

# 基于荧光多光谱融合的水质化学需氧量的检测

周昆鹏<sup>1</sup>, 白旭芳<sup>1</sup>, 毕卫红<sup>2\*</sup>

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** 为了对水中的有机污染物进行绿色、快速、准确的检测, 提出了一种基于荧光多光谱融合的水质化学需氧量(Chemical Oxygen Demand, COD)的检测方法。实验样本为包含近岸海水和地表水在内的实际水样53份, 采用标准化学方法获取样本的化学需氧量的理化值, 利用荧光分光光度计采集样本的三维荧光光谱并对光谱数据进行处理和建模。在200~300 nm(间隔5 nm)的激发波长范围内将三维光谱展开成二维的发射光谱(发射波长范围250~500 nm, 间隔2 nm)。采用ACO-iPLS(蚁群-区间偏最小二乘)算法提取发射光谱特征, PSO-LSSVM(粒子群优化的最小二乘支持向量机)算法建立预测模型, 分别建立了单激发波长下的荧光发射光谱数据预测模型、多激发波长下发射光谱的数据级融合(LLDF)预测模型以及多激发波长下发射光谱的特征级融合(MLDF)预测模型, 通过对预测效果的对比, 得出结论。实验结果表明, 对于不同激发波长下荧光发射光谱数据而言, 265 nm激发光作用下的发射光谱数据的预测模型最优, 其检验集决定系数 $R_p^2$ 和外部检验均方根误差RMSEP分别为0.990 1和1.198 6 mg·L<sup>-1</sup>; 对于荧光多光谱数据级融合模型(简称为: LLDF-PSO-LSSVM)而言, 在235, 265和290 nm激发光作用下的发射光谱的LLDF模型效果最优, 其检验集的 $R_p^2$ 和RMSEP分别为0.992 2和1.055 1 mg·L<sup>-1</sup>; 对于荧光多光谱特征级融合模型(MLDF-PSO-LSSVM)而言, 在265, 290和305 nm激发光作用下的荧光发射光谱的MLDF模型效果最优, 其 $R_p^2$ =0.998 2, RMSEP=0.534 2 mg·L<sup>-1</sup>。综合比较各类建模结果可知, MLDF-PSO-LSSVM的模型效果最优, 说明基于荧光发射光谱数据, 采用多光谱特征级融合模型检测水质COD时, 检测的精度更高, 预测效果更好。

**关键词** 荧光; 多光谱融合; 预测模型; 化学需氧量; 水质检测

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)03-0813-05

## 引言

光谱分析技术具有快速、灵敏等优点<sup>[1-2]</sup>, 随着科技的进步, 该技术在很多领域均得到广泛的应用<sup>[3-5]</sup>。将光谱技术用于检测水质化学需氧量(chemical oxygen demand, COD)具有检测快速、绿色无污染和重现性好的特点, 在水质参数检测应用方面具有很强的发展潜力。紫外-可见光谱(UV-Vis)法是目前分析水质参数时最常用的一种光谱学分析方法<sup>[6]</sup>。此外, 基于近红外光谱(NIR)和荧光光谱(Fluorescence)分析水质参数的方法也有报道<sup>[7-9]</sup>。也有研究者采用两种不同类别光谱融合的方法对不同水质的有机污染物指

标进行检测<sup>[10-11]</sup>, 但检测结果均不理想。

本文利用实际水样的荧光光谱对其水质化学需氧量进行分析, 建立了基于荧光发射光谱的单光谱数据模型和多光谱信息融合模型, 并对各方法所得到的水质COD预测模型进行了模型效果对比, 进一步得到了基于荧光多光谱信息融合技术检测水质COD的新方法。

## 1 实验部分

### 1.1 样本和仪器

所用实验样本分为地表水和近岸海水2类, 共53份。其中, 地表水样本31份, 分批采集于河北省秦皇岛市的天然河

收稿日期: 2018-05-02, 修订日期: 2018-10-12

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项目(BS432), 内蒙古民族大学科研项目(NMDYB17162)和内蒙古自治区高等学校科学研究项目(NJZY17202)资助

作者简介: 周昆鹏, 1983年生, 内蒙古民族大学物理与电子信息学院讲师 e-mail: kunpeng032@126.com

\* 通讯联系人 e-mail: bwhong@ysu.edu.cn

流和人工湖等地；海水样本 22 份，采集地为秦皇岛市的渤海近岸海域和入海口。对水样检测前需先静置 30 min 后提取静置后的水样上层液体，取出的水样上层液体分为两部分，一部分进行三维荧光光谱的采集，另一部分对其 COD 的理化值进行测量。三维荧光光谱利用日本 Hitachi 公司的 F-7000 型荧光分光光度计采集得到，地表水的 COD 理化值测量采用美国 HACH 公司的相关设备(消解器型号：DRB-200 型，分光光度计型号：DR-6000)检测得到，海水 COD 理化值的测量采用碱式高锰酸钾滴定法得到。

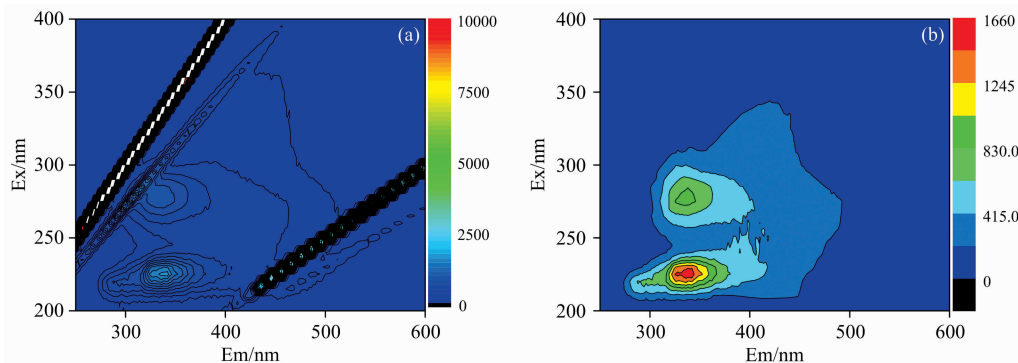


图 1 Delaunay 插值法去除荧光散射峰前后的效果

(a): 去除散射前的效果图; (b): 去除散射后的效果图

Fig. 1 The effect of removing the fluorescence scattering peaks before (a) and after (b) by Delaunay interpolation

由三维荧光光谱图确定有效的激发光波长在 200~330 nm(间隔 5 nm)范围内，发射波长范围为 250~500 nm(间隔 2 nm)，将每组三维荧光光谱数据在有效的激发波长范围内展开，得到一个 3 402 维的二维向量，截取有效激发波长下的发射光谱数据用于光谱特征的提取和预测模型的建立。

### 1.2.2 水样 COD 的理化值测量

海水样品的水质 COD 测量采用碱式高锰酸钾滴定法(GB17378.4—2007)在实验室内进行；地表水样品的水质 COD 测量采用快速消解分光光度法，消解水样时，消解器温度设置为 150 °C，消解时间设置为 120 min。两种方法均为标准方法，为了减小测量误差，采用这两种方法检测水样的 COD 值时，均须取平行双样同时进行检测，且与空白(参比)试液进行对比检测，取两个检测值的平均值作为最终结果。

经标准方法测定，在实验室环境下得到 53 份实验水样的 COD 理化值的范围为 0.64~44.5 mg·L<sup>-1</sup>。

### 1.3 光谱数据处理与数学模型的建立和验证

对实际水样的荧光光谱数据进行建模时，采用 ACO-iPLS 算法进行特征光谱提取，采用 PSO-LSSVM 算法进行建模。ACO-iPLS 算法是基于蚁群算法的波长优选方法，能有效地防止算法在收敛时出现局部最优解的情况发生；PSO-LSSVM 算法具有全局搜索能力和小样本学习能力强的特点，而且计算过程简单，收敛速度快。ACO-iPLS 算法和 PSO-LSSVM 算法的联合使用，可以提高所建模型的泛化能力。

对于采集的 53 份水样的荧光光谱数据，随机分为 2 组，一组作为校正集数据，另一组作为检验集数据。校正集数据的作用是建立准确的训练模型，为了提高校正模型的准确

## 1.2 方法

### 1.2.1 荧光光谱的采集与处理

荧光光谱采集时，激发光的波长在 200~400 nm 范围内，间隔为 5 nm，发射光设置在 220~600 nm 范围内采集，采样间隔设置为 2 nm。三维荧光光谱的采集过程中会出现瑞利散射和拉曼散射，这会影响到数据的可靠性，本文采用 Delaunay 三角内插值法<sup>[12]</sup>对原始的三维荧光光谱数据进行处理，以去除相关的散射峰。散射峰去除前后的效果如图 1 所示。

性，选择 37 份水样作为校正集样本；其余的 16 份水样(海水、地表水样本各 8 份)作为检验集样本，其数据对校正集数据所建模型的预测性能进行实际检验。

## 2 结果与讨论

### 2.1 单激发波长下的荧光发射光谱数据模型的建立

本文所用实验水样的荧光发射光谱数据建模后共得到 23 组结果，限于篇幅，这里只给出部分结果，如表 1 所示。表 1 中， $R_c^2$  代表校正集的决定系数，RMSECV 代表校正集交叉验证均方根误差， $R_p^2$  代表检验集的决定系数，RMSEP 为外部检验均方根误差。 $R_c^2$  和  $R_p^2$  的值越大越好，RMSECV

表 1 实验水样的部分建模结果

Table 1 Partial modeling results of water samples

激发波长 Ex/nm	校正集		检验集	
	$R_c^2$	RMSECV/ (mg·L <sup>-1</sup> )	$R_p^2$	RMSEP/ (mg·L <sup>-1</sup> )
225	0.999 9	0.170 3	0.978 3	1.778 3
<b>230</b>	<b>0.999 9</b>	<b>0.046 8</b>	0.977 9	1.845 9
235	0.998 0	0.635 1	0.987 6	1.372 5
250	0.998 0	0.654 5	0.975 0	1.957 0
<b>265</b>	0.999 8	0.178 9	<b>0.990 1</b>	<b>1.198 6</b>
285	0.999 9	0.048 6	0.972 3	2.266 2
290	0.999 9	0.048 0	0.983 9	1.413 3
295	0.999 6	0.296 5	0.973 7	2.107 0
305	0.999 9	0.091 2	0.965 6	2.359 1
310	0.999 3	0.366 5	0.978 9	1.781 1

和 RMSEP 值越小越好。

由表 1 可知,对于校正集而言,Ex=230 nm 时的模型最优,其决定系数  $R_c^2 = 0.999\ 9$ ,交叉验证均方根误差 RMSECV=0.046 8 mg · L<sup>-1</sup>;对于检验集而言,Ex=265 nm 时的模型最优, $R_p^2$  和 RMSEP 分别为 0.990 1 和 1.198 6 mg · L<sup>-1</sup>。需要注意的是,在进行多光谱融合建模时,应利用表 1 中不同激发光作用下的发射光谱数据模型结果,合理选择不同激发光作用下的发射光谱数据进行融合建模,以得到最优的模型效果。

## 2.2 多激发波长下发射光谱融合模型的建立

表 2 水样荧光的 LLDF 建模结果(部分)

Table 2 LLDF modeling results of water samples (Partial)

激发波长组合	校正集		检验集	
	$R_c^2$	RMSECV/(mg · L <sup>-1</sup> )	$R_p^2$	RMSEP/(mg · L <sup>-1</sup> )
225+230	0.999 9	0.025 8	0.975 5	1.907 3
235+265	0.999 9	0.094 0	0.983 5	1.548 5
<b>235+265+290</b>	0.999 9	0.068 8	<b>0.992 2</b>	<b>1.055 1</b>
265+290+305	0.992 3	1.560 8	0.976 9	2.228 0
230+285+290+305	0.999 8	0.186 7	0.967 4	2.295 3
230+235+265+285+290+305	0.999 9	0.047 8	0.982 5	1.502 8
all Ex	0.999 3	0.525 9	0.988 3	1.414 2

表 2 显示,激发光波长分别在 Ex=235, 265 和 290 nm 处发射光谱的 LLDF-PSO-LSSVM 模型检验集的预测效果最优,其  $R_p^2$  和 RMSEP 分别为 0.992 2, 1.055 1 mg · L<sup>-1</sup>。

### 2.2.2 多激发波长下的发射光谱 MLDF 建模

表 3 水样荧光的 MLDF 建模结果(部分)

Table 3 MLDF modeling results of water samples (Partial)

激发波长组合	校正集		检验集	
	$R_c^2$	RMSECV/(mg · L <sup>-1</sup> )	$R_p^2$	RMSEP/(mg · L <sup>-1</sup> )
225+230	0.999 9	0.038 5	0.980 1	1.700 6
235+265	0.999 9	0.045 6	0.988 6	1.347 1
235+265+290	0.999 9	0.048 1	0.989 1	1.293 6
<b>265+290+305</b>	0.999 9	0.094 5	<b>0.998 2</b>	<b>0.534 2</b>
230+285+290+305	0.999 9	0.040 7	0.985 0	1.593 3
230+235+265+285+290+305	0.999 9	0.039 4	0.986 0	1.554 8
all Ex	0.999 9	0.032 3	0.981 5	1.834 2

表 3 显示,激发光波长分别在 Ex=265, 290 和 305 nm 处的荧光多光谱特征级融合模型(简称为 MLDF-PSO-LSSVM)检验集的预测效果最优,其  $R_p^2$  和 RMSEP 分别为 0.998 2, 0.534 2 mg · L<sup>-1</sup>。

信息融合的方式有数据级融合、特征级融合和决策级融合三个层次,其融合级别由低到高,本文采用前两种融合方式,即数据级融合(LLDF)和特征级融合(MLDF)方式对实验水样的荧光发射光谱进行分析。

### 2.2.1 多激发波长下的发射光谱 LLDF 建模

对照表 1 中实验水样荧光光谱数据的建模结果,综合分析,选择出 10 个不同波长的激发光作用下的荧光发射光谱进行 LLDF 建模,荧光多光谱数据级融合模型(简称为 LLDF-PSO-LSSVM)的部分建模结果如表 2 所示,表 2 中评价指标的定义同表 1。

为了与 LLDF-PSO-LSSVM 的结果进行对比,选择与 2.2.1 节中相同激发/发射光谱数据,且激发波长 Ex 的组合方式也相同,在此基础上,对所选数据进行 MLDF 建模,部分建模结果如表 3 所示。

## 2.3 各模型的预测效果对比

为比较各模型的预测效果,将各方法所得最优模型的评价指标进行对比,不同方法所得最优模型的评价指标如表 4 所示。

表 4 各建模方法所得最优模型的建模效果

Table 4 Modeling results of optimal models built by different modeling methods

模型类别	校正集		检验集	
	$R_c^2$	RMSECV/(mg · L <sup>-1</sup> )	$R_p^2$	RMSEP/(mg · L <sup>-1</sup> )
单激发波长下的发射光谱模型	0.999 8	0.178 9	0.990 1	1.198 6
LLDF-PSO-LSSVM 模型	0.999 9	0.068 8	0.992 2	1.055 1
MLDF-PSO-LSSVM 模型	0.999 9	0.094 5	0.998 2	0.534 2

对比三种不同建模方法的效果可知：本文所采用的基于荧光多光谱融合的 MLDF-PSO-LSSVM 模型具有最优的预

测效果，其校正集和检验集的模型效果如图 2 所示。

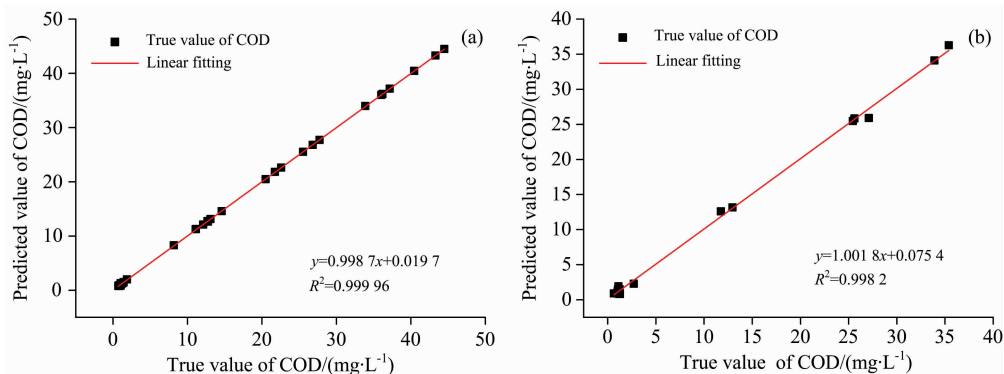


图 2 MLDF-PSO-LSSVM 模型的效果

(a): 校正集效果; (b): 检验集效果

Fig. 2 Effect of calibration set (a) and validation set (b) of the MLDF-PSO-LSSVM model

#### 2.4 本文方法与传统方法的检测效果对比

为了检验本研究的最优模型的预测效果，重新采集近岸海水和地表水样本，在实验室环境下分别采用标准方法和最优荧光多光谱融合法对新采集水样的 COD 进行检测，进而得到水质 COD 标准化验值和荧光多光谱特征级融合预测分析值的比较结果，如表 5 所示。

表 5 本文方法与传统方法的检测效果对比  
Table 5 Comparison between the presented method and the traditional method

水样类别	标准检测法 /(mg · L <sup>-1</sup> )	荧光多光谱 MLDF 融合 预测/(mg · L <sup>-1</sup> )	相对误差 /%
海水 1	1.24	1.45	16.94
海水 2	0.93	1.06	13.98
地表水 1	10.29	10.01	-2.72
地表水 2	12.84	13.06	1.71
地表水 3	6.33	5.87	-7.27
地表水 4	27.34	26.75	-2.16
地表水 5	9.94	10.15	2.11

由表 5 可知，基于荧光多光谱特征级融合预测所得的水质 COD 分析值与标准法的化验值相比能很好的反映出各水样 COD 的变化情况，且能够很好的逼近标准方法检测所得的化验值，相对误差较小，反映出本文方法的准确性。海水的检测相对误差较大的原因可能是由于海水的复杂组分对荧

光光谱造成了干扰。

总之，基于荧光多光谱融合模型具有较高的预测能力和较强的通用性，可以快速、准确的检测出各类水样的水质 COD。

### 3 结 论

以实际水样的水质 COD 为对象，通过建立基于荧光发射光谱数据的单激发波长下的荧光光谱的 PSO-LSSVM 模型、多激发波长下荧光光谱的 LLDF-PSO-LSSVM 模型和多激发波长下荧光光谱的 MLDF-PSO-LSSVM 模型，通过各类模型预测效果的对比可知，采用多种激发光作用下的荧光发射光谱的 MLDF 模型的效果最优，表明对于水质 COD 的检测来说，多激发波长下的荧光发射光谱数据的特征级融合模型预测水质 COD 更加准确。同时，采用荧光数据进行建模，分析方法可行性强，最优模型的预测效果优秀，可作为一种新型的检测水质化学需氧量的有效解决方法。由于实验过程中的采集水样类型较少，水样采集地点所处地域距离较近，因此本方法所得最优模型的适应能力有待进一步提高，后续工作拟针对不同地区不同类型的水样进行荧光光谱数据建模，以提高模型检测不同类型水体的适应性；本研究是在实验室环境下完成的，而实际水体中所含的物质及水体所处的环境较为复杂，水样本身含有荧光敏感物质时及荧光光谱受水体环境影响所带来的干扰有待进一步探讨。

### References

- [1] ZHU Yi-ning, YANG Ping, YANG Xin-yan, et al(朱毅宁, 杨平, 杨新艳, 等). Chinese Journal of Analytical Chemistry(分析化学), 2017, 45(3): 336.
- [2] YUAN Jing-ze, LU Qi-peng, WU Chun-yang, et al(袁境泽, 卢启鹏, 吴春阳, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(1): 42.
- [3] Kumar A, Jain S K. Analytical Chemistry Letters, 2016, 6(6): 894.
- [4] Liu Rongxia, Chen Liangliang, Zhang Hongyan, et al. Optoelectronics Letters, 2013, 9(3): 225.
- [5] Dong Ming, Sui Yue, Li Guolin, et al. Optoelectronics Letters, 2015, 11(6): 469.

- [ 6 ] WU De-cao, WEI Biao, TANG Ge, et al(吴德操, 魏彪, 汤戈, 等). Acta Optica Sinica(光学学报), 2017, 37(2): 0230007.
- [ 7 ] Dahlbacka J, Nyström J, Mossing T, et al. Spectral Analysis Reviews, 2014, 2(4): 19.
- [ 8 ] Wang Xiaoping, Zhang Fei, Kung Hsiangte, et al. Catena, 2017, 155: 62.
- [ 9 ] Qian Chen, Wang Longfei, Chen Wei, et al. Analytical Chemistry, 2017, 89(7): 4264.
- [10] Cao Hong, Qu Wentai, Yang Xianglong. Analytical Methods, 2014, 6(11): 3799.
- [11] WU Xiao-li, LI Yan-jun, WU Tie-jun(武晓莉, 李艳君, 吴铁军). Chinese Journal of Analytical Chemistry(分析化学), 2007, 35(12): 1716.
- [12] Zepp R G, Sheldon W M, Moran M A. Marine Chemistry, 2004, 89(1): 15.

## Detection of Chemical Oxygen Demand (COD) of Water Quality Based on Fluorescence Multi-Spectral Fusion

ZHOU Kun-peng<sup>1</sup>, BAI Xu-fang<sup>1</sup>, BI Wei-hong<sup>2\*</sup>

1. School of Physics and Electronic Information, Inner Mongolia University for Nationalities, Tongliao 028000, China

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

**Abstract** To conduct green, rapid and accurate detection of organic pollutants in water, the current paper proposes a detection method of Chemical Oxygen Demand (COD) based on fluorescence multi-spectral fusion. The experimental samples consist of 53 actual water samples including inshore seawater and surface water. The physicochemical values of COD of the samples are obtained by standard chemical methods. A fluorescence spectrophotometer is used to collect the three-dimensional fluorescence spectra of the samples, and the spectral data are processed and modeled. The three-dimensional fluorescence spectrum is spread at the excitation wavelength in the excitation wavelength range of 200~330 nm and the emission wavelength range of 250~500 nm, with excitation wavelength interval being 5 nm, and the emission wavelength interval 2 nm. With the ant colony optimization-interval partial least squares (ACO-iPLS) as the feature extraction algorithm and the least squares support vector machine algorithm optimized by particle swarm optimization (PSO-LSSVM) as the modeling method, the prediction model of fluorescence emission spectral data at a single excitation wavelength, the fluorescence multi-spectral data-level fusion (Low-Level Data Fusion, LLDf) model and the fluorescence multi-spectral feature-level fusion (Mid-Level Data Fusion, MLDF) model are built respectively, and the prediction effects of various models are compared. The results show that there exist some differences in the prediction effect of the models for the fluorescence emission spectrum data at different excitation wavelengths. The prediction model of the fluorescence emission spectrum data at the excitation wavelength of 265 nm is optimal, with determinant coefficient ( $R_p^2$ ) and the root mean square error in prediction (RMSEP) of the calibration set being 0.990 1 and 1.198 6  $\text{mg} \cdot \text{L}^{-1}$  respectively. For fluorescence multi-spectral data-level fusion models, fluorescence emission spectra at excitation wavelengths of 235, 265, and 290 nm (abbreviated as: LLDf-PSO-LSSVM) have the best prediction effect, with the results of  $R_p^2$  and RMSEP being 0.992 2 and 1.055 1  $\text{mg} \cdot \text{L}^{-1}$  respectively. For fluorescence multi-spectral feature-level fusion models, fluorescence emission spectra at excitation wavelengths of 265, 290, and 305 nm (abbreviated as: MLDF-PSO-LSSVM) have the best prediction effect, with the  $R_p^2$  being 0.998 2 and the RMSEP being 0.534 2  $\text{mg} \cdot \text{L}^{-1}$ . A comprehensive comparison of various modeling results shows that the model of MLDF-PSO-LSSVM has the best performance, indicating that the multi-spectral feature-level fusion model based on fluorescence emission spectrum data is more accurate and more effective for predicting COD of water quality.

**Keywords** Fluorescence spectrum; Multi-spectral fusion; Prediction model; Chemical oxygen demand(COD); Water quality detection

\* Corresponding author

(Received May 2, 2018; accepted Oct. 12, 2018)