

## 基于紫外-可见光谱的水质 TOC 定量分析方法

李庆波<sup>1</sup>, 魏源<sup>1</sup>, 崔厚欣<sup>2</sup>, 冯浩<sup>2</sup>, 郎嘉晔<sup>2</sup>

1. 北京航空航天大学仪器科学与光电工程学院, 精密光机电一体化技术教育部重点实验室, 北京 100191
2. 河北先河环保科技股份有限公司, 河北 石家庄 050035

**摘要** 地表水资源安全关系到国民健康、生态环境稳定和经济可持续发展, 具有重要战略意义。总有机碳 (TOC) 是一种衡量水体中有机物含量的综合指标, 其在水环境监管和治理中具有重要价值。传统检测方法通过高温催化氧化测定水样中 TOC 含量具有耗时较长、操作复杂的局限性, 紫外-可见光谱技术具有检测速度快、操作简单的优势, 因而在水质在线检测中具有较好的应用前景。国内外对地表水中 TOC 浓度的在线检测目前大多采用与 COD 浓度间的相关关系进行间接推算得到, 这类方法对水体成分的稳定性要求较高。相比于常规的间接推算方法, 采用光谱定量分析方法建立 TOC 与紫外-可见光谱间的分析模型具有更好的鲁棒性和分析精度, 便于实现水质无人值守在线监测。实验配置了 TOC 样本溶液, 设计了为期两天的实验, 在 4 个时间段采集得到样品光谱数据集 (分别记为 D1, D2, ..., D6)。首先, 通过分组实验将 D1 作为训练集建立 TOC 偏最小二乘 (PLS) 回归模型, 预测同一时间段测试集 D2 的 TOC 浓度, 得到平均绝对相对误差 (MAPE) 不超过 0.78%, 表明建立的 TOC 定量分析模型具有较高的精度。然后, 为验证 PLS 建立的 TOC 模型对仪器状态变化的鲁棒性, 选择不同时间段采集的光谱数据分别作为训练集和测试集, 进行不同仪器状态交叉实验, 4 组实验中测试集样品 TOC 浓度预测值的 MAPE 分别为 3.82%, 3.75%, 3.43% 和 0.98%。实验表明, 采用 PLS 算法建立的 TOC 紫外-可见光谱定量分析模型具有较好的分析精度和鲁棒性, 分组实验和不同仪器状态交叉实验中预测浓度的 MAPE 均不超过 3.82%, 优于常规的间接推算法。此外, 建立的光谱定量分析模型不依赖 COD 与 TOC 间的推算关系, 因此在水环境变化时较常规推算方法具有更好的适应能力。最后, PLS 算法建模过程简单, 运算速度快, 为浸入式在线检测设备的开发和维护提供了便利。

**关键词** 紫外-可见光谱; TOC; 偏最小二乘回归; 水质在线检测

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)02-0376-05

### 引言

总有机碳 (total organic carbon, TOC) 是从水体中有机成分含碳量的角度反映水体中有机物污染程度的重要指标。现有的 TOC 在线分析仪采用高温催化氧化法将水样中的有机物消解为二氧化碳, 通过红外检测设备计量二氧化碳的浓度, 然后根据不同浓度的 TOC 标准液消解产生的 CO<sub>2</sub> 浓度建立 TOC-CO<sub>2</sub> 标准曲线, 进而推算待测水样中 TOC 浓度<sup>[1]</sup>。紫外-可见光谱技术作为一种新兴技术应用在水质监测中, 具有检测快速、不需要添加化学试剂、无二次污染、检测设备结构简单、便于制作小型化原位在线检测设备并且维护成本较低等优势, 得到水环境保护领域的广泛研究<sup>[2-4]</sup>。

使用紫外-可见光谱技术对水中 TOC 定量分析目前大多通过 COD 浓度间接计算得到。通常认为在同一稳定的水体中 COD 浓度和 TOC 浓度间存在相关关系<sup>[5]</sup>, 有学者使用化学分析方法或专用分析仪器分别测定水样中 COD 和 TOC 的浓度, 从而建立起两种参数间的回归关系, 然后使用紫外-可见光谱法计算水样中的 COD 浓度进而推算 TOC 浓度<sup>[6]</sup>。但这种方法的适用前提是实验水体成分相对稳定, 一旦被测水体受到降雨或季节性水文变化影响导致水体成分改变, 使用推算算法得到的 TOC 浓度可能会有较大偏差, 需要重新校正 COD 和 TOC 间回归曲线的参数以保证 TOC 定量分析精度。

目前鲜有不经化学处理直接用水样的紫外-可见光谱对样品中 TOC 浓度进行定量分析的相关研究。本文使用偏最小二乘 (partial least squares, PLS) 回归模型建立被测水样

收稿日期: 2021-01-19, 修订日期: 2021-04-10

基金项目: 国家自然科学基金项目 (61575015) 和河北省重点研发计划资源与环境专项项目 (19271704D) 资助

作者简介: 李庆波, 女, 1975 年生, 北京航空航天大学仪器科学与光电工程学院副教授 e-mail: qbleebuaa@buaa.edu.cn

光谱与 TOC 浓度间的定量模型，可以有效避免因 COD 与 TOC 推算关系变化导致的 TOC 定量不准确的问题，提高 TOC 定量分析方法在不同水环境中的适用性和鲁棒性。

## 1 实验部分

### 1.1 样本

根据国标 TOC 标准液配置方法<sup>[1]</sup>，实验使用天津市光复科技公司生产的分析纯邻苯二甲酸氢钾试剂配置有机碳标准溶液，试剂纯度为 100% ± 0.05%。使用花潮高科公司的精度为 0.001 g 的电子分析天平称量 2.125 g 邻苯二甲酸氢钾，将其溶于蒸馏水中并移至 1 000 mL 容量瓶，用蒸馏水稀释定容，配置得到浓度为 1 000 mg · L<sup>-1</sup> 的 TOC 标准母液，将 TOC 标准母液稀释，配置成 25~150 mg · L<sup>-1</sup> 范围内浓度梯度均匀的 35 个 TOC 样品，进行分组实验，训练集和测试集分别记为 D1 和 D2。为进一步验证 TOC 定量模型的有效性，使用前述 1 000 mg · L<sup>-1</sup> 的 TOC 标准母液稀释配置浓度为 37, 49, 61, 76, 88, 109, 127 和 145 mg · L<sup>-1</sup> 的 TOC 样品作为验证集。在前述 35 个样品光谱采集完成后，关机半小时重新开机并初始化仪器状态，仪器参数与分组实验保持一致，然后采集验证集中 8 个样品的光谱，将光谱数据集记为 D3。在第二实验日重新采集上述 TOC 样品光谱形成对照实验，第二实验日的分组实验训练集、测试集和验证集分别记为 D4, D5 和 D6，两个实验日共计 86 个 TOC 样品。实验中，使用同一份 TOC 标准母液进行稀释，配置一定程度上可以减小实验试剂纯度对样品 TOC 浓度真值的影响。

### 1.2 仪器

采用河北先河环保科技股份有限公司研制的 XHWPFT 型浸入式水质在线分析设备作为主要实验设备。该设备的光程长为 2 mm，使用氙灯作为光源，光谱仪模块可以检测 256 个波段的吸光度值，设置光谱仪扫描波段为 188~722 nm，则样品光谱的平均波长间隔为 2 nm。为减小随机误差，每个浓度的 TOC 样品连续扫描 20 次，扫描间隔为 10 s，将测得的 20 条光谱取平均，作为该 TOC 样品对应的光谱。所有实验均采用相同的扫描方法，并且在相同的实验参数下进行。

### 1.3 方法

PLS 回归是一种基于多波长吸光度值的定量回归方法，通过将光谱矩阵投影到隐变量空间，利用隐变量间接地将光谱数据和被测浓度联系起来，可以有效解决不同波长光谱吸光度数据间的多重共线性问题<sup>[7]</sup>。

使用 PLS 建立 TOC 定量模型的原理是将中心化后的光谱响应矩阵  $\mathbf{X}_0$  和对应浓度矩阵  $\mathbf{Y}_0$  同时进行主成分分解

$$\mathbf{X}_0 = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

$$\mathbf{Y}_0 = \mathbf{UQ}^T + \mathbf{G} \quad (2)$$

其中  $\mathbf{T}$  和  $\mathbf{U}$  分别为  $\mathbf{X}_0$  和  $\mathbf{Y}_0$  的得分矩阵，而  $\mathbf{P}$  和  $\mathbf{Q}$  分别为  $\mathbf{X}_0$  和  $\mathbf{Y}_0$  相应的载荷矩阵， $\mathbf{E}$  和  $\mathbf{G}$  分别为拟合矩阵  $\mathbf{X}_0$  和  $\mathbf{Y}_0$  时产生的拟合误差。用光谱数据  $\mathbf{X}_0$  的线性组合表达得分矩阵  $\mathbf{T} = \mathbf{X}_0 \mathbf{W}$ ，然后建立  $\mathbf{T}$  和 TOC 浓度矩阵  $\mathbf{Y}_0$  间的回归关系，将水样光谱响应与目标水质参数浓度联系起来，其计算过程如式(3)和式(4)

$$\mathbf{Y}_0 = \mathbf{TC}^T + \mathbf{F} = \mathbf{X}_0 \mathbf{WC}^T + \mathbf{F} = \mathbf{X}_0 \mathbf{B} + \mathbf{F} \quad (3)$$

$$\mathbf{C}^T = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y}_0, \mathbf{B} = \mathbf{WC}^T \quad (4)$$

其中  $\mathbf{W} = [\omega_1, \omega_2, \dots, \omega_p]$  为权重向量矩阵， $p$  为主成分个数。 $\mathbf{C}$  为  $\mathbf{X}_0$  的各主成分和  $\mathbf{Y}_0$  间的关联矩阵， $\mathbf{B}$  为回归系数矩阵， $\mathbf{F}$  为矩阵  $\mathbf{T}$  对  $\mathbf{Y}$  的拟合残差。最后根据训练得到的回归系数矩阵  $\mathbf{B}$ ，得到未知样本的 TOC 浓度预测值  $\hat{\mathbf{Y}}$  为

$$\hat{\mathbf{Y}} = [\mathbf{1}, \mathbf{X}_{\text{new}}] \begin{bmatrix} \mathbf{Y}_{\text{mean}} - \mathbf{X}_{\text{mean}} \mathbf{B} \\ \mathbf{B} \end{bmatrix} \quad (5)$$

式(5)中， $\mathbf{1}$  为各元素均为 1 的列向量， $\mathbf{X}_{\text{new}}$  为未知样品的原始光谱矩阵， $\mathbf{X}_{\text{mean}}$  和  $\mathbf{Y}_{\text{mean}}$  分别为训练集样品光谱矩阵的均值和对应浓度矩阵均值。

## 2 结果与讨论

### 2.1 光谱特征曲线及模型评价指标

使用紫外-可见光谱技术进行水质参数定量分析时，根据目标水质参数的光谱吸收特性选择恰当的光谱分析区间可以有效减小不同水质参数定量分析时的交叉干扰。水体中的有机物在 200~320 nm 的波段有较强的特征吸收，本文选择该波段作为 TOC 定量分析的光谱分析区间，不同 TOC 浓度的样品光谱如图 1 所示，图中不同颜色的曲线代表不同 TOC 浓度样品的光谱。

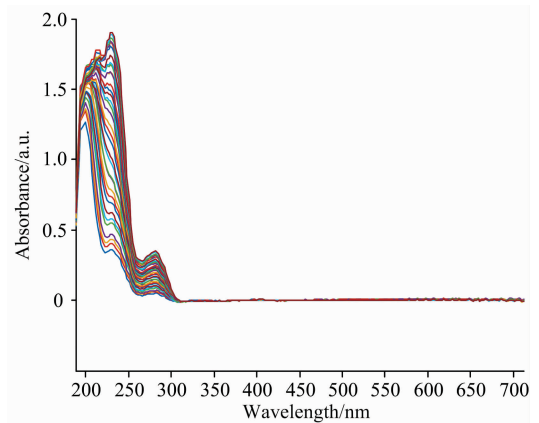


图 1 不同浓度 TOC 样品吸光度光谱  
Fig. 1 Absorbance spectra of TOC samples with different concentrations

为评价模型的定量分析精度，使用相对误差、相对误差绝对值的平均数(mean absolute percentage error, MAPE)和预测均方根误差(predict root mean square error, RMSEP)作为衡量建立的 TOC 定量模型精度评价指标，RMSEP 和 MAPE 的计算方法如式(6)和式(7)

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (7)$$

其中， $n$  为测试集样品数， $y_i$  为测试集中第  $i$  个样品的 TOC 实际浓度， $\hat{y}_i$  为对应样品的 TOC 预测浓度， $\bar{y}$  为样品预测浓

度的均值。RMSEP 和 MAPE 作为综合指标用来评价定量模型对所有测试样品 TOC 浓度预测的准确性和稳定性。

## 2.2 TOC 定量模型分析结果

### 2.2.1 单日样品分组实验与验证实验预测结果

首先进行分组实验,使用 spxy 算法<sup>[8]</sup>将 35 个不同浓度的 TOC 样品划分为训练集 D1 和测试集 D2,训练集样品容量为 20 个,测试集样品容量为 15 个,使用 PLS 建立水样 TOC 浓度定量分析模型。对训练集样品使用留一交叉验证法用来确定 PLS 模型的最佳主成分数,使用 TOC 浓度预测值的残差平方和(predictive residual error sum of squares, PRESS)作为评价指标,选择 PRESS-主成分数曲线的第一个极小值点对应的主成分个数为最佳主成分数。采用不同主成分数预测测试集样品对应的 PRESS 值曲线如图 2 所示,最佳主成分数预测测试集样本的相对误差如图 3 所示。

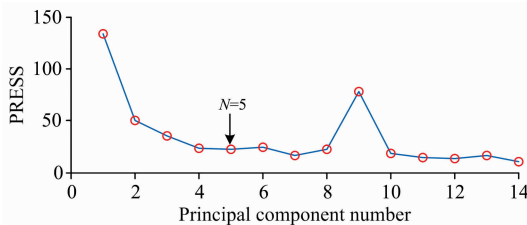


图 2 采用不同主成分数预测测试集样品浓度的 PRESS 曲线  
Fig. 2 PRESS curve of the predicted concentration of the test set samples using different principal component numbers

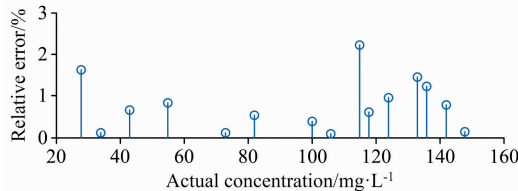


图 3 采用最佳主成分数预测测试集样品浓度的相对误差  
Fig. 3 Relative error of sample predicted concentration in test set using the optimal principal component number

由图 2 可知,最佳主成分数选为 5 时 PLS 模型对分组实验测试集样品 TOC 浓度定量精度最高,TOC 浓度预测值最大相对误差不超过 2.22%,此时 MAPE 为 0.78%,RMSEP 为 1.06。随着主成分个数增大,光谱中的噪声和其他与 TOC 浓度无关的信息随之进入模型,导致 PLS 模型定量精度有所下降。

由于实验设备每次开机会产生不同的随机误差,进而影响定量模型的预测精度,为进一步评价 TOC 定量模型的鲁棒性,在不同仪器状态下采集了验证集样本光谱 D3,用分组实验建立的 TOC 定量模型预测 D3 样品的 TOC 浓度,构成不同仪器状态验证实验,TOC 浓度预测结果如表 1 所示。

由表 1 结果可知,不同仪器状态下验证集样本 TOC 浓度预测值的 MAPE 为 0.98%,RMSEP 为 1.14。可见模型对不同仪器状态有较好的鲁棒性和预测精度。

表 1 第一实验日不同仪器状态验证实验 TOC 浓度预测结果  
Table 1 TOC concentration prediction results of validation experiment with different instrument states in the first experimental day

实际浓度/ ( $\text{mg} \cdot \text{L}^{-1}$ )	预测浓度/ ( $\text{mg} \cdot \text{L}^{-1}$ )	相对误差/ %
37.00	37.45	1.21
49.00	49.52	1.07
61.00	61.57	0.93
76.00	75.95	0.06
88.00	88.63	0.72
109.00	106.46	2.33
127.00	127.59	0.46
145.00	146.55	1.07

### 2.2.2 不同实验日样品交叉实验预测分析

为进一步评价 TOC 定量模型的精度和鲁棒性,使用 spxy 算法将第二实验日采集的 35 个不同浓度的 TOC 样品光谱同样分为 20 个样品的训练集 D4 和 15 个样品的测试集 D5,类似的在仪器关机半小时后重新开机,采集 8 个不同浓度的 TOC 样品光谱记为验证集 D6。然后进行不同实验日不同仪器状态交叉实验,具体分为用训练集 D1 建立的模型分别预测 D5 和 D6 样品的 TOC 浓度(记为实验 1 和实验 2)、用训练集 D4 建立的模型分别预测 D2 和 D3 样品 TOC 浓度(记为实验 3 和实验 4),上述所有实验结果汇总整理如表 2 所示。

表 2 不同实验日不同仪器状态交叉实验 TOC 浓度预测结果  
Table 2 TOC concentration prediction results of cross experiment with different instrument status on different experimental days

实验类型	预测浓度最大 相对误差/%	MAPE/%	RMSEP
实验 1	7.19	3.82	3.75
实验 2	15.91	3.75	2.99
实验 3	6.48	3.43	3.91
实验 4	2.33	0.98	1.14

由表 2 可知,PLS 建立的 TOC 定量分析模型在不同仪器状态交叉实验中具有较好的预测精度和鲁棒性,四组实验中 MAPE 均不超过 3.9%,除实验 2 的测试集 D5 中 TOC 实际浓度接近本文 TOC 量程下限的样品外,其余样品 TOC 预测值的最大相对误差均不超过 7.2%。实验 2 中 28 和 31  $\text{mg} \cdot \text{L}^{-1}$  样品 TOC 预测值相对误差较大,其原因为第二实验日上述样品光谱较第一实验日相同浓度样品光谱吸收较弱,且存在一定的基线漂移,两个实验日上述浓度的 TOC 样品光谱对比如图 4 所示。除上述两个样品外,实验 2 中其余样品 TOC 浓度预测值最大相对误差均不超过 5%,且实验 2 的 MAPE 和 RMSEP 分别为 3.75% 和 2.99。实验 2 中测试集样品 TOC 浓度预测值相对误差如图 5 所示。

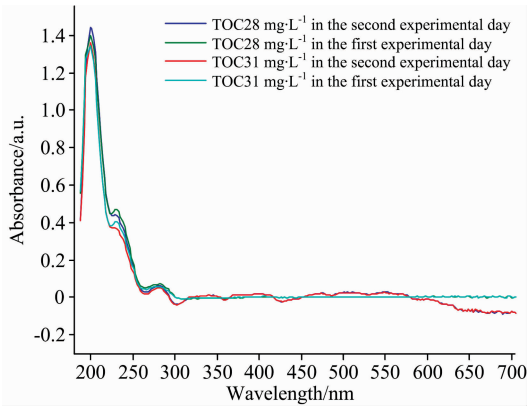


图 4 不同实验日 TOC 实际浓度为 28 和 31 mg · L<sup>-1</sup> 光谱对比  
 Fig. 4 Spectral comparison of TOC actual concentration of 28 and 31 mg · L<sup>-1</sup> on different experimental days

### 3 结 论

TOC 是反映水体受有机物污染程度的重要水质指标，

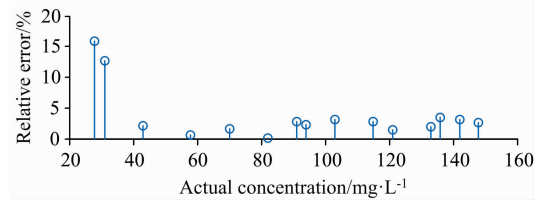


图 5 实验 2 中 TOC 浓度预测值相对误差  
 Fig. 5 Relative error of predicted TOC concentration in experiment 2

本文基于浸入式水质在线分析设备，使用 PLS 模型直接建立水体 TOC 浓度与紫外-可见吸收光谱间的定量分析模型，提高了 TOC 预测精度和鲁棒性，在一定程度上解决了目前通过测定水体中 COD 浓度间接推算 TOC 浓度引起的稳定性和适用性差的问题。实验结果表明，本文模型对 25~150 mg · L<sup>-1</sup> 范围内样品 TOC 预测浓度值的 MAPE 不超过 3.82%，通过不同仪器状态交叉实验充分验证了模型的鲁棒性和适用性。

### References

[1] National Standard of the People's Republic of China(中华人民共和国国家标准). GB/T 32116—2015, Determination of Total Organic Carbon (TOC) in Industrial Circulating Cooling Water[循环冷却水中总有机碳(TOC)的测定]. Beijing: Standardization Administration of the P. R. C. (北京: 国家标准化管理委员会), 2015.

[2] ZHAO You-quan, LI Xia, LIU Xiao, et al(赵友全, 李霞, 刘潇, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(11): 3592.

[3] Lepot M, Torres A, Hofer T, et al. Water Research, 2016, 101: 519.

[4] Carré E, Pérot J, Jauzein V, et al. Water Science and Technology, 2017, 76(3): 633.

[5] Lee J, Lee S, Yu S, et al. Environmental Monitoring and Assessment, 2016, 188(4): 252.

[6] Wu X, Tong R, Wang Y, et al. Sensors, 2019, 19(9): 2153.

[7] Guo Y, Liu C, Ye R, et al. Applied Sciences, 2020, 10(19): 6874.

[8] Galvao R K H, Araujo M C U, José G E, et al. Talanta, 2005, 67(4): 736.

# Quantitative Analysis of TOC in Water Quality Based on UV-Vis Spectroscopy

LI Qing-bo<sup>1</sup>, WEI Yuan<sup>1</sup>, CUI Hou-xin<sup>2</sup>, FENG Hao<sup>2</sup>, LANG Jia-ye<sup>2</sup>

1. Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

2. Hebei Sailhero Environmental Protection Hi-Tech Co., Ltd., Shijiazhuang 050035, China

**Abstract** The safety of surface water resources is of great strategic significance. It is related to national health, ecological environment stability and sustainable economic development. Total organic carbon (TOC) is a comprehensive index to reflect the content of organic matter in water. Hence, it has significant value in water environment supervision and treatment. This method is time-consuming and complex. UV-Vis spectroscopy technology has the advantages of fast detection speed and simple operation. Therefore it has a good application prospect in online detection of water quality. At present, the online detection methods of TOC in surface water mostly are indirectly calculated at home and abroad. These methods depend on the correlation between the concentration of COD and TOC, and they require high stability of water composition. Compared with the indirect calculation methods, the spectral quantitative analysis method has better robustness and accuracy. Moreover, this method is convenient for realizing unattended online monitoring of water quality. The experiment was equipped with TOC sample solutions, and a two-day experiment was designed. Six spectral data sets of the samples (denoted as D1, D2, ..., D6) were collected in 4 time periods. Firstly, D1 was used as the training set to establish a partial least squares (PLS) regression model in the group experiment. This model was used to predict the TOC concentration of D2, and the mean absolute percentage error (MAPE) was less than 0.78%. In addition, D1 and D2 were collected in the same period. The results show that the established TOC quantitative analysis model has high accuracy. Then, to verify the robustness of the TOC model established by the PLS method to the change of instrument state, the spectral data collected in different periods were selected as the training set, the test set and the validation set. Furthermore, the cross experiments of different instrument states were performed. The MAPE of the predicted TOC concentration in the four experiments were 3.82%, 3.75%, 3.43% and 0.98%, respectively. The results show that the UV-Vis spectroscopy quantitative analysis model of TOC established by the PLS algorithm has good accuracy and robustness. The MAPE of predicted concentration in the grouping experiment and cross experiments of different instrument states are all less than 3.82%. These results are better than the conventional indirect calculation method. Moreover, the established spectral quantitative analysis model does not depend on the calculation relationship between COD and TOC. Thus, it has better adaptability than the conventional indirect calculation method when the water environment changes. Finally, the PLS algorithm has the advantages of a simple modeling process and fast operation speed. It provides convenience for the development and maintenance of submersible online detection equipment.

**Keywords** UV-Vis spectroscopy; TOC; PLS regression; Water quality online detection

(Received Jan. 19, 2021; accepted Apr. 10, 2021)