

地表水亚硝酸盐氮浓度紫外可见光谱检测方法的基础研究

李庆波¹, 何林倩¹, 崔厚欣², 郝龙腾², 孙冬生²

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 随着我国经济的高速发展, 地表水污染问题日趋严重, 实现地表水水质连续监测对于保障人类健康和保护环境至关重要。亚硝酸盐氮浓度是水质评估的一项重要指标, 污染的水体对人畜及水产构成很大威胁。利用紫外可见吸收光谱检测有机污染物已经成为水质检测的重要方法。国内关于紫外可见光谱法检测亚硝酸盐氮的文献并不多, 一般采用对水样进行化学前处理然后再利用紫外分光光度计预测浓度。这种方法不仅检测步骤繁琐, 耗时耗力, 对环境进一步造成污染, 而且无法实现实时连续检测。无化学预处理的基于紫外可见吸收光谱法的亚硝酸盐氮浓度检测鲜有报道。采用紫外可见光谱法针对地表水水质的无人值守自动连续监测方法开展基础研究。配置了亚硝酸盐氮标准溶液, 设计了为期三天的实验, 每天分别测量所有样本的紫外可见光谱(记为 D1 组, D2 组, D3 组)。首先将前两天的样本分别进行偏最小二乘回归法(PLSR)建模, 利用交互验证得到平均绝对相对误差(MAPE)分别为 1.19% 和 1.85%, 这说明 PLSR 模型具有优良的预测精度。其次, 为了验证 PLSR 模型在不同条件下的适应性, 取 D1, D2 的实验数据进行互相预测分析。两天互相预测的 MAPE 分别为 3.36% 和 4.51%, 低于 5%, 说明 PLSR 模型具有良好的鲁棒性。最后, 将 D1, D2 的全部样本用于建立最终的 PLSR 实测模型, D3 的样本作为测试集。测试集的 MAPE 为 2.19%。结果表明, 基于紫外可见光谱分析技术的 PLSR 算法对溶液中亚硝酸盐氮浓度的算法检测结果的 MAPE 均控制在 5% 以下, 优于同类文献的测量精度。此外, PLSR 模型建模过程简单, 运算时间短; 建好的模型结构简单更容易移植并固化到嵌入式系统中, 为后期开发设计便携式装置带来便利。作为地表水亚硝酸盐氮浓度检测的基础研究, 可为日后地表水水质的精准快速检测提供指导。

关键词 紫外可见光谱; 亚硝酸盐氮; 偏最小二乘回归; 水质; 在线检测

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)04-1127-05

引言

地表水是人们生活和生产用水的主要来源, 随着我国经济发展和城镇化的加剧, 地表水的污染严重威胁社会用水安全, 因此加强地表水的监测迫在眉睫^[1]。“十三五”期间, 国家出台了“河长制”、“水十条”、以及《水污染防治法》等一系列政策, 推进水污染防治工作的发展。

亚硝酸盐氮是氮循环的中间产物, 对人和牲畜具有较大的毒性。当其进入血液时, 会将正常的携带氧气的血红蛋白氧化为不具备携带氧气能力的高铁血红蛋白, 进而导致组织缺氧。在 pH < 6.5 时, 它会与仲氨反应形成具有强致癌性的亚硝胺。

紫外-可见光谱法具有检测速度快、维护成本低、无二次

污染等优点, 近年来被广泛用于水质监测等各个领域。测定亚硝酸盐的方法主要有: 离子色谱法、气相分子吸收光谱法、流动分析法、紫外可见分光光度法等^[2-3]。离子色谱法、气相分子吸收光谱法和流动分析法适用于实验室检测, 不适合于地表水的现场连续检测。国内关于紫外可见光谱法检测亚硝酸盐氮的文献中, 多需要在酸性介质中通过添加化学试剂生成显色染料再利用紫外分光光度计预测浓度^[4]。其中国标水质亚硝酸盐氮的测定分光光度法^[5]采用的试剂盐酸萘乙二胺具有致癌性, 不受操作者欢迎。这种方法需要化学前处理的操作, 不仅检测步骤繁琐, 耗时耗力, 还会对环境进一步造成污染。此外, 这种方法也无法实现地表水水质的无人值守自动实时检测。无化学预处理的基于紫外可见吸收光谱法的亚硝酸盐氮浓度预测鲜有文献提及。

采用偏最小二乘回归法(partial least squares regression,

收稿日期: 2019-06-08, 修订日期: 2019-10-12

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

PLSR), 选择 190~500 nm 谱段建立模型, 对不同浓度的亚硝酸盐氮溶液进行分析。该方法无需添加任何化学试剂, 操作简便, 可为日后地表水无人值守自动实时检测提供参考。

1 实验部分

1.1 仪器设备与溶液配制

采用美国海洋光学公司的 USB2000+ 型号的微型光谱仪, 波长范围为 190~700 nm, 波长分辨率为 1 nm。将实验室配制的 1 000 mg · L⁻¹ 的亚硝酸钠母液(以 N 计)稀释, 配成 0.2, 0.4, 0.6, 0.8, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0 和 10.0 mg · L⁻¹ 系列标液, 分别在第一天, 第二天, 第五天用 1cm 比色皿用紫外可见光谱计多次进行光谱扫描, 扫描范围 190~700 nm, 记为 D1 组, D2 组, D3 组, 一共 42 个样本。

1.2 数学模型

采用偏最小二乘回归法(PLSR)建立模型^[6], 其原理是对浓度矩阵 Y 和对应的光谱矩阵 X 同时进行主成分分解

$$X = TP + E \quad (1)$$

$$Y = UQ + F \quad (2)$$

式(1)和式(2)中 T 和 U 是 X 和 Y 矩阵的得分矩阵, P 和 Q 分别表示 X 和 Y 矩阵的载荷矩阵, E 和 F 分别是误差矩阵。

经过主成分分解后对得到的各列正交的特征矩阵 T 和 U 矩阵进行回归:

$$U = TB \text{ 依据式(3)计算权重矩阵 } B$$

$$B = (T^T T)^{-1} T^T U \quad (3)$$

在实际的算法实现过程中, PLSR 方法将矩阵的分解和回归运算合并, 即同时分解 X 和 Y 矩阵, 并将 Y 中包含的信息传递至 X 的分解中, 使得 X 主成分直接与 Y 关联。将构建完成的模型用于未知样本预测时, 首先获取该样本(记为 X')的得分矩阵 T' , 而后基于式(3)计算预测结果, 见式(4)

$$Y = TBQ \quad (4)$$

2 结果与讨论

2.1 光谱整体特征与评价指标

水样的紫外可见吸收特征光谱在 210 nm 附近具有很强的吸收, 吸收带主要分布在 190~500 nm, 而在 500 nm 之后的吸光度接近于零。为了更清晰的分析水样的紫外可见光谱, 除去接近于零的波段, 即保留 190~500 nm 的紫外可见波段进行建模(见图 1)。

基于 PLSR 模型建立样本指标与光谱吸收率的关系, 采用均方根误差(root-mean-square error, RMSE)来衡量建模的精度, 计算公式如式(5)^[7]

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^* - y_i)^2} \quad (5)$$

其中 N 为样本个数, y_i^* 为第 i 个样本预测值, y_i 为第 i 个样本实测值, 能够反映出预测值与真值偏差平均数值。然而, 本研究的样本中, 亚硝酸盐氮的浓度梯度设置为 0.2, 0.4, 0.6, 0.8 以及 1.0~10.0 mg · L⁻¹。计算建模精度时,

RMSE 指标无法准确反映本研究中建立的模型对低浓度样本的预测能力。采用相对误差能够很好地解决这一问题, 不仅考虑预测值与真实值的误差, 还考虑了误差与真实值之间的比值, 充分反映模型对不同浓度水样的预测能力。因此, 本文选用相对误差绝对值的平均值(mean absolute percentage error, MAPE)作为模型精度的评价指标, 计算方法如式(6)^[8]

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i^* - y_i|}{y_i} \times 100\% \quad (6)$$

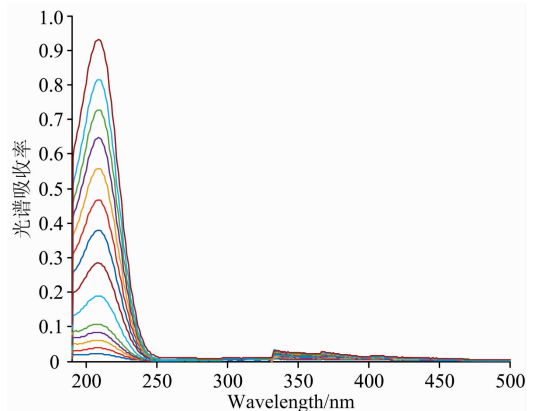


图 1 亚硝酸盐氮溶液样本的紫外可见吸收光谱

Fig. 1 Ultraviolet-visible spectrum absorption spectrum of nitrous nitrogen solution samples

2.2 PLSR 模型的可行性分析

2.2.1 单日的样本交互验证预测分析

由于实验样本个数较少, 采用留一交互验证法对数据进行建模预测。即对单日的 14 个样本, 每次用 13 个样本进行建模, 预测剩下的一个样本。重复试验 14 次, 直到所有的样本都被预测一次为止。分别取前两天的各 14 个样本, 用 14 次试验的 MAPE 作为评价指标。

D1 组的实验样本通过留一法交互验证得到最佳的主成分个数为 6, 每一个样本的相对误差范围在 -1.67%~5.00% 之间, RMSE 结果为 0.055 mg · L⁻¹, 平均绝对相对误差(MAPE)为 1.19%; 除了在 0.2 mg · L⁻¹ 处相对误差为 5.00%, 其余浓度预测结果均低于 5%。D2 组样本集通过留一法交互验证得到最佳的主成分个数为 4, 每一个样本的相对误差范围在 -5.00%~6.00% 之间, RMSE 结果为 0.075 mg · L⁻¹, 平均绝对相对误差为 1.85%; 除了在 1.0 mg · L⁻¹ 处相对误差为 6.00% 高于 5%, 其余浓度预测结果均低于 5%。此结果说明 PLSR 模型具有优良的预测精度。

2.2.2 不同日之间互相预测分析

由于实验设备每次开机产生不同的随机误差, 可能会对建模预测产生影响。为了验证 PLSR 模型在不同条件下的适应性, 取前两天的实验数据进行互相预测分析。即用一天的 14 个样本建模, 预测另一天的 14 个样本的浓度。

实验结果表明, 除了低浓度样本的预测结果的相对误差在 20% 左右, 模型在其他浓度范围内都有很高的预测精度, 两天互相预测的相对误差平均值分别为 3.36% 和 4.51%,

低于 5%，RMSE 结果分别为 0.108 和 0.083 $\text{mg} \cdot \text{L}^{-1}$ ，证明 PLSR 模型的适应性强，鲁棒性好。

2.3 PLSR 模型的实际水质预测

2.2.1 节和 2.2.2 节的实验表明 PLSR 模型的鲁棒性好，预测精度高。现将 D3 组的 14 个样本作为测试集，测试所建立的 PLSR 模型的预测精度。将 D1 组和 D2 组的 28 个样本用于建模，利用留一法交互验证调整模型参数，再用最佳模型预测 D3 组的 14 个样本。

为模型设置不同个数的主成分，通过留一法交互验证，计算选用不同的主成分个数的情况下得到的训练集建模精度。如图 2 所示，随着主成分个数 N 的增加，MAPE 呈下降趋势，在主成分个数 $N=7$ 时，训练集的平均相对误差最小，为 2.02%，最优参数下的 PLSR 模型在 3D 上预测结果如图 3 所示，测试集样本的真实浓度和预测集浓度具有很强的相关性，见图 3(a) 测试集平均相对误差为 2.19%，RMSE 为 0.044 $\text{mg} \cdot \text{L}^{-1}$ ，见图 3(b)。测试集的平均绝对相对误差为 2.19%，RMSE 为 0.044 $\text{mg} \cdot \text{L}^{-1}$ 。

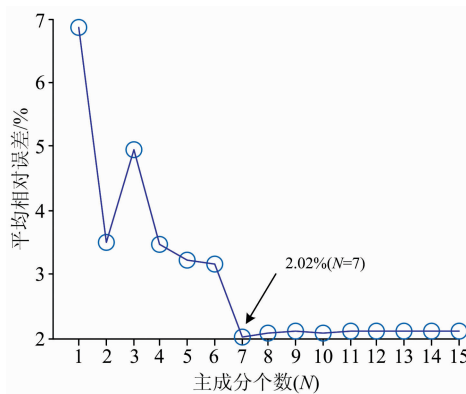


图 2 PLSR 模型选用的主成分个数与建模精度的关系
Fig. 2 Relationship between modeling accuracy and the number of principal components selected in PLSR model

表 1 PLSR 模型对 D3 组 14 个样本的预测结果

Table 1 The prediction result of 14 samples in D3 dataset

样本	实际浓度/ ($\text{mg} \cdot \text{L}^{-1}$)	预测浓度/ ($\text{mg} \cdot \text{L}^{-1}$)	相对误差/ %
1	0.2	0.22	10.00
2	0.4	0.403	0.75
3	0.6	0.63	5.00
4	0.8	0.801	0.13
5	1.0	1.07	7.00
6	2.0	2.04	2.00
7	3.0	3.02	0.67
8	4.0	4.05	1.25
9	5.0	5.02	0.40
10	6.0	6.10	1.67
11	7.0	7.05	0.71
12	8.0	8.03	0.38
13	9.0	9.04	0.44
14	10.0	10.03	0.30

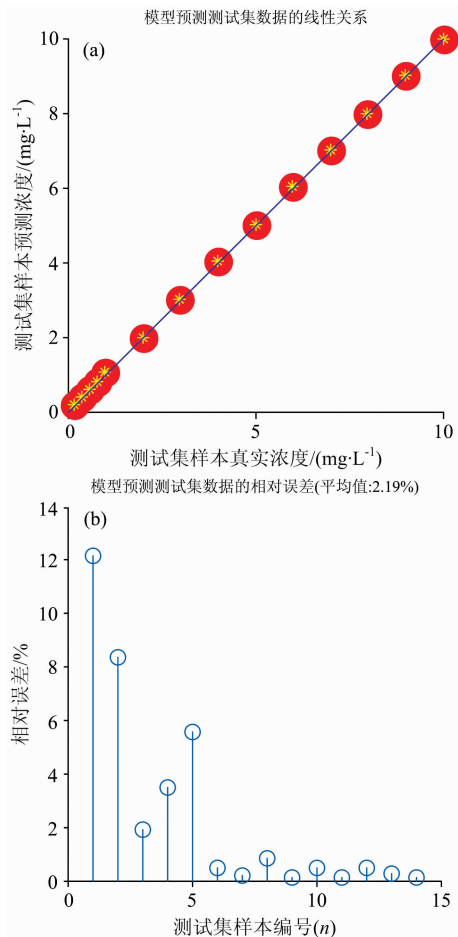


图 3 建立的 PLSR 模型在测试集 D3 上的预测结果分析
(a): 模型预测测试集线性关系; (b): 模型预测测试集数据的相对误差

Fig. 3 The analysis of the prediction result on D3 dataset achieved by the calibrated PLSR model

(a): The linear relationship of test set data;
(b): The relative error of test set data

由表 1 可知，除了测试集在 0.2 和 1.0 $\text{mg} \cdot \text{L}^{-1}$ 的绝对相对误差相对较差，其他浓度范围都有低于 5% 的不错的预测误差，平均绝对相对误差为 2.19%。

3 结 论

将传统的线性建模方法偏最小二乘回归法应用于亚硝酸盐氮溶液的浓度检测，通过对数据进行偏最小二乘回归，从数百维复杂的光谱数据提取出主要信息，并压缩维度，降低整个模型的复杂度并用留一交互验证法选出最优模型。通过同一天交互验证，不同天互相预测验证了 PLSR 非常适合亚硝酸盐氮检测，鲁棒性强，预测精度高。最终确定的模型在 0.2~10.0 $\text{mg} \cdot \text{L}^{-1}$ 浓度范围的平均绝对相对误差为 2.19%，除了在低浓度 0.2 和 1.0 $\text{mg} \cdot \text{L}^{-1}$ 的平均相对误差为 10% 和 7%，其余浓度范围都具有低于 5% 的较好的预测精度，PLSR 可以用于亚硝酸盐氮的快速无损检测。

References

- [1] Wang J, Hu M, Zhang F, et al. *Stochastic Environmental Research and Risk Assessment*, 2018, 32(9): 2633.
- [2] WANG Hong-liang, REN Guo-xing, MA Ran, et al(王洪亮, 任国兴, 马 然, 等). *Environmental Science & Technology(环境科学与技术)*, 2010, (s2): 372.
- [3] ZHANG Da-fen, TAO Hong-xia, ZHU Ao-lan, et al(张大芬, 陶红霞, 朱敖兰, 等). *Chinese Journal of Health Laboratory Technology(中国卫生检验杂志)*, 2015, (5): 639.
- [4] YANG Bing, PENG Chuan-you, PENG Bin, et al(杨 兵, 彭传友, 彭 斌, 等). *Chinese Journal of Spectroscopy Laboratory(光谱实验室)*, 2011, 28(6): 2952.
- [5] National Standard of the People's Republic of China(中华人民共和国国家标准). GB 7493—87. *Water Quality-Determination of Nitrogen(Nitrite) Spectrophotometric Method(水质亚硝酸盐氮的测定分光光度法)*. Beijing: State Environmental Protection Bureau(北京: 国家环境保护局), 1987.
- [6] LU Wan-zhen(陆婉珍). *Modern Near Infrared Spectroscopy Analytical Technology(现代近红外光谱分析技术)*. 2nd Ed. (第 2 版). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2007. 44.
- [7] Chai T, Draxler R R. *Geoscientific Model Development*, 2014, 7(3): 1247.
- [8] Kim S, Kim H. *International Journal of Forecasting*, 2016, 32(3): 669.

Determination of Nitrite Nitrogen Concentration in Surface Water Based on UV-Vis Spectral Analysis Method

LI Qing-bo¹, HE Lin-qian¹, CUI Hou-xin², HAO Long-teng², SUN Dong-sheng²

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 With the rapid development of China's economy, surface water pollution has become increasingly serious. Therefore, it is of vital importance to realize the continuous monitoring of surface water quality to ensure human health and protect the environment. Nitrite nitrogen concentration is an important index in water quality assessment. Polluted water poses a great threat to human, livestock and aquatic products. The detection of organic pollutants by UV-Visible absorption spectrum has become an important method for water quality detection. A few papers about the detection of nitrite nitrogen concentration in water by UV-Vis spectroscopy in China can be found. Most methods require chemical pretreatment of water samples and then use UV spectrophotometer to predict the nitrite nitrogen concentration. These methods are tedious, time-consuming and labor-consuming so that they can't realize real-time continuous detection. Besides, they will cause further environment pollution. The detection of nitrite nitrogen concentration without chemical pretreatment based on UV-Vis absorption spectrometry is rarely mentioned in the literature. Therefore, UV-Vis spectroscopy was adopted in this paper to carry out basic research on unattended automatic continuous monitoring of surface water quality. Nitrite nitrogen solution samples were prepared and a three-day experiment was designed to measure the ultraviolet and visible spectra of all samples (group D1, group D2, and group D3) respectively every day. Firstly, samples from group D1 and group D2 were modeled respectively by partial least squares regression method (PLSR). The mean absolute percentage error (MAPE) obtained by full cross validation was 1.19% and 1.85% respectively. The result shows that the PLSR model has good prediction accuracy. Secondly, in order to verify the adaptability of PLSR model under different measurement conditions, the experimental data of group D1 and group D2 were used for mutual prediction analysis. The MAPE was respectively 3.36% and 4.51%, less than 5%, indicating that PLSR model has good robustness. Finally, all samples from group D1 and group D2 were used for PLSR modeling, and samples from group D3 were used as the test set. The MAPE of the test set was 2.19%. The results show that the MAPE of the PLSR algorithm based on UV-Vis spectral analysis technique for detecting the nitrite nitrogen concentration in solution is controlled under 5%, better than similar reports. In addition, the modeling process of PLSR model is simple and the operation time is short. The model is simple in structure and easier

to be transplanted and solidified into embedded systems, bringing convenience to the later development and design of portable devices. As the basic research of the detection of nitrite nitrogen concentration in surface water, this paper can provide guidance for the accurate and rapid detection of surface water quality in the future.

Keywords UV-Vis; Nitrite nitrogen; Partial least squares regression; Water quality; Online inspection

(Received Jun. 8, 2019; accepted Oct. 12, 2019)