

# 水体 COD 激光诱导击穿光谱快速测量方法研究

赵贤德<sup>1,2</sup>, 陈肖<sup>2</sup>, 董大明<sup>2\*</sup>

1. 中国农业大学信息与电气工程学院, 北京 100083
2. 北京农业智能装备技术研究中心, 北京 100097

**摘要** 水体化学需氧量(COD)是一个重要的水体质量指标,一般用来衡量有机物的污染程度。对 COD 的检测长期依赖采样后的实验室化学分析方法,目前应用最普遍的是重铬酸钾氧化法与酸性高锰酸钾氧化法。化学分析的方法操作复杂,耗时费力,且引入新的化学药剂,造成二次污染,因此,急需一种能够实现水体 COD 快速测量的检测技术。在前期研究基础上,对水体 COD 的激光诱导击穿光谱检测方法进行深入探索,重点是优化模型预测速度,目的是研究激光诱导击穿光谱技术用于对水体 COD 的快速测量方法。采集了不同 COD 浓度的 99 个水体样本,分为训练集和测试集两组,通过重铬酸钾氧化法测定各水样的 COD 值,作为真实值,利用实验室自建的激光诱导击穿光谱采集系统采集各水样波长在 200~1 000 nm 的光谱信息,利用偏最小二乘算法建立训练集水样 COD 的定量化测量模型,然后对测试集光谱数据进行预测,将预测结果与实验室化学方法测定的真实值进行对比,评估预测效果。通过对原始光谱建立的预测模型进行分析,发现在建模过程中,大量的激光诱导击穿光谱数据与 COD 浓度相关性很差,而这些无用数据参与计算,浪费了计算资源,延长了检测时间,造成系统负载过大,不利于便携式检测设备的开发。重点研究贡献度最大的前几个主成分,通过对 COD 测量原理和 PLS 模型载荷分析,找到 LIBS 光谱中与水样 COD 浓度相关性最高的主要特征峰,经过分析发现,主要为来源于水中有有机物中的 C, H, O, N 以及水中一些还原性离子元素的特征峰,这些特征峰对 COD 的模型预测能力贡献最大,而 COD 的定义正是衡量水体中这些元素的多少,这与该研究分析结论相吻合。为了实现检测速度的提升,提取这些特征峰,对光谱数据进行降维,剔除大量无关或相关性较低的数据,经过多次筛选和降维,最终将原来参与计算的每条光谱的 13 622 个数据降到 28 个,大大降低系统的运算量,却依然能够保留不错的预测能力。筛选出的 28 个特征波长最能反映水体 COD 浓度,为水体 COD 便携式的多波段检测设备的研发,实现对 COD 的快速测量奠定了基础。

**关键词** 激光诱导击穿光谱; 化学需氧量; 测量; 偏最小二乘法

**中图分类号:** X832 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)09-2907-05

## 引言

水是生命之源,水源直接关系到人类的可持续发展。水质污染成为人们越来越关注的问题,而化学需氧量(chemical oxygen demand, COD)是评价水体有机物污染的一个重要指标。目前,在检测 COD 浓度时,普遍应用的方法是重铬酸钾氧化法与高锰酸钾氧化法<sup>[1-2]</sup>。然而这两种方法都需使用化学试剂,耗时长,会产生二次污染,并且不能实现实时检测水质的需求<sup>[3-4]</sup>。近年来,有学者提出使用紫外光谱法和近红外光谱法测量水体 COD<sup>[5]</sup>,这两种方法具有实时、无损测

量、环境友好的优点,但由于水体透过率差异大,导致在同一光程下的量程范围较小,因此很少用于实际的水质快速检测。

激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)是一种新型原子光谱分析技术,此方法的工作原理是将一束高强度的激光聚焦在样品表面,瞬间内将样品激发至等离子体态,获取等离子体云团的发射谱线。LIBS 具有快速、实时、且无需样品预处理等优点。几乎可以测量所有物质中的化学成分,在目前的材料分析中具有广泛的应用<sup>[7-8]</sup>。LIBS 在液态物质测量中的应用也逐渐获得了发展,研究者们尝试了多种液态物质的 LIBS 测量方法,如直接用

收稿日期: 2018-08-15, 修订日期: 2018-12-20

基金项目: 国家自然科学基金项目(31622040, 31600417)和北京市农林科学院创新能力建设专项(KJCX20170701)资助

作者简介: 赵贤德, 1984 年生, 北京农业智能装备技术研究中心助理研究员 e-mail: zhaoxd@nrcita.org.cn

\* 通讯联系人 e-mail: dongdm@nrcita.org.cn

光纤浸入液体测量<sup>[9]</sup>, 在空气中直接击穿液滴测量<sup>[10]</sup>, 或将液态转换为固态测量<sup>[11]</sup>, 并不断提高检测的灵敏度和精度。此外, LIBS 的性质导致其主要用于元素测量, 但近几年来也有研究者结合化学计量学, 研究 LIBS 对分子的测量方法<sup>[12]</sup>, 本课题组之前也研究了利用 LIBS 结合偏最小二乘法测量水体中 COD 的浓度<sup>[13]</sup>, 主要分析不同的预处理后的建模效果, 并初步发现了 LIBS 光谱与 COD 相关性较高的特征波段。

在前期研究的基础上, 尝试利用 LIBS 实现水体 COD 的快速测量, 验证利用相关性较高的特征波段建模对 COD 浓度的预测能力, 并继续对特征波长进行筛选, 找到 LIBS 光谱对 COD 最相关的主要特征波长, 剔除大部分相关度低的数据, 实现对模型自变量的降维, 大大降低运算量, 有利于小型速测设备的研制和开发。

## 1 实验部分

### 1.1 样品

实验所需水样取自北京市朝阳区亮马河的不同河段, 选择污物和杂质较少的河段取水, 共取水样 10 组, 每个水样各取约 500 mL, 水样运输过程中采用棕色瓶低温密封保存, 测试时间距离样本采回不超过 48 h。

### 1.2 样品制备与化验

水样采回后, 静置 1 h 后, 取上清液, 用去离子水进行稀释, 每组水样稀释十个梯度, 10 组水样得到 100 个不同浓度的被测样本, 并编号, 每个样本分成两份, 一份用于 LIBS 光谱法测量, 另一份送到清华大学环境质量检测中心进行化验, 化验采用的方法是利用重铬酸钾氧化法测定 COD 的值。经过化验, 得到 100 个样本的 COD 浓度, 剔除一个明显的异常样本值, 最终得到 99 个有效样本浓度数据。所有有效样本数据被分为两组, 训练集包括 70 个, 测试集包括 29 个, 两组样本数据的 COD 信息如表 1 所示。

表 1 有效水体样本 COD 浓度信息  
Table 1 COD concentration information of effective water samples

参数	训练集	测试集
样本数量	70	29
最小值/(mg · L <sup>-1</sup> )	7.52	9.11
最大值/(mg · L <sup>-1</sup> )	90.70	89.18
平均值/(mg · L <sup>-1</sup> )	30.81	28.43
样本标准差	18.61	16.75

### 1.3 LIBS 光谱采集

本实验光谱采集使用的 LIBS 采集系统为实验室自己搭建, 包括激光器、光谱仪、三维精密运动平台、信号延时器。激光器为镭宝公司 (Beamtech) 生产的型号为 Dawa-200 的 Dawa Series Q-Switched Nd: YAG Laser System。基频光波长 1 064 nm, 脉冲激光输出的最大能量为 200 mJ, 激光器的脉冲最高重复频率为 20 Hz, 脉冲宽度为 3~5 ns。光谱仪采

用 Ocean Optics 公司的 HR2000+, 光谱范围为 200~1 100 nm, 分辨率为 0.2 nm, 信噪比 250:1。系统搭载有三维精密运动平台, 利用步进电机实现平台的三个轴向运动, 从而对样品的不同位置进行测量。实验装置结构示意图如图 1 所示。激光器出射功率设置为 160 mJ。光谱范围设置为 200~1 000 nm, 积分时间设置为 2 ms。激光器触发和光谱采集之间的延时时间设置为 3 μs。实验前用光谱仪采集背景光谱, 在后续的检测中均扣除背景。

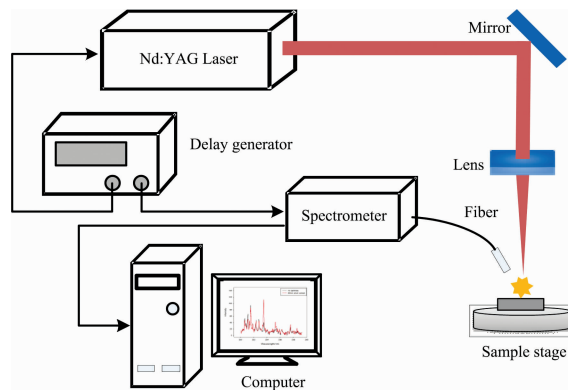


图 1 LIBS 系统结构示意图

Fig. 1 The schematic diagram of LIBS system

测量过程是, 将水样滴于滤纸上, 静置 3 min, 水样均匀浸润滤纸, 将该滤纸置于样品台。激光器发出的激光经反射镜反射后由凸透镜聚焦在滤纸表面产生等离子体。为避免滤纸表面的不均匀性对实验的干扰, 每个样本以重新取样的方式各测 2 次, 每次测量选择 5 个不同的位置, 两次总共得到不同位置的 10 条光谱, 然后对 10 条光谱取平均, 得到一条有效光谱。99 个水体样本全部测完, 共得到 99 条有效光谱。

## 2 结果与讨论

### 2.1 水体样本的 LIBS 光谱

实验采集的水体样本杂质较少, 因此 LIBS 系统采集到的 LIBS 光谱主要包含水体中常见离子和可溶性有机物等, 光谱谱峰相对比较简单, 如图 2 所示。

对水样的 LIBS 光谱的谱峰进行了辨别, 发现主要为 C (247.8 nm), H (656.4 nm), O (777.3 nm), N (742.3, 744.1, 746.7 nm) 以及一些水中常见的 Na (589.0, 589.6 nm), Mg (279.5 nm), Ca (393.4, 396.9 nm) 等元素的谱线, 这些元素主要来自于溶解在水中的有机物分子, 和一些可溶性无机盐, 这些都可以对 COD 的浓度产生影响, 但主要影响因素还是水中的还原性有机物。同时, LIBS 光谱中还有一些分子的谱线, 比如 CN (386.2, 387.1, 388.3 nm) 和 H<sub>2</sub> (422.7 nm), 这些分子片段主要来自于水中的有机物在被高能激光激发时, 有机物分子的部分化学键没有完全被打断, 形成分子片段部分分子, 比如部分 CN 基团, 是由于被激发成等离子体的 C 和 N 元素在冷却过程中重新结合形成新的 CN 基团。这些分子的 LIBS 特征峰能够间接反映水中的 COD 浓度。但是经过对不同浓度 COD 水样的 LIBS 的特

征峰的对比研究,发现这些特征峰的峰强并不随着 COD 的浓度变化表现出相对应的变化,主要是因为水体中 COD 是一个比较复杂的参数,受水中多种因素综合影响,单一的元素变化并不能完全代表水体中 COD 的变化。因此利用化学计量学方法,通过光谱数据建模,对水体的 COD 进行分析。

归系数较大,从而可以找到水体中对 COD 浓度密切相关的特征波长,对其中主要的特征波长进行了辨别,如图 5 所示,可以看到,回归系数变化主要与 C, H, O, N 及一些离子相关,基本与水体 LIBS 的特征峰吻合,这也是原子发射光谱与近红外光谱的不同之处。

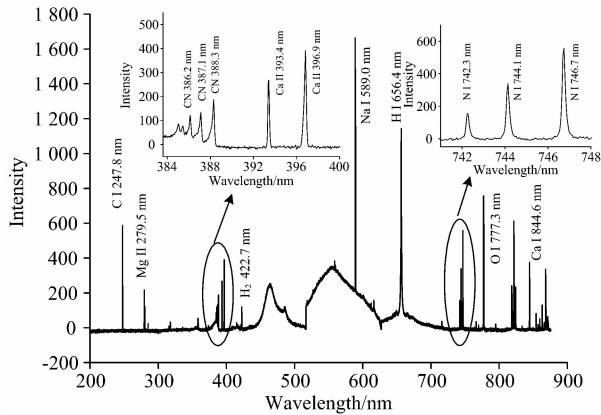


图 2 水样 LIBS 光谱

Fig. 2 LIBS spectrum of water sample

2.2 LIBS 光谱全波段 PLSR 分析

偏最小二乘回归(partial least squares regression, PLSR)是化学计量学中应用最广泛的一种多元校正方法,它的优点在于对光谱进行分解的同时,考虑数据与目标分析物的性质之间的关系。

对训练集的 70 个样本光谱基于光谱的全波段(200~890 nm)进行 PLSR 建模,然后利用建好的模型对测试集的 29 个样本进行预测,以评估模型的预测效果。为了研究各个波长对模型的贡献大小,从而筛选典型特征波段,没有对 LIBS 光谱做任何预处理,以防止预处理导致的单个波长特征的模糊和均匀,更好的保留光谱的原始信息。

建模和预测的效果如表 2 所示,利用全光谱建模,此时主成分设为 9,训练集模型  $R^2$  为 0.969, RMSE 为 3.049,测试集的  $R^2$  为 0.889, RMSE 为 6.391。测试集的 29 个样本的真实值和预测值的拟合图如图 3 所示。

表 2 全光谱 PLS 模型性能

Table 2 Performance of PLS model based on full spectrum

类别	样本数	$R^2$	RMSE
训练集	70	0.969	3.049
测试集	29	0.889	6.391

通过对水体 LIBS 全光谱建模,可以得到不错的预测效果。但是全光谱计算量太大,对处理系统的计算能力要求较高,因此希望能够在保证一定准确率的前提下对模型输入量进行降维,减少计算量,便于将模型移植到便携式测量设备中,以实现现场快速测量。对全光谱模型进行研究,如图 4 所示,从主成分得分图可以看出,主要的贡献度来自于前几个主成分,分析了前 5 个主成分。对各个主成分每个波长的回归系数进行了研究,对结果影响比较大的波长所对应的回

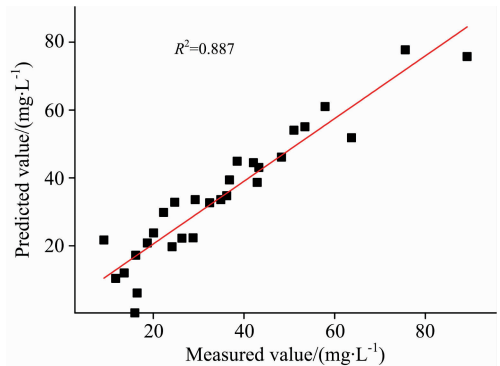


图 3 预测集真值与预测值线性拟合图

Fig. 3 Linear fitting graph of truth value and prediction value of prediction set

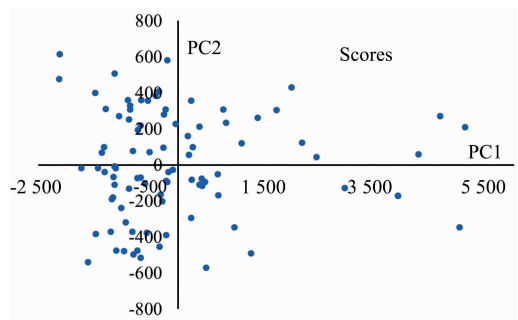


图 4 模型 PC1 和 PC2 得分图

Fig. 4 The scores of PC1 and PC2 in the model

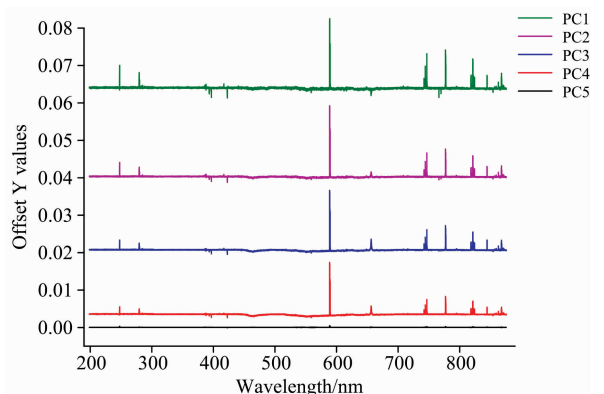


图 5 前 5 个主成分各波长的回归系数

Fig. 5 The regression coefficients of the first five principal components at different wavelengths

在以上研究基础上,根据各波长的载荷对水体 COD 的 LIBS 光谱进行了特征波段筛选,剔除大量贡献度极小的数据,保留相关度高的特征波长数据,然后再次进行建模并对

测试集数据的预测效果进行评估。

### 2.3 波段筛选后的 PLSR 分析

通过数据筛选,主要选择了 C, H, O, N, CN, H<sub>2</sub> 及 Na, Mg, Ca 等特征波长数据,最后将每条光谱的数据量由 13 622 降到 877,此时的模型预测效果如表 3 所示。测试集数据的预测效果并没有下降,如图 6 所示,说明数据降维后,并不影响测量效果,对此时的回归系数进行分析,继续对数据进行筛选。

表 3 877 个特征波长数据筛选后模型性能

Table 3 Model performance of 877 feature wavelength data after screening

类别	样本数	R <sup>2</sup>	RMSE
训练集	70	0.932	4.770
测试集	29	0.885	6.511

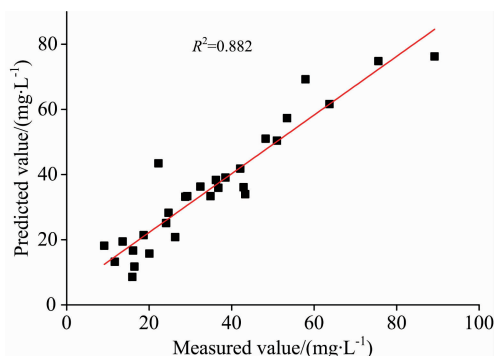


图 6 877 个特征波长筛选后预测集  
真值与预测值线性拟合图

Fig. 6 Linear fitting of true and predicted value of 877 characteristic wavelengths after screening

经过再次筛选,最终将参与建模的光谱波段定为 28 个,主要是每个特征峰的最高峰位波长,这 28 个特征波长分别是 247.8, 279.5, 280.2, 386.1, 387.1, 388.3, 393.4, 396.8, 422.7, 589.0, 589.6, 656.4, 742.2, 744.1, 746.7, 777.1, 794.7, 818.4, 818.7, 821.0, 821.6, 824.2, 844.6, 854.2, 862.9, 868.0, 868.3, 868.6 nm。然后利用这 28 个波长的数据建立 PLSR 模型,主成分选择 28 个,但实际效果

最好的为 18 个,利用建立的模型对 29 个测试集样本进行预测,效果如表 4 所示,从图 7 中也能看到,28 个特征波长筛选后预测集的真值与预测值线性拟合依然为强相关性。虽然均方根误差有所增大,但依然在可接受的范围内。

表 4 28 个特征波长数据筛选后模型性能

Table 4 Model performance of 28 characteristic wavelengths data after screening

类别	样本数	R <sup>2</sup>	RMSE
训练集	70	0.759	8.564
测试集	29	0.676	10.907

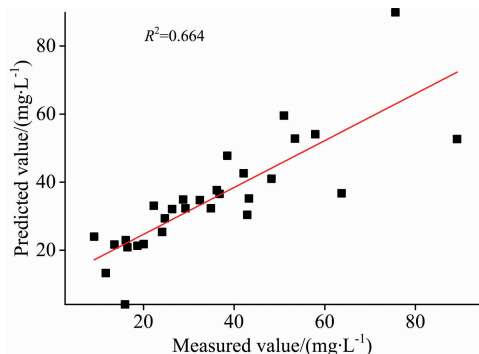


图 7 28 个特征波长筛选后预测集  
真值与预测值线性拟合图

Fig. 7 Linear fitting of true and predicted values of 28 sets of characteristic wavelengths after screening

## 3 结 论

研究了水体 COD 的 LIBS 光谱测量方法,由于 COD 受影响因素众多,并不能根据某单一元素的特征峰进行预测,因此结合化学计量学方法分析,发现水体 COD 主要与水体中有机物及一些无机还原性离子的特征峰相关性较强。根据对特征元素特征波段的筛选和模型输入量的降维,将每条光谱的 13 622 个数据降到 28 个,仍然具有一定的预测能力,这说明利用 LIBS 光谱对水体 COD 进行分析,不仅可行,而且具有很好的开发便携式现场快速测量设备的潜力。

## References

- [1] Goh C P, Lim P E. Asean Journal on Science and Technology for Development, 2017, 25(2): 383.
- [2] ZHANG Ji-hua(张继华). Environmental Science and Technology(环境科学与技术), 2004, 27(8): 37.
- [3] YUAN Han-hong, HU Yan-ping, QI Hui-liang(袁汉鸿, 胡艳平, 戚惠良). Water Purification Technology(净水技术), 2005, 24(4): 67.
- [4] LUO Guo-bing(罗国兵). Rock and Mineral Analysis(岩矿测试), 2013, 32(6): 860.
- [5] Pan T, Chen Z, Chen J, et al. Analytical Methods, 2012, 4(4): 1046.
- [6] Fortes F J, Moros J, Lucena P, et al. Analytical Chemistry, 2013, 85(2): 640.
- [7] Zhang D C, Hu Z Q, Su Y B, et al. Optics Express, 2018, 26(14): 18794.
- [8] Costa V C, Aquino F W B, Paranhos C M, et al. Polymer Testing, 2017, 59: 390.
- [9] Hoehse M, Paul A, Gornushkin I, et al. Analytical & Bioanalytical Chemistry, 2012, 402(4): 1443.

- [10] Janzen C, Fleige R, Noll R, et al. *Spectrochimica Acta Part B, Atomic Spectroscopy*, 2005, 60(7): 993.
- [11] Bilge G, Sezer B, Boyaci I H, et al. *Spectrochimica Acta Part B, Atomic Spectroscopy*, 2018, 145.
- [12] Mousavi S J, Farsani M H, Darbani S M R, et al. *Applied Physics B*, 2016, 122(5): 1.
- [13] YE Song, GU Ya-hui, DU Xiao-fan, et al(叶松, 顾亚辉, 杜晓凡, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2017, 37(11): 3585.

## Study of the Rapid Measurement of COD by Laser-Induced Breakdown Spectroscopy

ZHAO Xian-de<sup>1,2</sup>, CHEN Xiao<sup>2</sup>, DONG Da-ming<sup>2\*</sup>

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Beijing Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China

**Abstract** Chemical Oxygen Demand (COD) is an important water quality parameter, which is generally used to reflect the degree of organic pollution. The detection of COD has long relied on laboratory chemical analysis methods after sampling. The most commonly used method is potassium dichromate oxidation and acid Potassium Permanganate oxidation. However, the chemical analysis methods are complicated, time-consuming and labor-intensive. Moreover, the introduction of new chemicals by these methods causes secondary pollution. Therefore, there is an urgent need for a detection technique that enables rapid measurement of COD in water. Based on the previous research, the test method of COD by the laser-induced breakdown spectroscopy was explored in this paper. The focus was to optimize the model prediction speed in order to study the rapid measurement of COD in water by the technique of laser-induced breakdown spectroscopy. We collected 99 water samples with different COD concentrations, and divided them into two groups: training set and testing set. The COD concentration of each water sample was measured by potassium dichromate oxidation method. And the spectral information of each water sample at the wavelengths of 200~1 000 nm was collected by the laser-induced breakdown spectroscopy acquisition system built by our laboratory. Partial least squares (PLS) algorithm was used to establish a quantitative measurement model for COD of training set samples and the spectral data of test set were predicted. The predicted results were compared with the real values measured by laboratory chemical methods to evaluate the predicted results. By analyzing the prediction model established by the original spectrum, it was found that a large number of laser-induced breakdown spectral data have poor correlation with COD concentration during the modeling process, and these useless data participated in the calculation, wasting computing resources, dragging the detection time, causing the system load to be too large, which was not conducive to the development of portable detection equipment. We focused on the first few principal components with the largest contribution. By analyzing the principle of COD measurement and the load of PLS model, we found the main characteristic peaks of LIBS spectrum which have the highest correlation with the concentration of COD in water. These characteristic peaks belong to C, H, O, N and some reductive ion elements in water. Most of C, H, O and N come from organic matter in water whose characteristic peaks have the greatest contribution to the prediction ability of COD model. The definition of COD reflects the amount of these elements in the water body, which is consistent with our analysis conclusion. In order to improve the detection speed, we extracted these characteristic peaks, and eliminated a large number of unrelated or low-correlation data. After many times of screening and dimensionality reduction, the original 13 622 data of each spectrum were reduced to 28, which greatly reduced the computational complexity of the system, but still retains good prediction ability. The 28 characteristic wavelengths selected are the best ones to reflect the concentration of COD in water, which lays a foundation for the development of portable multi-band detection equipment for COD in water and the rapid measurement of COD.

**Keywords** Laser-induced breakdown spectroscopy; Chemical oxygen demand; Measurement; Partial least square

(Received Aug. 15, 2018; accepted Dec. 20, 2018)

\* Corresponding author