

基于紫外-荧光多光谱融合的水质化学需氧量检测

周昆鹏¹, 白旭芳¹, 毕卫红^{2*}

¹内蒙古民族大学物理与电子信息学院, 内蒙古 通辽 028000;

²燕山大学信息科学与工程学院, 河北省特种光纤与光纤传感重点实验室, 河北 秦皇岛 066004

摘要 提出了一种基于紫外和荧光多光谱融合的水质化学需氧量(COD)的检测方法。实验样本为包含近岸海水和地表水在内的 53 份水样, 采用标准化学方法获取样本的 COD 理化值, 利用紫外-可见光谱仪和荧光分光光度计采集样品的紫外吸收光谱和三维荧光光谱, 对光谱数据进行处理后建模。采用蚁群-区间偏最小二乘法(ACO-iPLS)作为特征提取算法, 采用粒子群优化的最小二乘支持向量机(PSO-LSSVM)算法作为建模方法, 分别建立基于紫外吸收光谱数据和单激发波长下的荧光发射光谱数据的预测模型, 以及紫外-荧光多光谱数据级融合模型和特征级融合模型, 并对各类模型的预测效果进行对比。结果表明: 基于紫外-荧光多光谱特征级融合模型的预测效果最优, 该模型预测水质 COD 的精度更高, 其校正集决定系数为 0.9999, 检验集的预测决定系数为 0.9912, 外部检验均方根误差为 1.1297 mg/L。本研究为水质 COD 的快速检测提供了一种新的研究思路和解决方法。

关键词 光谱学; 紫外吸收光谱; 荧光发射光谱; 多光谱融合; 化学需氧量; 水质检测; 中等水平数据融合

中图分类号 O433.4; O657.3

文献标识码 A

doi: 10.3788/LOP55.113003

Detection of Chemical Oxygen Demand in Water Based on Multi-Spectral Fusion of Ultraviolet and Fluorescence

Zhou Kunpeng¹, Bai Xufang¹, Bi Weihong^{2*}

¹College of Physics and Electronic Information, Inner Mongolia University for Nationalities, Tongliao, Inner Mongolia 028000, China;

²Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

Abstract A method for detecting chemical oxygen demand (COD) in water based on ultraviolet and fluorescence multi-spectral fusion is proposed. The experimental samples are 53 actual water samples, including coastal seawater and surface water. The physicochemical values of the experimental samples are calculated by the standard chemical method, and the ultraviolet absorption spectra of the samples are collected by the ultraviolet-visible spectrometer, and the three-dimensional fluorescence spectra are collected by fluorescence spectrophotometer, then the processed spectral data are used to build model. Using the ant colony-interval partial least squares(ACO-iPLS) as feature extraction algorithm and the particle swarm optimization least squares support vector machine(PSO-LSSVM) as modeling method, we establish the prediction model based on ultraviolet absorption spectra and fluorescence emission spectra at single excitation wavelength, the data level fusion model and the feature level fusion (mid-level data fusion, MLDF) model based on ultraviolet and fluorescence multi-spectral information, respectively. And the prediction results of various models are compared. The results show that the prediction effect of the MLDF model based on ultraviolet and fluorescence multi-spectral information is optimal, and the prediction accuracy of COD in water is relatively high. The determination coefficient of calibration set is 0.9999, the prediction determination coefficient of validation set is 0.9912, and the root mean square error in prediction set is 1.1297 mg/L. It provides a new research idea and solution for the rapid detection of COD in water.

Key words spectroscopy; ultraviolet absorption spectrum; fluorescence emission spectrum; multi-spectral fusion;

收稿日期: 2018-04-24; 修回日期: 2018-05-28; 录用日期: 2018-06-07

基金项目: 国家重点研发计划项目(2017YFC1403800)、河北省重点研发计划项目(18273302D)、内蒙古民族大学博士科研启动基金(BS432)、内蒙古民族大学科研项目(NMDYB17162)

* E-mail: bwhong@ysu.edu.cn

chemical oxygen demand; water quality detection; mid-level data fusion

OCIS codes 300.6170; 300.6280; 300.6540

1 引言

化学需氧量(COD)是水质检测的常规检测指标之一,它反映了水体受还原性物质污染的程度。目前,测定水质 COD 的方法有电化学法^[1]、快速消解分光光度法^[2]、光催化氧化法^[3]、臭氧氧化法^[4]和光谱测量法^[5]等。我国测定水质 COD 的标准方法属于湿化学法,具有测定结果相对可靠、重现性好的特点,但检验时间长,且易造成二次污染。

近年来,光谱分析方法已在食品^[6]、医药卫生^[7-8]、生物^[9]及环境安全监测^[10]等领域得到广泛应用。目前,基于光谱技术的水质 COD 的检测方法主要是紫外-可见(UV-Vis)吸收光谱法^[11]。曾有研究者采用近红外(NIR)吸收光谱法^[12]对水质 COD 进行了检测,但水中有机物在近红外光谱区的吸收强度相对较弱、光谱信噪比较低等因素使得近红外吸收光谱法分析水质的研究基本上还处于实验室研究阶段。荧光光谱法可以实现对水质 COD 的检测^[13-14],但目前研究者主要是利用荧光光谱法检测水体中的溶解有机物(DOM)^[15-16]。此外,也有研究者采用紫外和近红外光谱融合的方法对水质 COD 进行检测^[17-18],但检测效果还有待进一步提高。

采用紫外吸收光谱法检测水中有机物时,若有有机物的浓度高,检测灵敏度就比较高;若有有机物的浓度低,检测误差就会变大。采用荧光检测低浓度有机物时的检测精度很高,当水中有机物的浓度过高时则可能会引起荧光猝灭,从而影响检测精度。因此,从紫外光谱和荧光光谱的特点出发,基于以上分析,本文联合紫外光谱法和荧光光谱法检测水中有机污染物的优点,采用信息融合理论,利用水样的紫外吸收光谱和荧光发射光谱数据对水质 COD 进行建模分析;基于化学计量学方法,分别建立紫外吸收光谱数据模型、荧光发射光谱数据模型以及紫外-荧光多光谱数据级融合模型和特征级融合模型,并对所建模型的预测效果进行对比,进而建立基于紫外-荧光多光谱融合的水质 COD 的检测方法。

2 实验部分

2.1 仪器与样品

2.1.1 实验仪器设备

采用荷兰 Avantes 公司的 AvaSpec-2048 型紫

外-可见光纤光谱仪完成紫外吸收光谱的采集工作,采用 10 mm×10 mm 的石英比色皿作为样品池;采用日本 Hitachi 公司的 F-7000 型荧光分光光度计采集荧光光谱数据,采用 10 mm×10 mm 石英比色皿作为样品池。

2.1.2 实验样本

本研究以实际的水样为实验样本,实验样本分别取自近岸海水和地表水,共计 53 份。其中在河北省秦皇岛市的渤海海域分批次采集包括入海口和近岸海域在内的海水样本 22 份,在秦皇岛市的大汤河、小汤河、污水河、森林公园内河、森林公园人工湖、燕山大学燕鸣湖等地分批次采集地表水样本 31 份。

2.2 光谱采集与水质 COD 理化值测量

在实验室条件下,将采集的所有水样静置 30 min 后取上层液体,然后分别进行紫外吸收光谱和三维荧光光谱的采集,以及水质 COD 理化值的测量工作。

2.2.1 水样的紫外吸收光谱

采集紫外吸收光谱时设置紫外可见光谱仪的积分时间为 2 ms,采样次数设置为 200,设置光谱扫描范围为 200~400 nm。全部实验水样的紫外吸收光谱如图 1 所示。

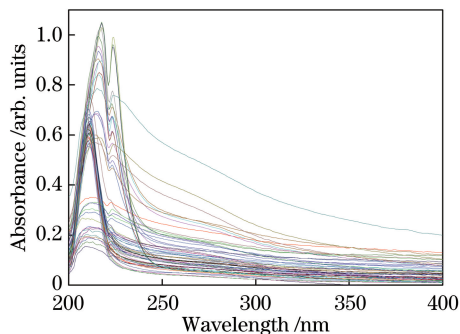


图 1 实验水样的紫外吸收光谱

Fig. 1 UV absorption spectra of experimental water samples

2.2.2 水样的三维荧光光谱

采集水样的三维荧光光谱时,设置 F-7000 荧光光谱仪的 PMT 电压为 700 V,激发光和发射光的狭缝宽度均为 5 nm,扫描速度为 12000 nm/min,激发波长间隔为 5 nm,发射波长间隔为 2 nm。样品的荧光光谱经 3 次采样后求平均值,其中激发波长 E_x 为 200~400 nm,发射波长 E_m 为 220~600 nm。由于水的三维荧光光谱中含有的瑞利散射和拉曼散射

的强度较大,会对荧光信息产生较大影响,因此采用 Delaunay 三角内插值法^[19]去除三维荧光光谱采集过程中的散射峰。某海水水样三维荧光光谱的散射

峰在去除前后的效果图分别如图 2(a)和图 2(b)所示。图 2 表明, Delaunay 三角内插值法可以有效去除散射峰的干扰。

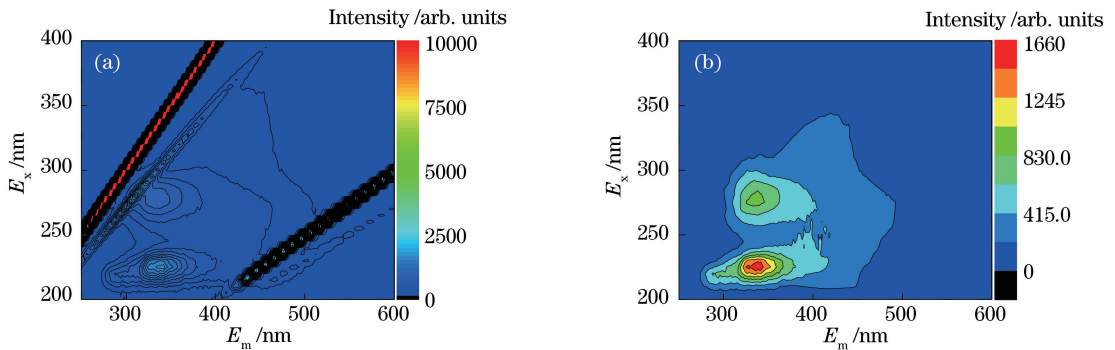


图 2 海水水样去散射前后的等高线光谱图。(a)去散射前;(b)去散射后

Fig. 2 Contour maps before and after removal of scattering of sea water samples.

(a) Before removal of scattering; (b) after removal of scattering

2.2.3 水质 COD 理化值的测量

海水样本水质 COD 的测量采用碱性高锰酸钾滴定法,该方法属于湿化学法,是我国国家标准规定的检测海水水质 COD 的方法。地表水的 COD 测量采用快速消解分光光度法,所用设备为美国 HACH 公司的 DRB200 型消解器和 DR6000 型分光光度计。消解地表水实验样品时,消解器温度设置为 150 ℃,消解时间为 120 min。

2.3 光谱数据的处理及建模方法

2.3.1 光谱数据的处理方法

由于实际水样的成分复杂,采集到的紫外吸收光谱曲线可能会因水体中复杂成分的影响而发生改变,如基线漂移、附加噪声等。因此需要对光谱数据进行预处理,本课题组采用的光谱预处理方法为 S-G 平滑、消除常数偏移量。S-G 平滑是一种常用的数据平滑处理方法,在进行平滑处理时使用最小二乘拟合对移动窗口内的光谱数据进行最小二乘多项式拟合。该方法的平滑过程也可以看作是一种加权平均的过程,原始光谱经过 SG 平滑后,光谱的形状不发生改变,而光谱曲线的噪声却明显降低(信噪比提高)。消除常数偏移量是将光谱平移,使每一列元素的最小值为零,经消除常数偏移量变换后,光谱数据矩阵的最小值为零。

荧光光谱数据的处理:根据采集到的三维荧光光谱图进一步确定有效发射波长范围为 250~500 nm,然后将三维荧光光谱数据按激发波长展开,激发波长范围为 200~330 nm,间隔 5 nm,则每组样本的荧光光谱数据可展开成 3402 维的向量。分别选取各激发波长下的荧光发射光谱数据进行特

征提取与模型建立。

2.3.2 模型的建立方法

将 53 份实验水样中的 37 份作为校正集数据,进行校正模型的建立;其余 16 份作为检验集数据,对所建模型的预测性能进行检验,其中海水和地表水数据各 8 份。对实际水样的紫外光谱和荧光光谱数据进行建模时,采用基于蚁群算法优化的间隔偏最小二乘法(ACO-iPLS)对光谱数据进行特征提取,采用基于粒子群优化的最小二乘支持向量机(PSO-LSSVM)算法进行模型的建立。分别建立基于紫外吸收光谱数据的预测模型、基于荧光发射光谱数据的预测模型、基于紫外-荧光多光谱数据级融合的预测模型以及基于紫外-荧光多光谱特征级融合的预测模型。将模型的效果进行比较,确定最优的模型为基于紫外-荧光多光谱特征级融合的预测模型。本研究得到的最优模型的建模过程如图 3 所示。

特征提取和模型建立的过程具体描述如下。

1) 设水样的紫外吸收光谱数据 x_{UV} 和三维荧光光谱数据 x_F 分别为

$$x_{UV} = \{x_{UV,1}, x_{UV,2}, \dots, x_{UV,s_1}\}, \quad (1)$$

$$x_F = \{x_{F,1}, x_{F,2}, \dots, x_{F,s_2}\}, \quad (2)$$

式中: x_{UV,s_1} 为第 s_1 个水样的紫外吸收光谱数据; x_{F,s_2} 为第 s_1 个水样的荧光光谱数。

2) 特征提取后得到的紫外特征光谱信息 z_{UV} 和荧光特征光谱信息 z_F 分别为

$$z_{UV} = \{z_{UV,1}, z_{UV,2}, \dots, z_{UV,n_1}\}, \quad (3)$$

$$z_F = \{z_{F,1}, z_{F,2}, \dots, z_{F,n_2}\}, \quad (4)$$

式中: n_1 和 n_2 分别为特征提取后紫外吸收光谱和

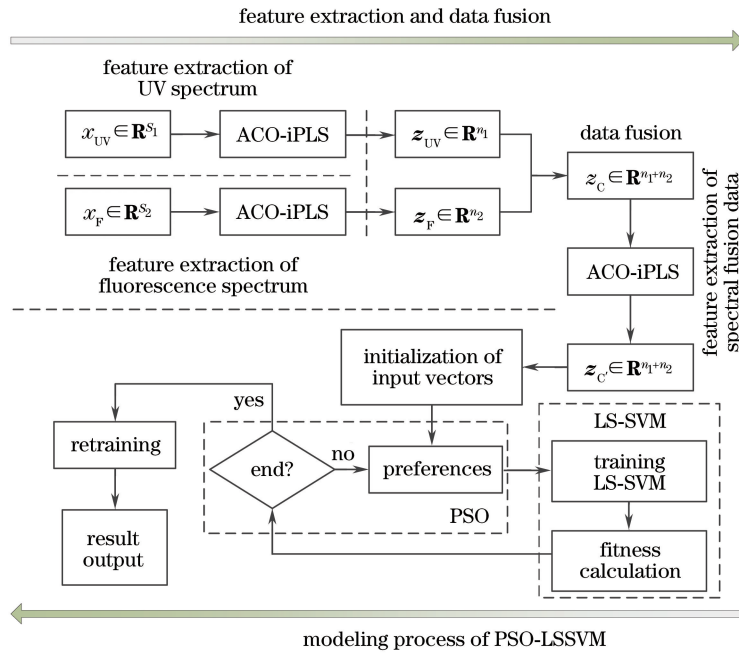


图3 本研究最优模型的建模过程

Fig. 3 Modeling process of the optimal model in this study

荧光光谱的数量。

3) 将两种光谱的特征数据联合构成向量 z_C , 再次利用 ACO-iPLS 对 z_C 进行特征提取, 得到信息融合后的输入向量 $z_{C'}$, 即

$$z_{C'} = [z_{UV}, z_F]^T \in \mathbf{R}^{n_1+n_2}. \quad (5)$$

4) 将 $z_{C'}$ 作为 PSO-LSSVM 的输入向量进行建模。

在本研究最终采用的建模方法中, 多次采用 ACO-iPLS 对光谱信息进行特征提取, ACO 算法属于全局搜索算法, 可以有效避免局部最优解的情况出现, 因此优选出的特征波长更具代表性。以 ACO-iPLS 的优选波长对应的光谱数据作为输入, 并采用 PSO 优化算法对 LSSVM 方法的关键参数进行寻优, 使最终模型的学习能力及泛化能力得到保证。本研究将 ACO-iPLS 算法和 PSO-LSSVM 算法联合应用, 也使得最终的模型具有更好的稳健性和预测能力。

2.3.3 模型的评价指标

采用校正集样本的决定系数 (R^2)、交叉验证均

方根误差 (RMSECV) 作为模型的评价指标, 采用验证集样本的外部检验均方根误差 (RMSEP) 来评价模型的预测精度。一般来说, R^2 值越大, RMSECV 和 RMSEP 值越小, 模型的预测效果越好。

3 结果与讨论

3.1 单谱源模型的建立

3.1.1 紫外吸收光谱数据建模

利用 PSO-LSSVM 算法建模时, 将 ACO-iPLS 算法提取的最优特征谱段作为 LSSVM 算法的输入, 其中 ACO-iPLS 的蚂蚁数量设定为 2000, 主成分数量为 5, 选择变量数目为 10; PSO 优化算法的 c_1 、 c_2 分别设定为 1.5 和 1.7, 种群规模为 30, 进化次数为 300。经预处理和特征提取后紫外吸收光谱数据的 PSO-LSSVM 建模结果如表 1 所示。其中, R_c^2 为校正集的决定系数, R_p^2 为检验集的决定系数。

由表 1 可以看出, 利用 ACO-iPLS 算法作为特征提取方法对 S-G 平滑和消除偏移量处理的光谱

表 1 实验水样紫外光谱数据的 PSO-LSSVM 建模结果

Table 1 Modeling results of UV spectra of experimental water samples by PSO-LSSVM

Feature extraction method	Optimal preprocessing method	Preferred band /nm	Calibration set		Validation set	
			R_c^2	RMSECV / (mg·L ⁻¹)	R_p^2	RMSEP / (mg·L ⁻¹)
ACO-iPLS	S-G smoothing & elimination constant offset	220.553, 222.934, 231.857, 247.308, 254.432, 256.213, 260.959, 266.889, 271.631, 287.619	0.9983	0.6582	0.9839	2.0794

数据进行特征提取后再进行 PSO-LSSVM 建模的模型效果较好。这是由于蚁群算法具有全局搜索能力,可以有效避免特征提取收敛过程中局部最优解的出现;而 PSO-LSSVM 算法同时发挥了 PSO 算法收敛快、全局搜索能力强和 LSSVM 算法的小样本学习能力强、计算简单的特点,结合 PSO 算法理论对 LSSVM 算法的两个关键参数 γ 和 σ 进行寻优运算,直至寻优结束,获得 LSSVM 算法的最优参数组合 (γ, σ) ,提高了最终模型的预测能力及泛化能力。

表 2 实验水样荧光光谱数据的 PSO-LSSVM 建模结果

Table 2 Modeling results of fluorescence spectra of experimental water samples by PSO-LSSVM

Excitation wavelength /nm	Calibration set		Validation set	
	R_c^2	RMSECV /($\text{mg}\cdot\text{L}^{-1}$)	R_p^2	RMSEP /($\text{mg}\cdot\text{L}^{-1}$)
220	0.9999	0.1004	0.9692	2.0987
225	0.9999	0.1703	0.9783	1.7783
230	0.9999	0.0468	0.9779	1.8459
235	0.9980	0.6351	0.9876	1.3725
240	0.9293	3.7334	0.9661	2.5271
245	0.9995	0.3051	0.9684	2.1227
250	0.9980	0.6545	0.9750	1.9570
255	0.9283	3.7744	0.9620	2.5084
260	0.9939	1.1336	0.9634	2.6259
265	0.9998	0.1789	0.9901	1.1986
270	0.9916	1.3183	0.9724	2.1348
275	0.9342	3.3273	0.9038	4.3444
280	0.9923	1.2832	0.9563	2.8420
285	0.9999	0.0486	0.9723	2.2662
290	0.9999	0.0480	0.9839	1.4133
295	0.9996	0.2965	0.9737	2.1070
300	0.9896	1.4185	0.9725	2.1539
305	0.9999	0.0912	0.9656	2.3591
310	0.9993	0.3665	0.9789	1.7811
315	0.9936	1.1119	0.9800	2.2497
320	0.9966	0.8122	0.9690	2.2926
325	0.9999	0.1123	0.9514	3.1902
330	0.9846	1.5396	0.9647	2.6610

由表 2 可知,单激发波长下的荧光发射光谱数据可以对实际水样的水质 COD 进行较为准确的检测,且不同波长激发光下的荧光发射光谱数据模型的预测效果存在较大差异。从建模结果可以看出,对于校正集,激发波长为 230 nm 时的荧光发射光谱数据模型最优,其决定系数 $R_c^2 = 0.9999$,交叉验证均方根误差 RMSECV 为 0.0468 mg/L;对于检验集,激发波长为 265 nm 时的荧光发射光谱数据模型最优,其决定系数 $R_p^2 = 0.9901$,外部检验均方根误差 RMSEP 为 1.1986 mg/L。因此,在后续建立多光谱融合模型时,需参考表 2 数据,选择合适的激发波长下的荧光发射光谱数据和紫外吸收光谱数据进行多光谱数据的融合建模。

3.1.2 荧光发射光谱数据建模

借鉴紫外吸收光谱数据建模经验,基于荧光发射光谱数据建立模型时,仍然采用 ACO-iPLS 作为特征提取算法。实验水样的三维荧光光谱按激发波长展开的荧光发射光谱数据经 ACO-iPLS 提取特征后再由 PSO-LSSVM 算法建模的结果如表 2 所示。其中,各行分别表示在不同的激发波长下的荧光发射光谱数据所建模型的评价指标,各评价指标的定义同表 1。

将实验水样荧光光谱数据的最优建模结果与基于紫外吸收光谱的最优建模结果进行比较可知:在校正集样本和检验集样本相同的划分情况下,荧光发射光谱数据最优模型检验集的预测结果的决定系数增加了 0.63%,预测均方根误差减小了 42.36%;荧光数据模型的预测误差明显低于紫外数据模型的预测误差。这说明,对于本研究所选的实验水样的 COD 检测方法而言,荧光光谱的预测效果更好。实验水样 COD 的紫外吸收光谱模型预测效果欠佳的原因分析如下:实验样本中的海水和地表水的盐度不同,海水的盐度远高于地表水的盐度,而海水中的部分溶解盐在紫外波段又有吸收,因此,紫外吸光度发生变化,导致建模结果误差变大。

3.2 信息融合与多光谱融合模型的建立

信息融合的方式分为三个层次,由低到高依次为数据级融合、特征级融合和决策级融合,这三种融合方式分别对应三种预测分析方法。为保证数据处理的有效性和实时性,本研究采用数据级融合和特征级融合方式对荧光发射光谱信息进行融合分析。根据信息数据来源的不同,光谱信息融合可分为两种:一是光谱信息与其他传感器信息的融合,二是光谱信息与光谱信息间的融合。本研究属于第二种。

3.2.1 紫外-荧光多光谱数据级融合建模

由于紫外吸光度值远小于荧光发射光谱的强度值,因此二者不存在可比性。若直接将紫外吸收光谱和荧光发射光谱数据进行融合,则荧光信息会在建模过程中起主导作用,而紫外吸收光谱信息就会被忽略掉,从而造成光谱信息融合的失败。因此,在紫外光谱和荧光光谱融合前,需将紫外吸收光谱和荧光发射光谱数据进行归一化处理,使二者具有可比性,然后再进行融合、提取特征、建模等操作。归一化的目的是将不同尺度上的评判结果统一到一个尺度上,从而可以作比较和计算。本研究采用矢量归一化(VN)方法对光谱数据进行归一化处理。矢量归一化处理是将原始光谱数据矩阵中每一个元素减去该元素所在列的均值后再除以所在列所有元素的平方和方根的操作,从而使不同尺度的光谱数据统一到相同的尺度上进行操作。

数据级融合为低水平数据融合(LLDF)。对于本研究而言,数据级融合过程为:先将需要融合的光谱数据归一化后进行直接融合,然后再进行特征光谱的提取和校正模型的建立。数据级融合建模流程图如图4所示。其中,“UF-LLDF”表示紫外光谱和

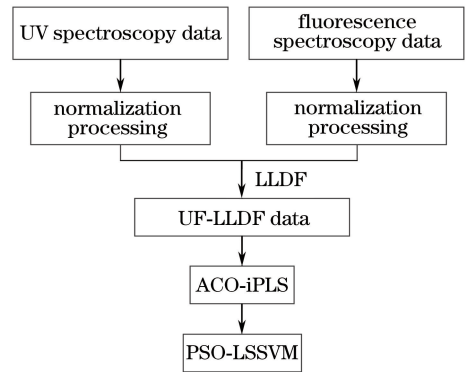


图4 紫外-荧光多光谱数据级融合建模流程图

Fig. 4 Flow chart of low level data fusion modeling based on UV and fluorescence multi-spectral information. 紫外-荧光多光谱数据级融合方式。

对照表2中荧光光谱数据在不同激发波长下荧光发射光谱的COD浓度建模结果,结合校正集和检验集的模型评价指标,选择激发波长230, 265, 290 nm处的荧光发射光谱数据(建模波长范围为250~500 nm)与对应COD浓度的紫外吸收光谱数据进行数据级融合,并采用ACO-iPLS算法进行紫外-荧光多光谱的特征提取,采用PSO-LSSVM算法进行建模。

根据特征提取结果,结合矢量归一化方法,得到基于PSO-LSSVM算法的建模结果。采用矢量归一化方法的紫外-荧光数据级融合的PSO-LSSVM建模结果如表3所示。其中:融合方式LLDF1~LLDF3依次代表紫外吸收光谱与激发波长230, 265, 290 nm下的荧光发射光谱数据进行紫外-荧光多光谱数据级融合;LLDF4~LLDF7代表紫外吸收光谱与两个或三个激发波长下的荧光发射光谱进行数据级信息融合。

表3 紫外-荧光多光谱数据级融合的PSO-LSSVM建模结果

Table 3 PSO-LSSVM modeling results of low level data fusion based on UV and fluorescence multi-spectral information

Fusion approach	Calibration set		Validation set	
	R_c^2	RMSECV / (mg·L ⁻¹)	R_p^2	RMSEP / (mg·L ⁻¹)
LLDF1	0.9999	0.1361	0.9821	1.6552
LLDF2	0.9999	0.0585	0.9886	1.2582
LLDF3	0.9999	0.3969	0.9876	1.4495
LLDF4	0.9999	0.0450	0.9857	1.5243
LLDF5	0.9999	0.0377	0.9766	1.8591
LLDF6	0.9999	0.0713	0.9860	1.5294
LLDF7	0.9999	0.0352	0.9875	1.4589

综合比较LLDF1~LLDF7的模型效果可知,对于检验集而言,LLDF2的PSO-LSSVM模型为紫外光谱与在265 nm激发下的荧光发射光谱进行数据级融合的最优模型,简称为LLDF2-PSO-

LSSVM模型。将LLDF2-PSO-LSSVM模型的预测效果与紫外吸收光谱的最优模型PSO-LSSVM的预测效果进行比较可知,在校正集和检验集相同的情况下,LLDF2-PSO-LSSVM模型的检验集预测

结果决定系数 R_p^2 增加了 0.48%，预测均方根误差 RMSEP 减小了 39.49%，因此 LLDF2-PSO-LSSVM 的效果优于紫外吸收光谱的最优模型预测效果。将 LLDF2-PSO-LSSVM 的模型预测效果与单激发波长 ($E_x = 265 \text{ nm}$) 下的荧光发射光谱最优模型的预测效果进行对比可知，其检验集的预测决定系数减小了 0.15%，均方根误差 RMSEP 却增大了 4.97%，其模型预测效果还稍劣于激发波长为 265 nm 时的荧光发射光谱最优模型的预测效果。原因在于相对于荧光光谱的最优模型，紫外最优模型的效果较差，两光谱信息直接融合建模时模型预测值的相对误差变大，降低了融合模型的预测效果。

3.2.2 紫外-荧光多光谱特征级融合建模

特征级融合为中等水平数据融合 (MLDF)。对于本研究而言，特征级融合过程为：首先将各类光谱数据归一化，再进行特征光谱的提取，然后将提取到的特征光谱数据进行融合，最后将特征级融合光谱数据再次提取特征后进行校正模型的建立。特征级融合建模流程图如图 5 所示。其中，“UF-MLDF”表示紫外光谱和荧光光谱的特征级融合方式。

结合紫外-荧光多光谱数据级建模的经验，选择激发波长为 230, 265, 290 nm 时荧光发射光谱数据和紫外吸收光谱数据进行特征级融合，依然采用矢

表 4 紫外-荧光多光谱特征级融合的 PSO-LSSVM 建模结果

Table 4 PSO-LSSVM modeling results of mid-level data fusion based on UV and fluorescence multi-spectral information

Fusion approach	Calibration set		Validation set	
	R_c^2	RMSECV / ($\text{mg} \cdot \text{L}^{-1}$)	R_p^2	RMSEP / ($\text{mg} \cdot \text{L}^{-1}$)
MLDF1	0.9999	0.0284	0.9863	1.4265
MLDF2	0.9999	0.0285	0.9864	1.4265
MLDF3	0.9999	0.0409	0.9886	1.4086
MLDF4	0.9999	0.0168	0.9856	1.5116
MLDF5	0.9999	0.0304	0.9912	1.1297
MLDF6	0.9999	0.0164	0.9833	1.5715
MLDF7	0.9999	0.0131	0.9848	1.5517

由表 4 可知，对于紫外吸收光谱和荧光发射光谱的特征级融合模型而言，MLDF5 (紫外吸收光谱与 $E_x = 265 \text{ nm}$ 、 $E_x = 290 \text{ nm}$ 处的荧光发射光谱进行多光谱特征级融合) 的 PSO-LSSVM 模型预测效果最优，其预测集决定系数 R_p^2 为 0.9912，外部检验均方根误差 RMSEP 为 1.1297 mg/L。通过对比可知，特征级融合最优模型的预测效果稍优于数据级融合最优模型。相比于 LLDF2-PSO-LSSVM 的效果，MLDF 的最优 PSO-LSSVM 模型的检验集决定系数 R_p^2 增大了 0.27%，而检验集的预测均方根误差 RMSEP 减小了 10.21%。因此，对于紫外-荧光

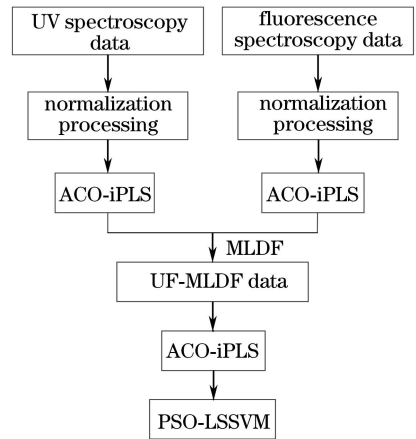


图 5 紫外-荧光多光谱特征级融合建模流程图

Fig. 5 Flow chart of mid-level data fusion modeling based on UV and fluorescence multi-spectral information

量归一化作为光谱数据归一化方法。采用 ACO-iPLS 算法提取特征，根据特征提取结果，按照不同的融合方式将特征光谱进行融合，再次提取特征后采用 PSO-LSSVM 算法对特征光谱数据进行建模，建模结果如表 4 所示。其中：MLDF1~MLDF3 依次代表紫外吸收光谱和激发波长分别为 230, 265, 290 nm 时的荧光发射光谱进行特征级融合；MLDF4~MLDF7 依次代表紫外吸收光谱与两个或三个激发波长下的荧光发射光谱进行特征级信息融合。

多光谱融合建模检测水质 COD，将紫外吸收光谱和 $E_x = 265 \text{ nm}$ 、 $E_x = 290 \text{ nm}$ 处的荧光发射光谱进行多光谱特征级融合得到的预测模型 (MLDF5-PSO-LSSVM) 最优。将 MLDF5-PSO-LSSVM 的模型预测效果与单激发波长 ($E_x = 265 \text{ nm}$) 下的荧光发射光谱的最优模型预测效果进行对比后可知，其检验集的预测决定系数增大了 0.11%，均方根误差 RMSEP 减小了 5.75%，其模型预测效果优于 $E_x = 265 \text{ nm}$ 激发下的荧光发射光谱最优模型的预测效果。

3.3 各类建模方法的综合对比

将各建模方法所得最优模型的评价指标进行综

合对比,可知:各模型校正集和检验集的决定系数以及均方根误差之间具有很好的一致性;对于实验水样的水质 COD 预测模型而言,紫外吸收光谱模型的预测效果最差,紫外和荧光多光谱的 MLDF-

PSO-LSSVM 模型的预测效果最优。最优模型 (MLDF5-PSO-LSSVM 模型)的预测结果如图 6 所示,其中图 6(a)为 MLDF5 融合模型的校正集预测结果,图 6(b)为该模型检验集的预测结果。

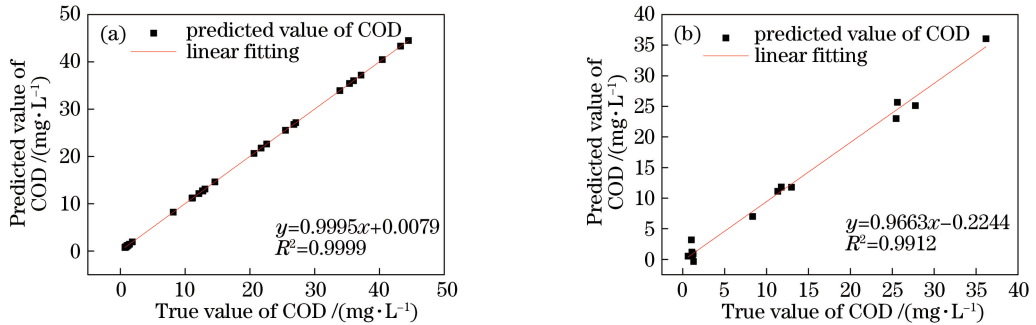


图 6 MLDF5-PSO-LSSVM 模型的预测结果。(a)校正集;(b)检验集

Fig. 6 Prediction results of MLDF5-PSO-LSSVM model. (a) Calibration set; (b) validation set

由图 6(a)可知,由校正集样本数据得到的定量分析模型具有非常好的预测效果,拟合曲线的决定系数 R^2 高达 0.9999,说明校正集样本选择恰当,建模方法正确。由图 6(b)可知,对于检验集数据,定量分析模型的预测效果仍然表现优秀,拟合曲线的决定系数 R_p^2 也大于 0.99,说明本研究方法所建校正模型具有较好的稳健性和适应能力,能很好地完成对水质 COD 的检测工作。

文献[17]采用的实验样本为 COD 标准液,其溶质为单一的邻苯二甲酸氢钾,与实际水样的组成成分有较大差异,因此不将其与本研究进行比较;对于文献[18],其采用的是紫外光光谱与近红外光谱进行融合建模检测水质 COD,而本研究采用的是紫外光谱与荧光发射光谱进行融合建模检测水质 COD,由于水溶液的近红外光谱重叠严重,易受环境影响等原因,使得近红外检测过程中的误差相对较大,从而造成紫外与近红外光谱融合模型的稳定性不佳,总体检测效果不如本研究所建模型的检测效果优秀。

此外,对比文中最好的两个结果:一个是采用 265 nm 激发下的荧光发射光谱检测时,获得的 $R_p^2 = 0.9901$,RMSEP 为 1.1986 mg/L;另一个是采用 MLDF5 算法检测时,获得的 $R_p^2 = 0.9912$,RMSEP 为 1.1297 mg/L。从算法上讲,MLDF5 算法更为复杂,但两模型的预测结果却相差不大。可见,对于本研究所取样本而言,采用 265 nm 激发下的荧光发射光谱对水质 COD 进行检测也可以获得较佳的效果。这不仅降低了算法的复杂度,提高了运算速度,还突出了光谱检测技术在线、无损、快速

的特点。

4 结 论

本研究以实际水样的水质 COD 为研究对象,建立了基于紫外吸收光谱数据模型、单激发波长下的荧光发射光谱数据模型、紫外-荧光多光谱数据级融合模型、紫外-荧光多光谱特征级融合模型,并将各模型的预测效果进行对比。结果表明,采用紫外-荧光多光谱特征级融合建模所得结果最优,说明采用多光谱融合建模是一种对水质 COD 分析的有效建模方法,可以更准确地检测水质 COD。本研究所采用的建模方法具有较好的可行性,实验所得的最优模型的预测效果较好,可为水质 COD 的检测提供一种新的解决方法,也为水质 COD 的光学分析方法提供了一种新的思路。本研究所采用的基于多光谱融合的建模方法也可推广应用于检测混合物中某种物质的含量,为基于光谱学的混合物中某成分的无损检测提供了一种可使用的测量方法,具有一定的参考价值。

参 考 文 献

- [1] Gutiérrez-Capitán M, Baldi A, Gómez R, *et al.* Electrochemical nanocomposite-derived sensor for the analysis of chemical oxygen demand in urban wastewaters [J]. *Analytical Chemistry*, 2015, 87 (4): 2152-2160.
- [2] Shen B J, Zhao Y Y, Xu Y, *et al.* Determination of low chemical oxygen demand of waste water with high chloride by fast digestion-spectrophotometric method[J]. *Chemical Analysis and Meterage*, 2016,

- 25(3): 69-72.
- 沈碧君, 赵洋甬, 徐运, 等. 快速消解分光光度法测定高氯废水中低浓度化学需氧量[J]. 化学分析计量, 2016, 25(3): 69-72.
- [3] Gimeno O, García-Araya J F, Beltrán F J, *et al.* Removal of emerging contaminants from a primary effluent of municipal wastewater by means of sequential biological degradation-solar photocatalytic oxidation processes [J]. *Chemical Engineering Journal*, 2016, 290: 12-20.
- [4] Jin B H, He Y, Shen J C, *et al.* Measurement of chemical oxygen demand (COD) in natural water samples by flow injection ozonation chemiluminescence (FI-CL) technique[J]. *Journal of Environmental Monitoring*, 2004, 6(8): 673-678.
- [5] Hou D B, Zhang J, Chen L, *et al.* Water quality analysis by UV-vis spectroscopy: a review of methodology and application [J]. *Spectroscopy and Spectral Analysis*, 2013, 33(7): 1839-1844.
- 侯迪波, 张坚, 陈冷, 等. 基于紫外-可见光光谱的水质分析方法研究进展与应用[J]. 光谱学与光谱分析, 2013, 33(7): 1839-1844.
- [6] Zhu Y N, Yang P, Yang X Y, *et al.* Classification of fresh meat species using laser-induced breakdown spectroscopy with support vector machine and principal component analysis[J]. *Chinese Journal of Analytical Chemistry*, 2017, 45(3): 336-341.
- 朱毅宁, 杨平, 杨新艳, 等. 支持向量机结合主成分分析辅助激光诱导击穿光谱技术识别鲜肉品种[J]. 分析化学, 2017, 45(3): 336-341.
- [7] Yuan J Z, Lu Q P, Wu C Y, *et al.* Noninvasive human triglyceride detecting with near-infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(1): 42-48.
- 袁境泽, 卢启鹏, 吴春阳, 等. 近红外光谱人体血液甘油三酯无创检测[J]. 光谱学与光谱分析, 2018, 38(1): 42-48.
- [8] Kumar A, Jain S K. Development and validation of UV-spectroscopy based stability indicating method for the determination of fluoxetine hydrochloride[J]. *Analytical Chemistry Letters*, 2016, 6(6): 894-902.
- [9] Liu R X, Chen L L, Zhang H Y, *et al.* A label-free single photonic quantum well biosensor based on porous silicon for DNA detection[J]. *Optoelectronics Letters*, 2013, 9(3): 225-228.
- [10] Dong M, Sui Y, Li G L, *et al.* Mid-infrared carbon monoxide detection system using differential absorption spectroscopy technique [J]. *Optoelectronics Letters*, 2015, 11(6): 469-472.
- [11] 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.
- 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.
- [12] Dahlbacka J, Nyström J, Mossing T, *et al.* On-line measurement of the chemical oxygen demand in wastewater in a pulp and paper mill using near infrared spectroscopy[J]. *Spectral Analysis Review*, 2014, 2(4): 19-25.
- [13] Chen M F, Wu J, Lü Y L, *et al.* Fluorescence properties of municipal wastewater[J]. *Acta Optica Sinica*, 2008, 28(3): 578-582.
- 陈茂福, 吴静, 律严励, 等. 城市污水的三维荧光指纹特征[J]. 光学学报, 2008, 28(3): 578-582.
- [14] Wang J, Zhang F, Wang X P, *et al.* Three-dimensional fluorescence characteristics by parallel factor method coupled with self-organizing map and its relationship with water quality[J]. *Acta Optica Sinica*, 2017, 37(7): 0730003.
- 王娟, 张飞, 王小平, 等. 平行因子法结合自组织映射神经网络的三维荧光特征及其与水质的关系[J]. 光学学报, 2017, 37(7): 0730003.
- [15] Qian C, Wang L F, Chen W, *et al.* Fluorescence approach for the determination of fluorescent dissolved organic matter[J]. *Analytical Chemistry*, 2017, 89(7): 4264-4271.
- [16] Yang L Y, Hur J, Zhuang W N. Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review [J]. *Environmental Science and Pollution Research*, 2015, 22(9): 6500-6510.
- [17] Wu G Q, Bi W H. Research on chemical oxygen demand optical detection method based on the combination of multi-source spectral characteristics [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(11): 3071-3074.
- 吴国庆, 毕卫红. 多源光谱特征组合的 COD 光学检测方法研究[J]. 光谱学与光谱分析, 2014, 34(11): 3071-3074.
- [18] Cao H. Research on rapid determination of organic matter concentration in aquaculture water using multi-source spectral data fusion [D]. Hangzhou: Zhejiang University, 2014: 58-68.

曹泓. 基于多源光谱数据融合的水产养殖水质有机物浓度快速检测研究 [D]. 杭州: 浙江大学, 2014: 58-68.

[19] Zepp R G, Sheldon W M, Moran M A. Dissolved

organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices[J]. Marine Chemistry, 2004, 89: 15-36.