

基于荧光发射光谱的水质化学需氧量的检测

周昆鹏¹, 刘双硕¹, 崔健¹, 张红娜¹, 毕卫红^{2*}, 唐维²

1. 内蒙古民族大学工学院, 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 采用光谱技术检测水质参数是当前研究的一个热点, 提出了一种基于荧光发射光谱的水质化学需氧量(COD)的检测方法。实验样本分为两组, 第一组为20份COD标准溶液, 第二组为63份实际水样。实验样本的化学需氧量检测采用快速消解分光光度法, 利用三维荧光分光光度计采集水样在EX=275 nm激发波长下的荧光发射光谱(荧光发射光谱范围为EM=325~450 nm), 并对两类水样的荧光发射光谱数据进行了处理和建模。分别采用主成分回归(PCR)和偏最小二乘回归(PLSR)对两类水样的荧光发射光谱数据进行了预测模型的建立, 并对模型效果进行了对比。为验证该方法的可行性和模型的预测能力, 将所建PLSR模型预测结果与标准方法的检测结果进行了对比。结果表明, 对于COD标准液来说, PLSR和PCR的主成分数分别取5和8时所得的模型的交叉检验效果最优, 校正模型的决定系数分别为 $R_{PLS}^2=0.9999$ 和 $R_{PCR}^2=0.9897$, 校正模型对检验集数据的预测误差不超过10%, 且PLSR所建模型优于PCR模型。对于实际水样而言, PLSR和PCR的主成分数分别为6和7时, 校正模型的交叉验证效果最优。PLSR法和PCR法的校正集的交叉检验均方差RMSECV_{PLS}和RMSECV_{PCR}分别为0.9322和0.9764 mg·L⁻¹。对于实际水样的检验集来说, PLSR法和PCR法的预测决定系数 R_{PLS}^2 和 R_{PCR}^2 分别为0.9402和0.9190, 说明PLSR法的预测效果更优, 基于荧光发射光谱数据的PLSR模型具有较高的预测能力和较强的适应性, 可以快速、准确的检测出水质COD。通过和传统检测方法的效果对比可知, 该方法可用于检测有机污染物浓度较低的水体, 有机物质浓度较高时采用该方法时检测误差会变大。该研究为水质检测光学传感器的研发提供了一种新的设计思路。

关键词 水质检测; 化学需氧量; 荧光; 发射光谱; 预测模型

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)04-1143-06

引言

光谱分析方法因其具有快速、绿色、无损等诸多优点, 已在多领域得到广泛的应用, 如食品^[1]、医药卫生^[2]、生命安全^[3]、环境监测^[4]等, 基于光谱法的水质参数检测技术^[5-6]也是其典型应用, 如紫外-可见(UV-Vis)吸收光谱法^[7]、近红外(NIR)光谱法^[8]和荧光(fluorescence)光谱法^[9]等。其中, UV-Vis方法最为常见, 采用NIR法检测水质还处于实验室研究阶段, 而荧光法目前主要用于检测水体中的溶解有机物(DOM)。

通过采集水样的荧光发射光谱, 结合相关的化学计量学

算法对水样的COD浓度值进行分析, 最终得出了基于荧光发射光谱对水质COD检测的方法, 该方法检测速度快、绿色环保、无二次污染, 具有较强的应用价值。

1 实验部分

1.1 样本

实验样本分为COD标准溶液和实际水样两类。其中, 不同浓度的COD标准溶液20份, 其最低浓度为2.5 mg·L⁻¹, 最高浓度为100 mg·L⁻¹; 另在不同的区域采集实际水样63份, 实际水样均取自湖泊、河流等地表水源, 采集实际水样后在室温下静置30分钟, 取上层液体进行实验。

收稿日期: 2019-11-05, 修订日期: 2020-01-10

基金项目: 国家自然科学基金项目(61963031), 内蒙古自治区自然科学基金项目(2019MS06017), 内蒙古自治区高等学校科学研究项目(NJZY20122), 河北省重点研发计划项目(18273302D)和内蒙古自治区大学生创新训练计划项目(201910136011)资助

作者简介: 周昆鹏, 1983年生, 内蒙古民族大学工学院讲师 e-mail: kunpeng032@imun.edu.cn

* 通讯联系人 e-mail: bwhong@ysu.edu.cn

1.2 仪器

荧光光谱采集设备采用日本 Hitachi 公司的 F-7000 型荧光分光光度计, 其分辨率可达 1 nm, 波长准确度为 ± 1 nm, 狭缝(光谱通带)为 1~20 nm, 其采用的色散单元为光栅, 光源为 150 W 的氙灯。实际水样的化学需氧量采用 DRB200 消解器和 DR6000 分光光度计(HACH 公司, 美国)进行标准化测量。

1.3 方法

1.3.1 荧光光谱的采集与处理

将 F-7000 荧光光谱仪的光电倍增管电压设置为 700 V, 以 5 nm 为采样间隔在 200~400 nm 的激发波长(EX)范围内采集荧光发射光谱, 发射光谱范围为 EM=220~600 nm, 间隔为 2 nm。

由于三维荧光采集过程中存在比较强的散射峰(主要为瑞利散射和拉曼散射), 对特征峰会造成干扰, 必须去除以提高检测精度、减小误差^[10]。某实际水样去除散射峰后的三维荧光光谱图如图 1 所示。

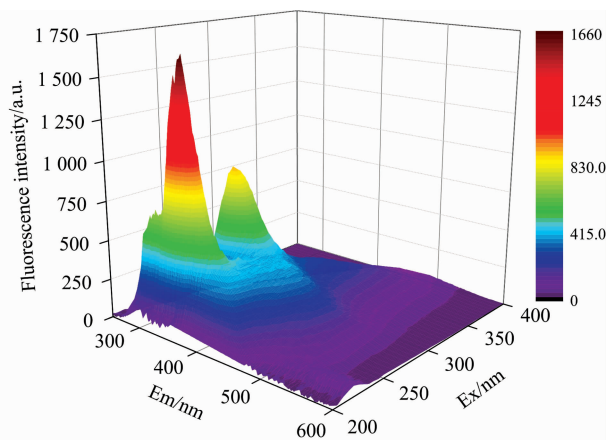


图 1 某实际水样去散射后的三维荧光光谱图
Fig. 1 EEMs of a real water sample after removal of scattering

去除干扰峰后发现, 在 EX=275 nm 激发波长处的荧光发射光谱变化最为明显, 275/350 nm 的激发/发射波长对对应的特征峰的荧光强度最强。因此将每组实验样本的三维光谱数据在 EX=275 nm 处进行荧光发射光谱的提取, 得到一组基于发射荧光(发射波长-荧光强度)光谱数据。为了保证光谱数据的准确性, 在光谱采集过程中将每一份实验样本采集三次光谱求均值后得到最终光谱数据进行数学模型的建立。

1.3.2 实际水样 COD 的理化值测量

为提高实际水样化学需氧量检测的准确性, 减小检测误差, 采用快速消解分光光度法测量时将包括参比试剂在内的各水样进行平行双样测量。测量完成后, 将各测量结果求均值后作为最终检测值。经测定, 在实验室环境下得到 63 份实验水样的 COD 理化值的范围为 0.64~44.5 mg·L⁻¹。

1.4 数学模型的建立和验证

对于 COD 标准液样本和实际水样的荧光发射光谱数据

进行建模时, 均采用基于 PCR 算法和 PLS 算法对水质 COD 浓度进行定量建模, 为了验证建模方法的有效性和准确性, 对实验样本进行如下处理:

(1) 在 20 组 COD 标准液的荧光光谱数据中随机抽取 15 组作为校正集数据建立模型, 其余 5 组数据作为检验集数据对校正模型的准确性进行验证。

(2) 在 63 组水样的荧光光谱数据中随机抽取 48 组作为校正集数据, 剩余 15 组作为检验集数据。

(3) 两类实验样本的荧光发射光谱范围均为 325~450 nm。

2 结果与讨论

2.1 COD 标准液光谱的模型建立

分别对 COD 标准液的荧光发射光谱的校正集数据基于 PCR 算法和 PLSR 算法建立模型。经比较, PCR 算法和 PLSR 算法的主成分数分别取 8 和 5 时, 两类模型的效果最优。建模结果如图 2 所示, 图中 R^2 为决定系数(又称拟合优度), 其决定了相关的密切程度。

由图 2 可看出, COD 标准液校正集数据的 PCR 与 PLSR 模型的 R^2 分别为: $R^2_{PCR}=0.9897$ 和 $R^2_{PLSR}=0.9999$ 。将检验集数据代入校正模型进行检验, 得到的预测结果如表 1 所示。

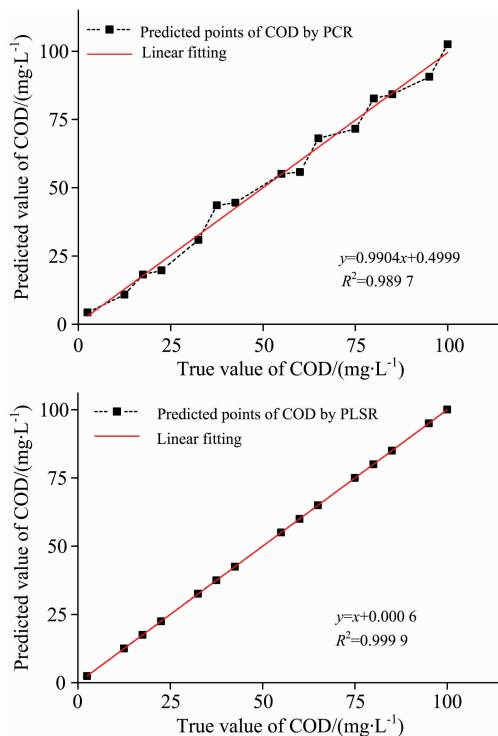


图 2 COD 标准液数据的 PCR 和 PLSR 建模结果
Fig. 2 Modeling results of PCR and PLSR for COD standard solution data

从表 1 可知, 检验集样本 COD 值的取值范围为 7.5~90 mg·L⁻¹, 且基于 PCR 和 PLSR 方法处理的数据结果误差都

不超过 10%，因此两种建模方法均能较好的对 COD 标准液的荧光发射光谱数据进行模型的建立，且 PLS 法处理的数据所建立的模型更加精确。经过重复性实验验证，个别预测结果误差较大的原因是由于配制标准溶液过程中引入了误差，或在采集样品光谱时比色皿表面洁净程度不同而引入了误差。

2.2 实际水样光谱的模型建立

由于实际水样的组分较为复杂，因此得到的光谱数据也存在不少干扰。先采用 S-G 平滑处理算法将提取出的实际水样的荧光发射光谱数据进行平滑处理，处理后的谱线变得平滑但不改变谱线的整体变化趋势。平滑处理后，将校正集数据进行 PCR 和 PLSR 建模，确定主成分数的方法采用的是目前在诸多方法中最常用的预测残差平方和 (PRESS) 的方法，PRESS 越小，所得的模型的预测能力越好。PCR 和 PLSR 算法的主成分数/隐含变量和 PRESS 的关系如图 3 所示。

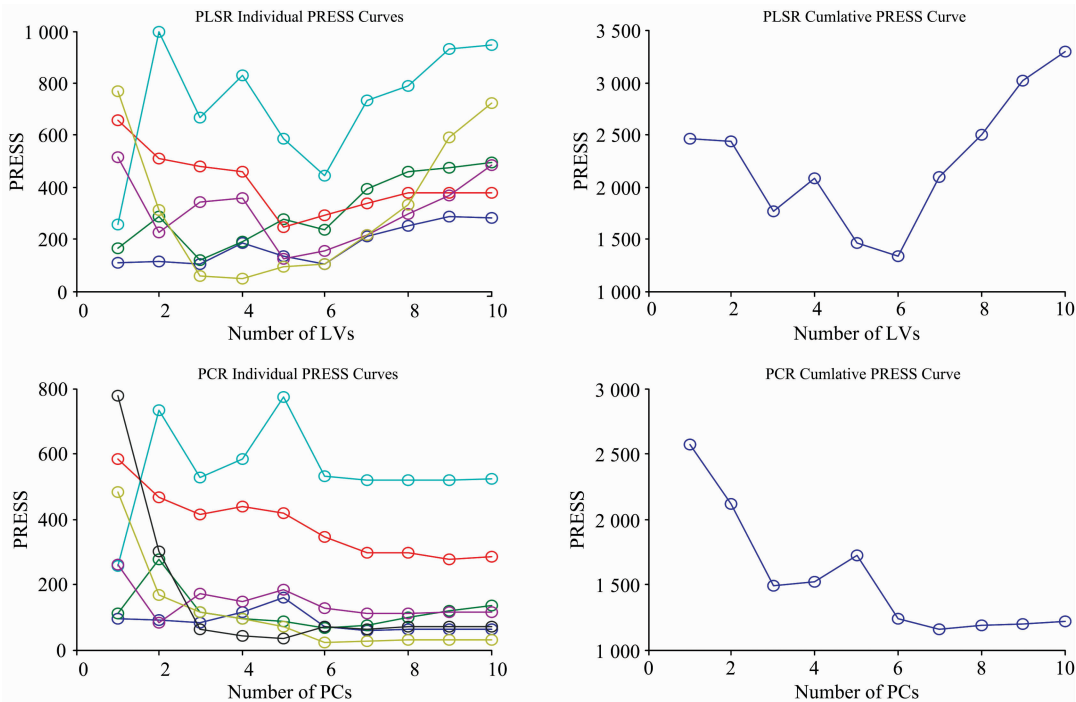


图 3 PRESS 与主成分数/隐含变量的关系

Fig. 3 The relationship between PRESS and principal component/latent variable

通过比较确定 PLS 与 PCR 的主成分数分别为 6 和 7。在此基础上所得校正模型的效果如图 4 所示，采用交叉验证方法对校正模型进行预测效果的检验。

由图 4 可以看出，两种方法预测的结果与真实值均有着很好的相关性。对于校正集数据而言，在交叉验证过程中，PCR 法和 PLSR 法的校正集决定系数分别为 $R_{\text{PCR}}^2 = 0.933 2$ 和 $R_{\text{PLS}}^2 = 0.976 4$ 。因此 PLSR 方法得到的预测值更加集中，更接近于真实值。显然，PLSR 法要比 PCR 法更加精确。

有一小部分预测值远离真实值，说明荧光光谱的采集过程中引入了一定的误差。此外，实际水样 COD 理化值检测过程中也可能引入误差，从而造成化学测量值与实际值之间

表 1 COD 标准液的预测值对比
Table 1 Comparison of prediction results of COD standard solution

序号	标准液 COD 理化值 /($\text{mg} \cdot \text{L}^{-1}$)	标准液 COD 预测值/($\text{mg} \cdot \text{L}^{-1}$)	
		PCR 预测值	PLSR 预测值
1	7.5	9.96	7.53
2	27.5	28.07	27.69
3	47.5	47.16	47.87
4	70	78.40	70.10
5	90	86.62	89.86

由图 3 可知，两种算法的主成分 (PC) 数目均设为 1~10，对于 PLSR 方法而言，PRESS 在达到最低后又开始上升，且在主成分数为 6 时达到了最小值。表明了此后加入的主成分是与样品无关的噪声主成分，使得模型产生了过拟合，从而使模型的预测能力变弱。

出现较大的偏差。

为了验证模型效果，将检验集的 15 组荧光发射光谱数据分别代入两模型中进行预测检验，得到的基于 PCR 和 PLSR 模型的实际水样 COD 的预测值如表 2 所示。

PCR 模型的预测均方根误差为 $\text{RMSEP}_{\text{PCR}} = 3.520 4 \text{ mg} \cdot \text{L}^{-1}$ ，PLSR 预测均方根误差为 $\text{RMSEP}_{\text{PLS}} = 2.092 2 \text{ mg} \cdot \text{L}^{-1}$ 。对比两算法所得 RMSEP 的值可看出 PLS 法得到的模型要优于 PCR 法得到的模型。其中，PCR 法的决定系数 $R_{\text{PCR}}^2 = 0.919 0$ ，而 PLSR 算法的决定系数 $R_{\text{PLS}}^2 = 0.940 2$ 。预测结果如图 5 所示。

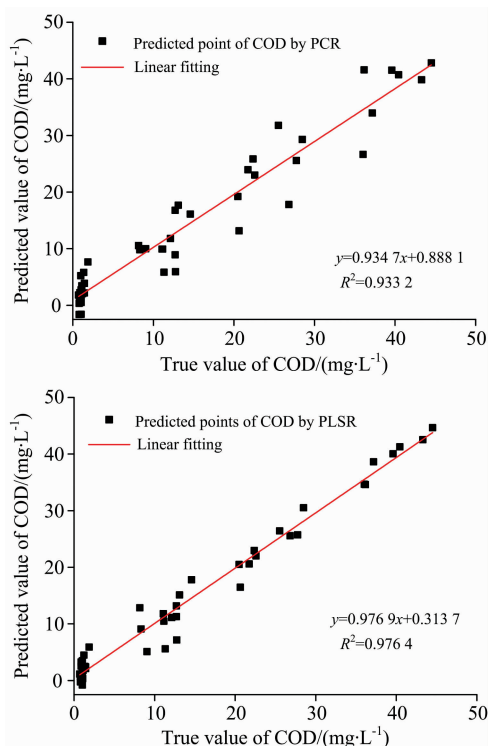


图 4 实际水样光谱数据的 PCR 和 PLSR 模型效果对比

Fig. 4 Comparison of PCR and PLSR modeling results of actual water samples spectral data

表 2 实际水样的检验集 COD 预测结果

Table 2 Results of predicted COD values of actual water samples

序号	检验集 COD 理化值 /($\text{mg} \cdot \text{L}^{-1}$)	检验集 COD 预测值/($\text{mg} \cdot \text{L}^{-1}$)	
		PCR 预测值	PLSR 预测值
1	0.64	-2.39	0.11
2	1.12	2	2.23
3	1.28	2.93	1.82
4	1.12	4.67	4.98
5	25.6	34.65	35.77
6	12.95	7.54	7.46
7	33.9	43.97	41.03
8	25.45	24.85	26.19
9	1.24	4.01	1.55
10	1.16	-1.36	1.67
11	0.76	4.16	3.57
12	2.72	8.02	7.63
13	27.1	35.13	33.03
14	11.75	5.04	7.82
15	35.4	40.01	38.56

2.3 本方法与传统方法的检测效果对比

为了检验本研究所建模型的预测效果,重新配制 COD 标准溶液和采集地表水样本,实验室环境下分别采用标准方法和实际水样的 PLSR 模型对实验水样的 COD 进行检测,进而得到水质 COD 标准化验值和荧光发射光谱数据的 PLSR 模型预测分析值的比较结果,如表 3 所示。

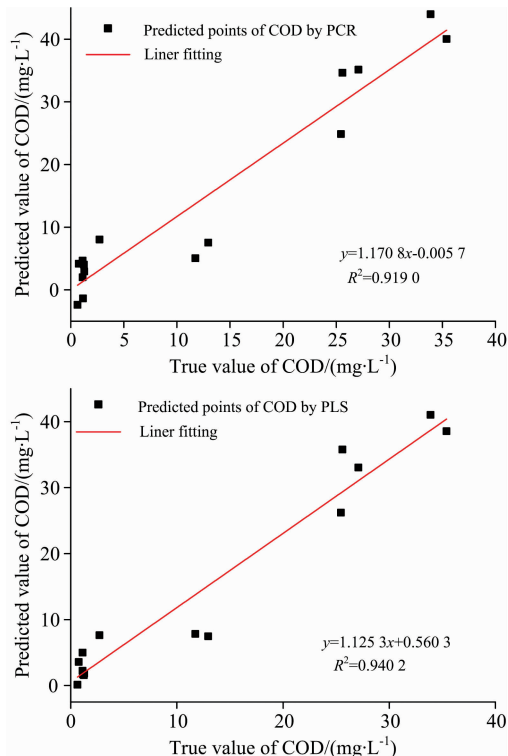


图 5 实际水样检验集的预测结果

Fig. 5 Prediction results of validation data of actual water samples

表 3 本方法与传统方法的检测效果对比

Table 3 Comparison between the presented method and the traditional method

检测方法	标准检测法 /($\text{mg} \cdot \text{L}^{-1}$)	本文方法 (PLSR)	相对误差 /%
标准液 1	0.33	0.42	27.27
标准液 2	400	334.6	-16.35
地表水 1	5.49	6.58	19.85
地表水 2	13.26	12.78	-3.62
地表水 3	4.75	5.42	14.11
地表水 4	23.58	24.21	2.67

由表 3 可知,基于荧光发射光谱预测模型所得的水质 COD 预测值与标准方法所得化验值相比能较好的反映出各水样 COD 的变化情况。对于 COD 标准溶液样本,预测值的相对误差稍大,原因为预测模型选用的是基于实际水样的荧光发射光谱数据的 PLSR 模型,而相对于实际水样,标准液水样所含物质单一,采集到的谱线变化规律可能不同于实际水样的谱线变化规律,因此在预测过程中可能出现预测误差稍大的现象。

此外,对于 COD 浓度为 $400 \text{ mg} \cdot \text{L}^{-1}$ 的标准溶液而言,预测值出现了较大的偏差,这是由于标准溶液的浓度效应导致了荧光强度降低,使得荧光强度值与实验水样的 COD 浓度值不再呈现线性关系。对于实际水样来说,预测值可以较好的反映出水质 COD 的变化情况,且与标准理化值相比具

有相对误差较小,这说明基于荧光发射光谱数据所建 PLSR 模型具有较好的预测能力。

基于荧光发射光谱数据的 PLSR 模型具有较高的预测能力和较强的适应性,可以快速、准确的检测出水质 COD。

3 结 论

分别以 COD 标准溶液和实际水样为研究对象,分别建立和验证了基于单激发波长下荧光发射光谱数据的 PCR 和 PLSR 模型,针对特定水样,将不同模型的预测效果进行了对比。得到如下结论:

(1)通过实验证明了 PCR 和 PLSR 方法用于建立水质

COD 的荧光发射光谱数据模型的可行性;表明两类模型均可以合理表征荧光强度与水质 COD 浓度之间的关系,且 PLSR 模型的预测效果更优。

(2)本方法可用于检测有机污染物浓度较低的水体,有机物浓度较高时采用本文方法时检测误差会变大。

(3)本方法利用特定激发波长下的荧光发射光谱进行 COD 分析,较三维荧光光谱(EEM)更加简单易行,可为水质 COD 的光学分析方法提供一种新的思路。

(4)实际水体组分和所处环境复杂,对建模的适应性、水体含有荧光敏感物质时及荧光光谱受水体环境影响对预测模型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] Kumar A, Jain S K. Analytical Chemistry Letters, 2016, 6(6): 894.
- [3] Liu Rongxia, Chen Liangliang, Zhang Hongyan, et al. Optoelectronics Letters, 2013, 9(3): 225.
- [4] Dong Ming, Sui Yue, Li Guolin, et al. Optoelectronics Letters, 2015, 11(6): 469.
- [5] CHEN Ying, DI Yuan-jian, TANG Xin-liang, et al(陈颖, 邸远见, 唐心亮, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(7): 2176.
- [6] ZHAO Xian-de, CHEN Xiao, DONG Da-ming(赵贤德, 陈肖, 董大明). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(9): 2907.
- [7] 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.
- [8] Dahlbacka J, Nyström J, Mossing T, et al. Spectral Analysis Reviews, 2014, 2(4): 19.
- [9] Zheng Hao, Yan Zhishen, Chen Jianfang, et al. Sustainability, 2018, 10(5): 1444.
- [10] 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 Emission Spectra

ZHOU Kun-peng¹, LIU Shuang-shuo¹, CUI Jian¹, ZHANG Hong-na¹, BI Wei-hong^{2*}, TANG Wei²

1. College of Engineering, 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 The detection of parameters for water quality with spectral technique is a research hotspot at present. This paper proposes a method for the determination of chemical oxygen demand (COD) based on the fluorescence emission spectrum. Two groups of experimental samples are provided in the experiment, among which the fundamental group is 20 COD standard solutions, and the remaining 63 are actual water samples as the other group. Rapid digestion spectrophotometry is utilized to detect the COD of experimental samples. Three dimensional fluorescence spectrophotometer is used to collect the fluorescence emission spectra of the water samples at EX=275 nm (all the range of fluorescence emission spectra are EM=325~450 nm), then the data of fluorescence emission spectra of two kinds of water samples are processed and modeled. Principal component regression (PCR) and partial least squares regression (PLSR) are utilized to establish the prediction models based on fluorescence emission data respectively, and the effects of the models are compared. In order to verify the feasibility of the proposed method and the prediction ability of the model, the results of the PLSR mode are compared with the standard method. The comparison results show that, for the COD standard solution, when the number of principal component of PLSR and PCR is 5 and 8 respectively, the optimal results are obtained for both models, of which the determination coefficients of the correction model are $R_{PLS}^2=0.999$

9 and $R_{\text{PCR}}^2 = 0.9897$, respectively. The prediction error of validation set data in the calibration model is less than 10%, and the PLSR model is better than the PCR model. While for the actual water samples, when the number of principal component of PLSR and PCR is 6 and 7 respectively, the cross-validation effect of the correction model is the best. Among them, root mean square error of cross-validation of the PLSR method is $\text{RMSECV}_{\text{PLS}} = 0.9322 \text{ mg} \cdot \text{L}^{-1}$, while for the PCR algorithm, $\text{RMSECV}_{\text{PCR}} = 0.9764 \text{ mg} \cdot \text{L}^{-1}$. For the validation set, the determination coefficient of PLSR is 0.9402, while for PCR method, it is 0.9190. It shows that PLSR method has better prediction effect. Consequently, the PLSR model based on fluorescence emission spectrum data has high prediction ability and strong adaptability, which can detect water COD quickly and accurately. Through the comparison of the method proposed in this paper and the traditional detection, we can see that proposed method can be used to detect the water with low concentration of organic pollutants, however, the detection error will increase if the concentration of organic pollutants in the detected water is high. This paper provides a new design idea for the research and development of water quality detection optical sensor.

Keywords Water quality detection; Chemical oxygen demand (COD); Fluorescence spectrum; Emission spectra; Prediction model

(Received Nov. 5, 2019; accepted Jan. 10, 2020)

* Corresponding author

敬告读者——《光谱学与光谱分析》已全文上网

从 2008 年第 7 期开始在《光谱学与光谱分析》网站 (www.gpxygpfx.com)“在线期刊”栏内发布《光谱学与光谱分析》期刊全文, 读者可方便地免费下载摘要和 PDF 全文, 欢迎浏览、检索本刊当期的全部内容; 并陆续刊出自 2004 年以后出版的各期摘要和 PDF 全文内容。2009 年起《光谱学与光谱分析》每期出版日期改为每月 1 日。

《光谱学与光谱分析》期刊社