分的图谱。整体检测流程如图 2 所示。

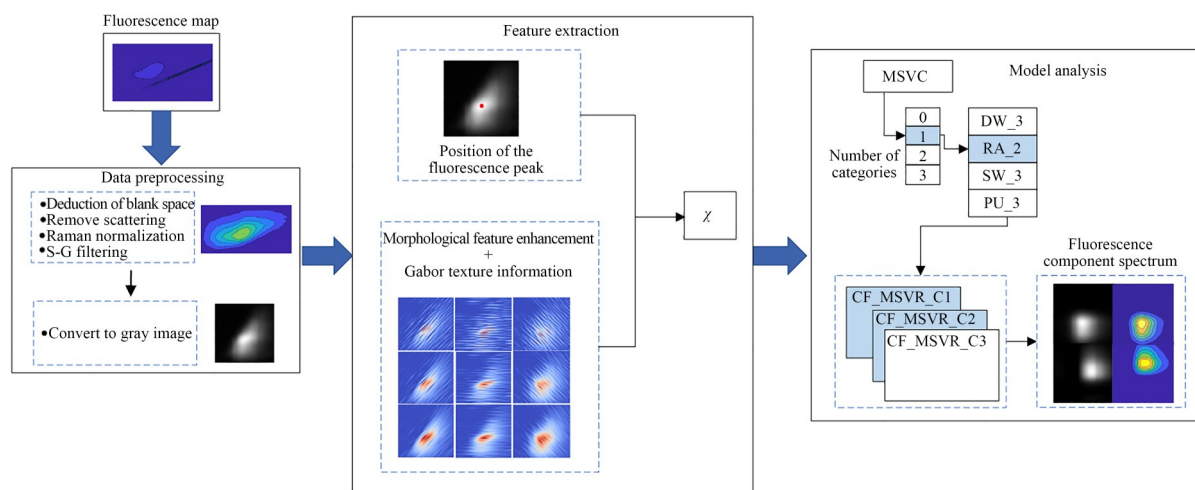


图 2 基于三维荧光光谱的快速检测识别流程

Fig. 2 Rapid detection and identification process based on three-dimensional fluorescence spectroscopy

2.4.1 二维 Gabor 小波

三维荧光光谱包含激发波长、发射波长和荧光强度这三个维度的信息,因此,在分析时可以将其视为图像信息来处理。激发波长和发射波长分别看作图像的两个空间维度,而荧光强度则相当于图像的灰度值,每个像素点对应于一个特定激发波长和发射波长下的荧光强度^[26]。由于二维 Gabor 小波具有良好的空间局部化和方向选择性,能够有效地提取三维荧光光谱中的局部结构和纹理信息,定位和分离光谱中的特征峰,特别是当荧光峰因为结构相似的同类有机物而重叠时,二维 Gabor 小波可以有效分离并定位每一个重叠的峰,并捕获其细微变化。通过准确地提取荧光峰,可以实现对水质中各种有机物质的精确识别和量化,进而对环境中的水质进行准确、高效的监测和分析。二维 Gabor 函数定义如下^[27]:

$$\varphi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|^2}{\sigma^2} - \exp\left(-\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}\right) \cdot \left[\exp(ik_{\mu,v}z) - \exp\left(-\frac{\sigma^2}{2}\right) \right], \quad (4)$$

其中: μ 和 v 分别表示 Gabor 滤波器的方向和尺

度, $z=(x,y)$ 为空间位置, $k_{\mu,v}$ 为波向量,表示为 $k_{\mu,v}=k_v \exp(i\phi_\mu)$,其中, $k_v=\frac{k_{\max}}{f_v}$, $\phi_\mu=\frac{\mu\pi}{8}$, $k_{\max}=\frac{\pi}{2}$ 为最大频率。

2.4.2 荧光峰位信息提取

类比于深度学习中的位置注意力机制^[28],本研究引入了一种基于像素强度权重的多峰关注机制,采用一种基于像素强度的方法来提取荧光峰位置,获得三维荧光峰的全局信息。首先,荧光光谱图像 I 被读取为灰度图,并经过二值化处理以突出潜在的峰值区域。二值化处理定义如下:

$$B(x,y) = \begin{cases} 255, & I(x,y) \geq T \\ 0, & I(x,y) < T \end{cases}, \quad (5)$$

其中: $I(x,y)$ 表示图像在位置 (x,y) 的像素强度, T 是预先设定的阈值, $B(x,y)$ 为二值化后的图像,突出了阈值高于 T 的区域。

对二值化图像进行连通区域分析,每个连通区域 R_k 定义为:

$$R_k = \{(x,y) | B(x,y) = 255 \text{ 且 } (x,y) \text{ 在二值图中相连}\}. \quad (6)$$

对于每个连通区域 R_k , 识别出具有最大像素强度的位置作为峰值位置:

$$(x_{\text{peak}}, y_{\text{peak}})_k = \arg \max_{(x, y) \in R_k} I(x, y). \quad (7)$$

峰值位置的像素强度 $I_{\text{peak}}(k)$ 计算如下:

$$I_{\text{peak}}(k) = I[(x_{\text{peak}}, y_{\text{peak}})_k]. \quad (8)$$

然后将所有峰值位置按照像素强度进行排序, 选取像素强度最大的前 N 个峰值。该多峰关注机制能够专注于输入荧光光谱中的峰值位置, 从而更加准确地获取这些区域的特征和信息。这一方法能够学习到荧光光谱中各个位置的重要性权重, 从而有助于强化模型对于峰值位置的响应和对于非峰值区域的抑制, 而忽略了不重要且无关的背景噪声和扰动。

2.4.3 基于 SVM 有机污染物识别

本研究中, 采用支持向量机 (SVM) [29] 作为核心的分析工具, 对水体 3D-EEM 数据进行有效分类和组分拟合。将提取的纹理特征和全局特征合并为一个特征向量 χ , 作为 SVM 的输入。

采用 SVM 多分类器对水样进行分类, 利用下面公式对模型训练进行优化:

$$\min_{w_k, b_k} \frac{1}{2} \sum_{k=1}^K \|w_k\|^2 + C \sum_{i=1}^N \xi_i, \quad (9)$$

其中: K 表示类别数量, w_k 和 b_k 分别是第 k 个类别的决策边界的法向量和偏置项, ξ_i 是松弛变量, C 是正则化参数。

为了进一步拟合每个类别的不同水样组分, 采用改进的 CF_MSVM (Component Fitting MSVM) 对多组分水样进行拟合。为每个组分创建一个子模型, 针对本文的四类水样, 平行因子分解得到每类水样最大的组分的数量是 3, 因此我们构建了 3 个 CF_MSVM 子模型, CF_MSVM 模

型的数学表述如下:

$$\min_{w_j, b_j} \frac{1}{2} \sum_{j=1}^M \|w_j\|^2 + C \sum_{i=1}^N (\xi_{i,j}^* + \xi_{i,j}), \quad (10)$$

其中: C 是正则化参数, M 表示输出维度, N 表示样本数。 w_j 表示第 j 个输出的权重向量, b_j 表示第 j 个输出的偏置项。 $\xi_{i,j}$ 和 $\xi_{i,j}^*$ 表示第 i 个样本在第 j 个输出上的正负误差的松弛变量。

约束条件为:

$$|Y_{i,j} - w_j^T \phi(\chi_{i,j}) - b_j| \leq \epsilon + \xi_{i,j}, \xi_{i,j} \geq 0. \quad (11)$$

CF_MSVM 模型采用高斯径向基函数 (Radial Basis Function, RBF) 作核函数:

$$\phi(\chi_{i,j}) = \exp(-\gamma \|x_i - x_j\|^2). \quad (12)$$

CF_MSVM 模型保留了传统 SVM 在处理非线性问题时的强大能力, 并针对每个组分建立子模型, 可以处理具有多种有机污染物的复杂水样。

3 结果和讨论

3.1 数据集的建立

将采集的三维荧光光谱数据按前述 2.3 节要求进行预处理, 使用 F98 荧光光谱仪扫描的样品显示了瑞利和拉曼散射。通过去除离子水的光谱数据, 消除了背景干扰和拉曼散射, 并进行了拉曼归一化, 采用 Delaunay 三角形内插值法消除了瑞利散射和拉曼散射。对散射消除前后的光谱进行比较表明, 污染物的特征信息在无散射的谱上比在有散射的谱上更明显, 更直观地反映了污染物的特征峰。通过对光谱进行 S-G 滤波, 消除了光谱的抖动, 使边缘更平缓, 预处理前后的三维荧光对比谱图如图 3 所示。原始三维荧光

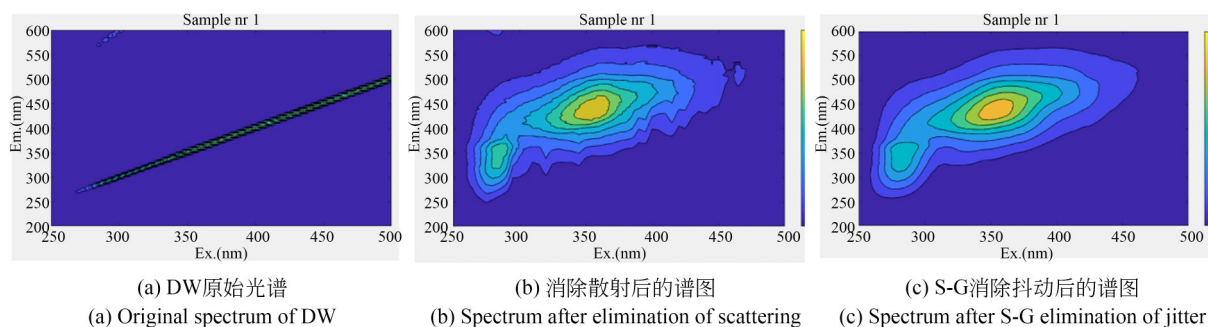


图 3 样品 DW 预处理前后的三维荧光光谱

Fig. 3 Three-dimensional fluorescence spectra of sample dw before and after pre-treatment

数据通过预处理生成用于平行因子分析和 EEM_MSVM 模型训练的规范数据集。通过平行因子分析得到荧光组分数量和荧光组分图谱,作为前期训练数据的标签,结合水样类别标签,进行水样荧光组分拟合模型 CF_MSVM 和水样分类模型 MSVC 的训练。不难看出,预处理对于增强荧光光谱特征效果明显,这为后续平行因子分析和模型训练奠定了基础。但目前预处理步骤较为繁琐,不仅需要利用 Delaunay 三角形内插值法消除瑞利散射和拉曼散射的影响,还需要借助 S-G 滤波改善光谱形状,凸显光谱特征峰。如何简化预处理步骤和增强预处理算法的泛化性是下一步方法走向实用性亟待解决的问题。

3.2 EEM_MSVM 分析结果

在该研究中,采用的 Gabor 滤波器组包含 4 个尺度 $\nu = \{10, 13, 15, 18\}$ 和 5 个方 $\mu = \{0, \pi/8, 3\pi/8, 5\pi/8, 7\pi/8\}$ 。这些不同的尺度和方向互相组合,生成了 20 个特定的滤波输出。

本文引入了一种基于像素权重的多峰关注机制^[28],利用此方法对荧光峰进行了提取,如图 8 所示。从图 4 中可以观察到 DW 具有两个荧光峰位,而 RA 和 SW 各具有一个荧光峰位,PU 则显示出三个荧光峰位。这些观测结果与上文的分析结果相互验证,进一步确保了本研究方法的准确性与可靠性。

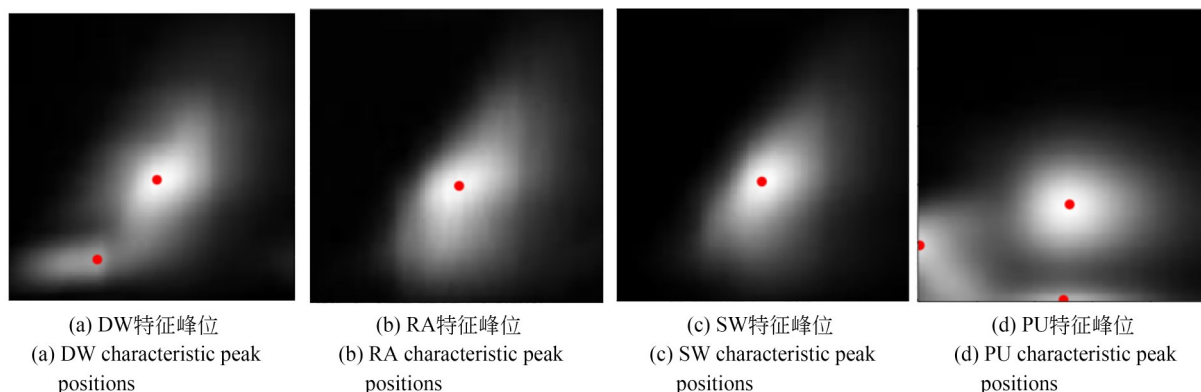


图 4 不同样本的特征峰位的提取

Fig. 4 Extraction of characteristic peak positions for different samples

3.3 EEM_MSVM 模型的性能

将提取的纹理特征和全局特征合并为一个特征向量 χ , 作为 EEM_MSVM 的输入。将特征向量 χ 输入支持向量机多分类器 (MSVC) 中进行建模。按 7:3 的比例将数据集划分为训练集和测试集,详细的训练过程及其混淆矩阵如图 5 所示。为了评估 MSVC 的性能,选择了分类准确率作为评价指标。由图中不难看出,随着训练样本量的增加,分类准确性逐渐提升,并趋于收敛,最终达到 97.6% 的总体分类准确度,表明 MSVC 具有良好的分类效果。同时,也为用于训练的样本数量确定了一个上限。

在 MSVC 的基础上,本文进一步开展了组分拟合模型的搭建。本文中 EEM_MSVM 组分拟合模型由三个 CF_MSVM 子模型组成,将 PARAFAC 分析结果作为训练标签,利用 3D-

EEM 数据训练得到每个子模型。为了评估 CF_MSVM 模型对组分图谱的拟合能力,采取了十折交叉验证策略。在模型评价指标方面,选取均方根误差来衡量模型的损失值,均方根误差越小,相关系数越接近于 1,说明模型预测结果越趋近于真实值。本文模型的均方根误差为 5.3,相关系数达到 0.94,说明在有限的样本下,本文模型能够有效实现对组分图谱的拟合,但仍有进一步提升的空间。这一系列实验数据充分证明了 CF_MSVM 模型在组分图谱拟合方面的出色表现与高度可靠性。

$$R = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)} \cdot \sqrt{\text{Var}(y)}}, \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N [y_i - x_i]^2}, \quad (14)$$

其中: R 表示相关系数, RMSE 表示均方误差,

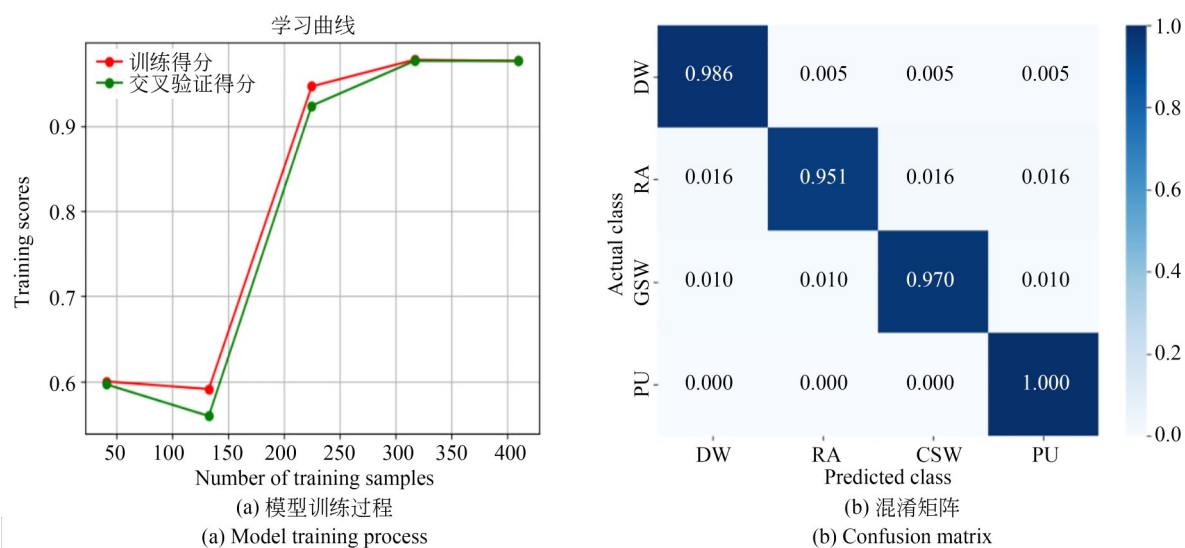


图 5 EEM_MSVM 分类模型结果

Fig. 5 EEM_MSVM classification model results

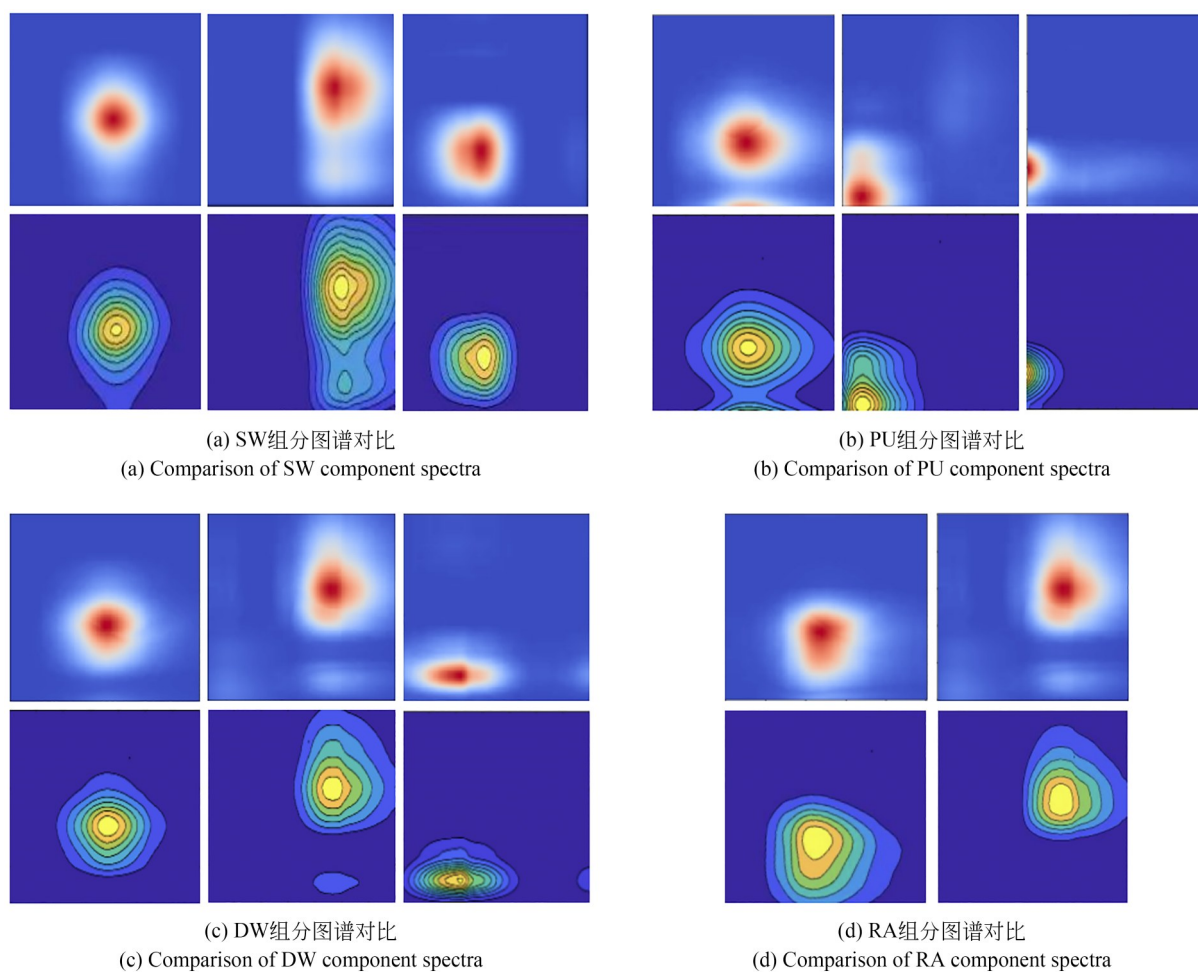


图 6 EEM_MSVM 模型拟合组分图和 PARAFAC 得到组分图的对比

Fig. 6 Comparison of component maps fitted by EEM_MSVM model and component maps obtained by PARAFAC

$\text{Var}(\cdot)$ 表示计算待测光谱的方差, $\text{Cov}(\cdot)$ 表示计算原信号(x)与目标信号(y)之间的协方差。

为了进一步证明经训练后的EEM_MSVM模型在分类和拟合方面的优越性能,选取了部分未训练的3D-EEM数据作为测试样本。CF_MSVM模型得到的拟合组分与PARAFAC的结果进行对比。如图6所展示,每个子图第一行CF_MSVM行是本文模型拟合出的组分图谱,每个子图第二行PARAFAC行则是由平行因子分析所得出的结果。对比发现,CF_MSVM模型的拟合效果与PARAFAC分析所得结果具有显著的相似性。因此EEM_MSVM模型可以很好地对水样的3D-EEM进行识别和拟合。

4 结 论

针对水污染检测耗时复杂,不利于水样在线监测与分析等问题,本文提出了一种基于EEM_MSVM快速检测模型,用于不同水样的识别和组分图谱的拟合。该方法将三维荧光光谱数据作为图像进行处理,首先采用二维Gabor小波提取每个样本的纹理特征,在此基础上,我们

通过提取荧光峰的位置信息,引入全局特征。基于改进的SVM,构建了EEM_MSVM模型。实验结果表明,本文构建的水样快速识别模型,在模型训练完成后,能够实现水样类别预测,并且拟合出水体荧光组分图谱。本文虽然在多个区域收集了3种水样,但并未细分讨论具体位置、季节、污染程度、复合污染物等参数对荧光数据造成的影响,这降低了本文方法的泛化性和鲁棒性。未来,可进一步加强复合污染物的荧光光谱识别研究,探讨不同污染物荧光光谱之间的相互影响以及相关识别算法的进一步改进,为该方法进一步走向生产应用奠定基础。

作者贡献声明:

姜赞成:论文构思,实验验证,论文初稿写作;

王瑞杰:论文构思,方法设计,数据采集及初稿写作;

项晓亮:数据分析及稿件修改;

叶彬强:方案设计,实验设计及论文修改;

冯鹏:论文构思,方案设计,实验数据分析及论文修改。

参考文献:

- [1] 刘文清,陈臻懿,刘建国,等. 环境监测领域中光谱学技术进展[J]. 光学学报, 2020, 40(5): 7-14.
LIU W Q, CHEN Z Y, LIU J G, *et al.* Advances with respect to the environmental spectroscopy monitoring technology[J]. *Acta Optica Sinica*, 2020, 40(5): 7-14. (in Chinese)
- [2] 最高人民检察院. 生态环境和资源保护检察白皮书(2018-2022). [R/OL]. (2023-06-05) https://www.spp.gov.cn/xwfbh/wsfbh/202306/t20230605_616291.shtml
The Supreme People's Procuratorate. *White Paper on Ecological Environment and Resource Protection Inspection* [R/OL]. (2023-06-05) https://www.spp.gov.cn/xwfbh/wsfbh/202306/t20230605_616291.shtml. (in Chinese)
- [3] HYNES N R J, KUMAR J S, KAMYAB H, *et al.* Modern enabling techniques and adsorbents based dye removal with sustainability concerns in textile industrial sector-a comprehensive review[J]. *Journal of Cleaner Production*, 2020, 272: 122636.
- [4] 贾文坤,张恒之,马洁,等. 基于紫外-可见光谱与深度学习CNN算法的水质COD预测模型研究[J]. 光谱学与光谱分析, 2020, 40(9): 2981-2988.
JIA W S, ZHANG H Z, MA J, *et al.* Study on the predication modeling of COD for water based on UV-VIS spectroscopy and CNN algorithm of deep learning [J]. *Spectroscopy and Spectral Analysis*, 2020, 40(9): 2981-2988. (in Chinese)
- [5] 尹航,俞巧君,侯迪波,等. 基于监督学习的紫外-可见光光谱水质在线异常检测方法研究[J]. 光谱学与光谱分析, 2019, 39(2): 491.
YIN H, YU Q J, HOU D B, *et al.* In-situ detection of water quality anomaly with UV/vis spectrum based on supervised learning [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(2): 491. (in Chinese)
- [6] 杨欣,吴支行,叶寅,等. 店埠河农业小流域水体溶解性有机质三维荧光光谱的平行因子分析[J]. 光谱学与光谱分析, 2022, 42(3): 978-983.
YANG X, WU Z H, YE Y, *et al.* Parallel factor

- analysis of fluorescence excitation emission matrix spectroscopy of DOM in waters of agricultural watershed of dianbu river[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(3): 978-983. (in Chinese)
- [7] 陈颖, 段玮靓, 杨英, 等. 基于三维荧光光谱和 GBDT-LR 的褐潮藻辨识[J]. 光学学报, 2022, 42(12): 1230001.
- CHEN Y, DUAN W L, YANG Y, *et al.* Identification of brown tide algae based on three-dimensional fluorescence spectra and GBDT-LR[J]. *Acta Optica Sinica*, 2022, 42(12): 1230001. (in Chinese)
- [8] 刘传晔, 柴一荻, 徐宪根, 等. 南方某河水水质荧光指纹特征及污染溯源[J]. 光谱学与光谱分析, 2021, 41(7): 2142-2147.
- LIU C Y, CHAI Y D, XU X G, *et al.* Aqueous fluorescence fingerprint characteristics and discharge source identification of a river in Southern China[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(7): 2142-2147. (in Chinese)
- [9] 钟婷婷, 林涛, 刘威. 饮用水处理过程中全氟化化合物的分布、转化及去向[J]. 环境科学, 2023, 44(5): 2613-2621.
- ZHONG T T, LIN T, LIU W. Distribution, transformation, and fate of per-and polyfluoroalkyl substances in drinking water treatment[J]. *Environmental Science*, 2023, 44(5): 2613-2621. (in Chinese)
- [10] SHI W X, ZHUANG W E, HUR J, *et al.* Monitoring dissolved organic matter in wastewater and drinking water treatments using spectroscopic analysis and ultra-high resolution mass spectrometry[J]. *Water Research*, 2021, 188: 116406.
- [11] CHEN Z Y, LI S Y. Seasonal variation of DOM spectral characteristics of rivers with different urbanization levels in the Three Gorges Reservoir area[J]. *Huan Jing Ke Xue*, 2021, 42(1): 195-203.
- [12] CHENG C, WU J, YOU L D, *et al.* Novel insights into variation of dissolved organic matter during textile wastewater treatment by fluorescence excitation emission matrix[J]. *Chemical Engineering Journal*, 2018, 335: 13-21.
- [13] CIFUENTES-CABEZAS M, GALINHA C F, CRESPO J G, *et al.* Nanofiltration of wastewaters from olive oil production: Study of operating conditions and analysis of fouling by 2D fluorescence and FTIR spectroscopy[J]. *Chemical Engineering Journal*, 2023, 454: 140025.
- [14] 高贝贝, 王欣, 胡术刚, 等. 畜禽养殖废水的三维荧光特征[J]. 环境监测管理与技术, 2022, 34(1): 60-63.
- GAO B B, WANG X, HU S G, *et al.* 3-D fluorescence characteristics of livestock and poultry breeding wastewater[J]. *The Administration and Technique of Environmental Monitoring*, 2022, 34(1): 60-63. (in Chinese)
- [15] NIE J X, YAN S W, LIAN L S, *et al.* Development of fluorescence surrogates to predict the ferate(VI) oxidation of pharmaceuticals in wastewater effluents[J]. *Water Research*, 2020, 185: 116256.
- [16] PELEATO N M, LEGGE R L, ANDREWS R C. Neural networks for dimensionality reduction of fluorescence spectra and prediction of drinking water disinfection by-products[J]. *Water Research*, 2018, 136: 84-94.
- [17] 侯茂泽, 马艳琼, 田森林, 等. 基于卷积神经网络识别三维荧光光谱的水污染溯源研究[J]. 中国环境监测, 2022, 38(5): 188-195.
- HOU M Z, MA Y Q, TIAN S L, *et al.* Research on water pollution traceability based on convolutional neural network identification of three-dimensional fluorescence spectrum[J]. *Environmental Monitoring in China*, 2022, 38(5): 188-195. (in Chinese)
- [18] XU R Z, CAO J S, FENG G Y, *et al.* Fast identification of fluorescent components in three-dimensional excitation-emission matrix fluorescence spectra via deep learning[J]. *Chemical Engineering Journal*, 2022, 430: 132893.
- [19] RUAN K, ZHAO S, JIANG X Q, *et al.* A 3D fluorescence classification and component prediction method based on VGG convolutional neural network and PARAFAC analysis method[J]. *Applied Sciences*, 2022, 12(10): 4886.
- [20] 陈庆, 汤斌, 缪俊锋, 等. 基于荧光光谱的水体分类与荧光组分识别方法[J]. 光学学报, 2023, 43(6): 326-336.
- CHEN Q, TANG B, MIAO J F, *et al.* Water sample classification and fluorescence component identification based on fluorescence spectrum[J]. *Acta Optica Sinica*, 2023, 43(6): 326-336. (in Chinese)
- [21] MURPHY K R, STEDMON C A, GRAEBER D, *et al.* Fluorescence spectroscopy and multi-way techniques. PARAFAC[J]. *Analytical Methods*,

- 2013, 5(23): 6557-6566.
- [22] YUAN Y Y, WANG X J, LIU X Y, *et al.* Novel automatic de-scattering method for three-dimensional fluorescence spectra based on deep learning [J]. *Microchemical Journal*, 2024, 204: 111145.
- [23] DIERKER S B, ARONSON M C. Reduction of Raman scattering and fluorescence from anvils in high pressure Raman scattering[J]. *The Review of Scientific Instruments*, 2018, 89(5): 053902.
- [24] 伍卓慧, 黄冰佳, 李学勤, 等. 一种自适应的三维荧光光谱去散射方法[J]. *光谱学与光谱分析*, 2024, 44(4): 969-976.
- WU Z H, HUANG B J, LI X Q, *et al.* A self-adapting method for removing scatterings in the excitation-emission matrix spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(4): 969-976. (in Chinese)
- [25] HUANG Y H, YANG R X, GENG X, *et al.* Two filters for acquiring the profiles from images obtained from weak-light background, fluorescence microscope, transmission electron microscope, and near-infrared camera[J]. *Sensors*, 2023, 23(13): 6207.
- [26] ATTALLAH O. Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and Gabor wavelets[J]. *Scientific Reports*, 2024, 14: 20637.
- [27] 陈柱, 吴帮玉, 邱辰禹, 等. 基于时间权重损失函数与注意力机制的卷积神经网络速度谱自动拾取方法[C]. 中国石油学会 2021 年物探技术研讨会论文集. 成都, 2021: 1042-1045.
- CHEN Z, WU B Y, QIU C Y, *et al.* Automatic velocity spectrum picking method for convolutional neural networks based on time-weighted loss function and attention mechanism [C]. *Proceedings of the 2021 Geophysical Exploration Technology Seminar of the Chinese Petroleum Society*. Chengdu, China, 2024, 14: 20637.
- [28] CHEN H Z, XU L L, AI W, *et al.* Kernel functions embedded in support vector machine learning models for rapid water pollution assessment via near-infrared spectroscopy[J]. *Science of the Total Environment*, 2020, 714: 136765.

作者简介:



姜赞成(1986—),男,四川广元人,博士研究生,高级工程师,2006年于成都理工大学获得学士学位,2009年于成都理工大学获得硕士学位,2022年于重庆大学攻读博士学位。主要方向为环境水质光谱检/监测研究以及产品研制。E-mail: agreenginger@qq.com

通讯作者:



叶彬强(1981—),浙江东阳人,博士研究生,讲师,2022年于重庆大学攻读博士学位。研究方向为环境水质光谱检测研究。E-mail: ybq@cqut.edu.cn

通讯作者:



冯 鹏(1981—),男,河南商城人,博士,教授,博士生导师。2007年于重庆大学获博士学位,主要研究方向为环境监测中的光谱信号检测及相关设备研制。E-mail: coe-fp@cqu.edu.cn