

基于紫外光谱的水体硝酸盐浓度混合预测模型研究

陈颖¹, 何磊¹, 崔行宁¹, 韩帅涛¹, 朱奇光², 翟应俭³, 李少华³

1. 燕山大学电气工程学院河北省测试计量技术及仪器重点实验室, 河北 秦皇岛 066004

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器重点实验室, 河北 秦皇岛 066004

3. 河北先河环保科技股份有限公司, 河北 石家庄 050000

摘要 水体中的硝酸盐浓度过高不仅会造成水环境污染而且会对人类身体健康造成很大威胁, 传统的检测硝酸盐的方法检测时间长且操作复杂。针对水体中硝酸盐氮难以快速在线检测的问题, 基于紫外吸收光谱, 提出了一种混合预测模型结合光谱积分快速定量检测水体中硝酸盐浓度的方法。混合预测模型为低浓度样本建立的双波长法预测模型与高浓度样本建立的偏最小二乘支持向量机 (LS-SVM) 预测模型数据融合之后的模型。按照合适的浓度梯度配备了 19 组硝酸盐氮标准溶液, 通过实验测得不同浓度硝酸盐氮样本的光谱数据。首先基于双波长法对所有样本进行回归分析, 按照 $A = A_{220} - 2A_{275}$ 计算不同实验样本的吸光度 A , 其中 A_{220} 和 A_{275} 是 220 和 275 nm 处样本的吸光度, 将吸光度 A 与样本浓度值进行线性回归, 拟合出样本浓度的预测值。结果显示当样本浓度较小时, 相关性很好, r 为 0.997 4, 随着实验样本浓度的上升, 曲线发生严重的非线性漂移, 因此双波长法只适合低浓度样本预测模型的建立。对于高浓度样本, 光谱重叠严重, 适合建立非线性的预测模型, 支持向量机 (SVM) 与 LS-SVM 都适合小样本的非线性数据建模, LS-SVM 预测精度稍高, 运算速度稍快。通过对所有的实验样本进行全波长光谱积分, 比较相邻样本光谱积分的变化率可以筛选出样本的临界浓度值, $4 \text{ mg} \cdot \text{L}^{-1}$ 的硝酸盐样本积分值前后变化率最大, 因此选择 $4 \text{ mg} \cdot \text{L}^{-1}$ 作为临界浓度值较为合适。浓度高于 $4 \text{ mg} \cdot \text{L}^{-1}$ 的实验样本建立 LS-SVM 预测模型, 通过交叉验证的方法选择出合适的参数, 正则化参数 $\gamma = 50$, 核函数选择高斯核, 核函数宽度 $\sigma^2 = 0.36$, 训练样本之后进行回归; 其余样本建立双波长法预测模型, 最后进行两种模型的数据融合, 形成从低浓度到高浓度的水体中硝酸盐浓度的检测。为了验证混合预测模型的预测精度, 另外建立了 SVM, LS-SVM, 偏最小二乘 (PLS) 等模型, 并求出 r , 预测值与真实浓度值平均绝对误差 (MAE) 和均方根误差 (RMSE) 对模型进行评价。验证结果表明, 相比于 SVM, LS-SVM 和 PLS 等模型, 提出的混合模型回归的相关系数为 0.999 86, 分别提高了 1.8%, 1.6% 和 0.45%, 预测值与真实浓度的平均绝对误差为 2.55%, 分别降低了 6.27%, 4.49% 和 1.01%, 均方根误差为 0.303, 为四种预测模型中最小, SVM 与 LS-SVM 的相对误差相对较高, PLS 模型相对误差上下波动比较大, 混合预测模型相对误差最为稳定, 并保持在较低水平, 由此可见混合预测模型的预测效果明显优于其他几种模型。并与文献 [5—7] 中的测量方法进行对比, 该混合预测方法可以简单快速的测量水体中硝酸盐氮的浓度, 且不需要试剂, 无二次污染, 与文献 [9] 中的预测模型相比, 预测精度明显提高。因此提出的混合模型可正确快速的预测水体中硝酸盐氮的浓度, 可为在线监测水体中硝酸盐浓度提供有效的技术参考。

关键词 紫外光谱; 混合预测; 双波长; LS-SVM; 数据融合; 在线监测

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1489-06

收稿日期: 2018-03-30, 修订日期: 2018-07-21

基金项目: 国家自然科学基金项目(61201112, 61475133), 河北省自然科学基金项目(F2016203188), 河北省普通高校学校青年拔尖人才计划项目(BJ2014056)资助

作者简介: 陈颖, 女, 1980年生, 燕山大学电气工程学院河北省测试计量技术及仪器重点实验室教授 e-mail: chenying@ysu.edu.cn

引 言

近些年来,随着我国工农业的发展,人民生活水平的不断提高,越来越多的含氮废物排放到水中,导致氮污染问题愈加严重,海洋、河流、地下水等都受到了不同程度的污染^[1]。水体中的含氮物质经过氧化转化为硝酸盐,含量过高容易造成水体富营养化,导致藻类大规模生长,形成赤潮等,破坏生态平衡;水体中硝酸盐被人体吸收可以在人体内被还原成亚硝酸盐,过量的亚硝酸盐会使血液运输氧气的的能力下降,当饮用水中硝酸盐含量过高时,会对人体造成很大伤害^[2]。国际饮用水标准规定,饮用水中硝酸盐氮含量的上限是 $10 \text{ mg} \cdot \text{L}^{-1}$ 。为防止氮污染进一步加重,有必要对水中的硝酸盐进行快速检测,并作为水质监测的重要指标之一^[3]。

目前,硝酸盐检测方法主要包括离子色谱法、镉柱还原法、离子电极法等,但大多数操作复杂而且需要很长时间,实验装置昂贵^[4]。Horioka 等利用离子色谱法同时选择性的检测了海水中的硝酸盐,亚硝酸盐及磷酸盐,测量周期在 15 min 左右^[5]。Lorrana 等采用物理的流动注射法结合镉柱还原法,测量硝酸盐及亚硝酸盐的浓度,其硝酸盐的检测下限可以达到 $0.15 \text{ mg} \cdot \text{L}^{-1}$ ^[6]。Manea 等利用电化学的方法,以银掺杂的柔性石墨作为电极,同时检测水中硝酸盐及亚硝酸盐的含量,但是测量精度相对较低^[7]。紫外吸收光谱法可以简单迅速检测水体中硝酸盐含量,不需要试剂并且无二次污染,国内外学者基于光谱吸收法作了很多研究,赵友全等基于紫外光谱,建立主成分分析结合偏最小二乘回归的模型,可以很好地对水样进行分类^[8]。杨鹏程等将紫外光谱与偏最小二乘相结合,对海水中硝酸盐的浓度进行回归计算,

并做了相关的实验进行验证,验证结果良好^[9]。袁月等采用基于紫外光谱的双波长法测定葡萄糖转化为 5-羟甲基糠醛的葡萄糖的转化率和 HMF 的产率,并通过实验验证了该方法的准确性^[10]。王帅等提出了基于 LS-SVM 算法的煤炭指标预测方法,近红外光谱数据与煤炭指标的相关性普遍在 0.9 左右,取得了较好的预测效果^[11]。

本文提出了一种基于紫外吸收光谱的混合回归预测模型,通过光谱积分筛选出样本的临界浓度值,将样本分为高浓度和低浓度两部分,当样本浓度低于临界值时,采用双波长法定量分析,当样本浓度较高时,光谱重叠严重,适合非线性分析,采用 LS-SVM 法进行回归预测,最后数据融合,并与其他几种定量分析模型进行了对比,通过实验对提出的混合模型进行验证,验证结果优于其他几种模型。

1 实验部分

1.1 装置

硝酸盐检测实验装置如图 1 所示,整个实验装置主要包括光源、耦合光纤、样品槽、光谱仪等。其中光源选用的是紫外氙光源(HPX-2000 闻奕光电),光源光谱范围 $190 \sim 2200 \text{ nm}$,并自带有效率耦合 SMA905 光纤接口输出,光源通过光纤耦合到样品槽。样品槽采用高通透的石英比色皿,为了提高入射光、透射光的传输效率,在样品槽两边加上凸透镜组。光谱仪采用光纤光谱仪(MAX2000 Pro 闻奕光电),负责光谱信号的采集,具有高灵敏度,高分辨率,可以通过同步脉冲调节氙灯光源和光谱仪,避免产生光路延迟对测量结果造成误差。测量的光谱数据通过 USB 传输到计算机中进行光谱数据处理,建立合适的数学模型。

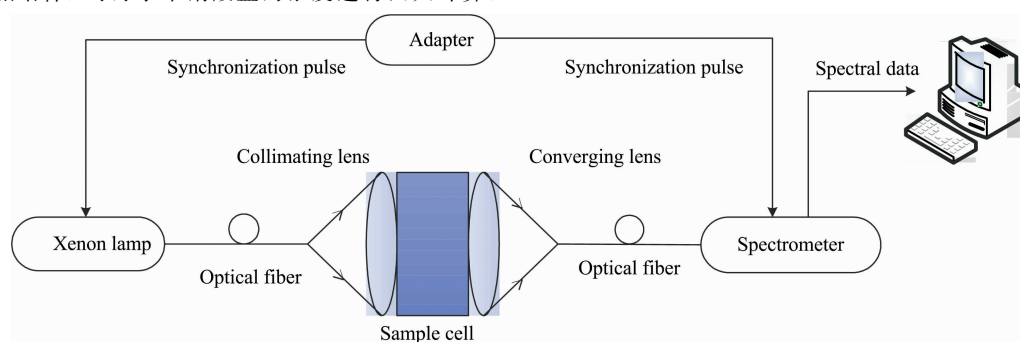


图 1 硝酸盐检测实验装置

Fig. 1 Nitrate detection experiment device

1.2 数学模型

采用紫外吸收光谱法是因为硝酸盐仅在紫外光波段有吸收,在其他波段几乎无吸收,且在特定的浓度范围内才符合朗伯-比尔定律。通过实验装置进行相关实验,测得所需的光谱数据,因变量为预测浓度值,自变量为样本的吸光度数据,建立合理的数学模型,利用双波长、LS-SVM 等算法反演出水体中硝酸盐的浓度作为预测浓度值。通过对所有的样本进行光谱积分,比较每个实验样本积分值前后的变化率,

筛选出临界浓度值,低于临界浓度值的样本采用双波长法进行回归预测,大于等于临界值的样本进行 LS-SVM 回归预测,最后进行数据融合,并求出预测模型评价指标。

LS-SVM 是在 SVM 基础上扩展的,主要是将 SVM 中的不等式约束扩展成一个等式的约束条件^[12]。训练样本通过映射函数 $\phi(x)$ 映射到高维特征空间,并构造最优决策函数,如式(1)所示

$$y = \omega^T \phi(x) + b \quad (1)$$

式(1)中 w 和 b 是未知参数, 根据结构风险最小化原则, 得到 LS-SVM 的约束条件以及目标函数, 如式(2)所示

$$\begin{cases} \min J(w, e) = 0.5w^T w + 0.5\gamma \sum_{k=1}^n e_k^2 \\ y_k = w^T \phi(x_k) + b + e_k \quad k = 1, 2, \dots, n \end{cases} \quad (2)$$

式(2)中 w 为权向量, γ 为正则化参数, e_k 为误差变量, b 为偏置参数, 相应的拉格朗日函数为

$$L(w, b, \alpha, e) = J(w, e) - \sum_{k=1}^n \alpha_k [w^T \phi(x_k) + b + e_k - y_k] \quad (3)$$

式(3)中, α_k 为拉格朗日乘子, 通过对式(3)求偏导得到方程组(4), 通过求解方程组(4), 将其转化成线性方程组(5), 其中 $\vec{1} = (1, 1, \dots, 1)$, $\alpha = (\alpha_1, \dots, \alpha_n)^T$, $Y = (y_1, \dots, y_n)^T$, $I = \text{diag}\{1, 1, \dots, 1\}$, $K = K(x, x_k)$ 是一个 $n \times n$ 的对称矩阵, 利用求解线性方程组的偏最小二乘法求解公式(5), 得到模型的决策函数, 如式(6)所示。LS-SVM 可以用来分类或者回归, 适用于小样本的非线性数据建模^[13], 主要参数包括核函数的选取以及正则化参数 γ , 本文中主要用 LS-SVM 进行回归, 利用光谱数据进行浓度反演。

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w - \sum_{k=1}^n \alpha_k \phi(x_k) = 0 \\ \frac{\partial L}{\partial b} = 0 \rightarrow -\sum_{k=1}^n \alpha_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \gamma e_k - \alpha_k = 0 \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow w^T \phi(x_k) + b + e_k - y_k = 0 \end{cases} \quad (4)$$

$$\begin{bmatrix} 0 & \vec{1} \\ \vec{1} & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (5)$$

$$y(x) = \sum_{k=1}^n \alpha_k K(x, x_k) + b \quad (6)$$

2 结果与讨论

2.1 样本光谱数据采集

一般水体的硝酸盐浓度为 $0.2 \sim 5 \text{ mg} \cdot \text{L}^{-1}$ 左右, 当发生较严重污染时, 硝酸盐氮浓度可以达到 $10 \sim 20 \text{ mg} \cdot \text{L}^{-1}$ 。因此通过 $100 \text{ mg} \cdot \text{L}^{-1}$ 的硝酸盐氮标准液(分析纯)按照不同的浓度梯度配制了 $0.2 \sim 20 \text{ mg} \cdot \text{L}^{-1}$ 的 19 组不同浓度的样本, 为了相对减少光谱重叠, 随着样本浓度的上升, 配制的浓度梯度逐渐增大。测试波段为 $190 \sim 350 \text{ nm}$, 变化步长为 1 nm , 每一组样本测量 3 遍, 吸光度数据取平均值, 对测量的光谱数据进行简单的数据预处理。

$4.0 \text{ mg} \cdot \text{L}^{-1}$ 的硝酸盐氮溶液的吸收光谱图如图 2(a)所示, 水体中硝酸盐的官能团的吸收带约为 $190 \sim 250 \text{ nm}$, 在 250 nm 之后的波长处几乎无吸收, 因此建立数学模型时选取 $190 \sim 250 \text{ nm}$ 为分析波段。19 组不同浓度溶液的吸收光谱图如图 2(b)所示。

2.2 双波长法回归分析

采用双波长法对所有样本进行线性回归拟合, A_{220} 和

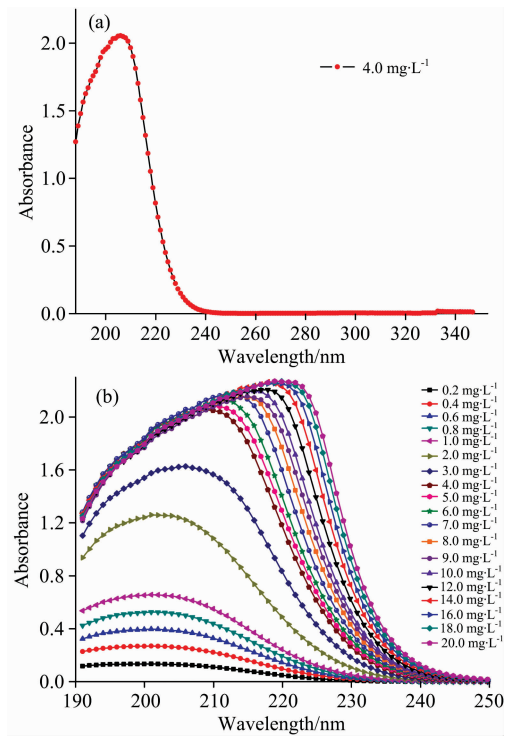


图 2 不同浓度样本的吸收光谱图
Fig. 2 Absorption spectra of different concentrations of samples

A_{275} 是 220 和 275 nm 处样本的吸光度, 按照 $A = A_{220} - 2A_{275}$ 计算不同样本的吸光度 A , 与样本浓度值进行线性回归, 最后拟合出样本浓度的预测值, 此方法也是测量水中总氮的国标方法。双波长法的吸光度 A 与样本浓度值的折线图如图 3 所示。可以发现, 当浓度较低时, 吸光度 A 与样本浓度的相关性较好, 随着样本浓度的升高, 曲线发生很大的非线性漂移, 这是因为吸光物质的浓度越大, 物质分子之间的相互作用越大, 导致偏离朗伯-比尔定律的程度也越严重。

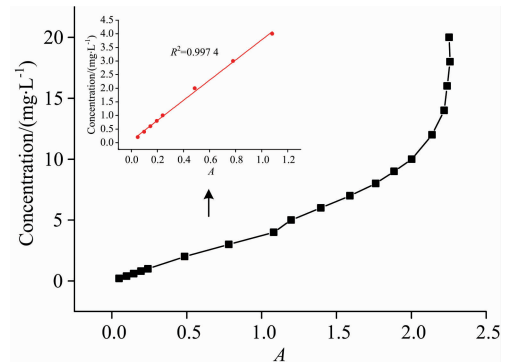


图 3 吸光度 A 与样本浓度值折线图
Fig. 3 Line graph of absorbance A and sample concentration values

图 3 的内插图为样本浓度较低时, 双波长法测量的吸光度 A 与样本浓度的回归曲线, 相关系数 r 为 0.9974 , 线性回归效果较好。由图 2(b)可知, 当样本浓度较高时, 大部分

区间的光谱重叠严重,不适合线性回归,所以有必要选择一个合适的临界浓度值,将样本分成两部分,分别使用不同的数学模型建模进行拟合预测。

2.3 混合预测模型分析

通过光谱积分的方法可以筛选出临界浓度值,所有样本的光谱积分如图 4 所示。图 4(a)为样本的光谱积分与波长的曲线,从下到上的样本浓度逐渐升高,图 4(b)为全波长光谱积分值与样本浓度的关系曲线。

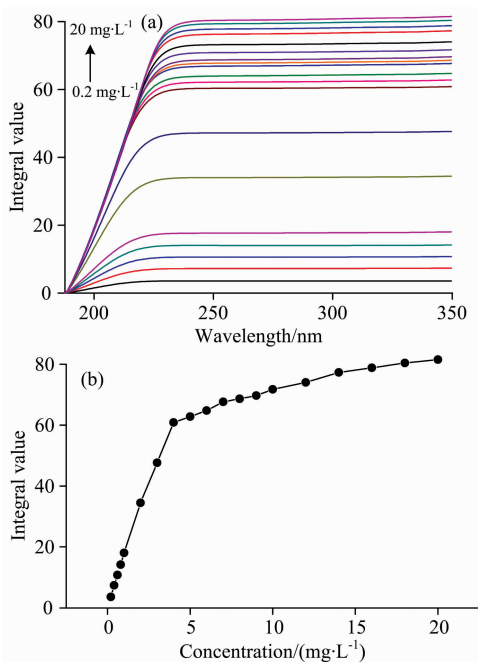


图 4 全波长光谱积分图

Fig. 4 Full wavelength spectral integration

当样本浓度大于 $4 \text{ mg} \cdot \text{L}^{-1}$ 时,光谱积分曲线在分析波段(190~250 nm)很大一部分重叠,并且全光谱积分值的变化率明显减慢,说明浓度大于 $4 \text{ mg} \cdot \text{L}^{-1}$ 的样本光谱数据适合非线性回归分析。因此,选择 $4 \text{ mg} \cdot \text{L}^{-1}$ 为划分样本的临界浓度值较为合适。对于浓度高于 $4 \text{ mg} \cdot \text{L}^{-1}$ 的样本采用适合非线性数据的 LS-SVM 法进行回归预测,通过交叉验证的方法选择出合适的参数,选择正则化参数 $\gamma=50$,核函数选择高斯核,核函数宽度 $\sigma^2=0.36$,训练样本之后进行回归分析,拟合出预测值,最后进行数据融合,形成从低浓度到高浓度的浓度值的预测。通过回归预测数据的 r ,预测值与真实浓度值的 MAE 和 RMSE 对模型进行评价。基于双波长法,LS-SVM 混合预测模型的回归结果如图 5 所示,预测浓度值与真实浓度之间的 r 为 0.999 86。

基于实验测得的光谱数据,另外建立了 PLS, SVM 和 LS-SVM,三种定量分析预测模型,并与本文提出的混合预测模型进行了对比,四种定量分析模型预测浓度值与样本浓度值相对误差如图 6 所示,图中从左至右的样本浓度依次升高。

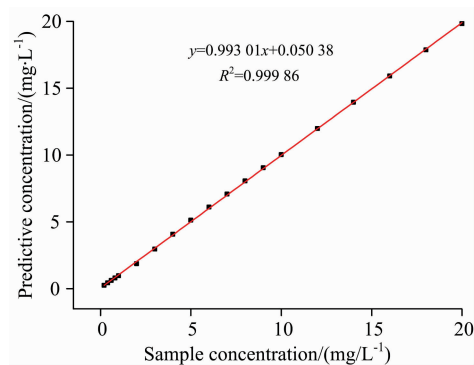


图 5 混合预测模型回归图

Fig. 5 Mixed prediction model regression

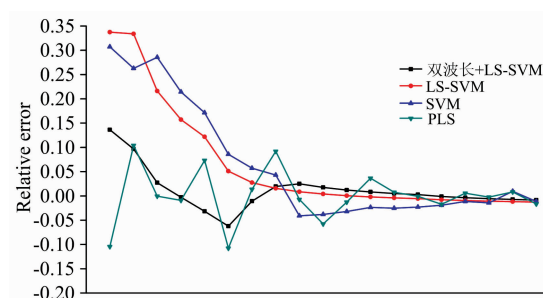


图 6 不同分析模型相对误差图

Fig. 6 Relative analysis error model of different analysis models

由图 6 和表 1 分析可知,当样本浓度较小时,SVM 与 LS-SVM 模型的预测值与真实浓度值的相对误差较高,PLS 模型的相对误差的上下波动较大,混合预测模型的相对误差变化最为平稳,并且保持在较低水平,相比于 SVM,LS-SVM 和 PLS 等模型,提出的混合模型的回归的 r 为 0.999 86,分别提高了 1.8%,1.6%和 0.45%,MAE 为 2.55%,分别降低了 6.27%,4.49%和 1.01%,RMSE 为 0.303,为四种分析模型中最小。比较可知,提出的混合预测模型预测效果明显优于其他几种分析模型。

表 1 不同分析模型评价参数比较

Table 1 Comparison of evaluation parameters of different analytical models

| 预测模型 | r | RMSE | MAE/% |
|------------|---------|-------|-------|
| SVM | 0.981 1 | 0.667 | 8.82 |
| LS-SVM | 0.983 2 | 0.656 | 7.04 |
| PLS | 0.995 3 | 0.352 | 3.56 |
| 双波长+LS-SVM | 0.999 8 | 0.303 | 2.55 |

3 结 论

提出了一种将紫外光谱与混合预测模型相结合的快速定量预测水体中硝酸盐浓度的方法,混合预测模型包括双波长

法与LS-SVM。

(1)采用双波长法对所有样本回归分析,整体回归效果不理想,样本浓度偏高时,回归曲线发生较大偏离,对低浓度样本回归分析,回归效果好,相关系数为0.997 4。

(2)对所有样本光谱积分,筛选出临界浓度值 $4 \text{ mg} \cdot \text{L}^{-1}$,将样本分为高浓度、低浓度两部分,分别建立LS-SVM与双波长回归预测模型,进行数据融合。

(3)相比于SVM,LS-SVM和PLS等模型,混合预测模型回归的 r 为0.999 86,分别提高了1.8%,1.6%和

0.45%;预测值与真实浓度值的相对误差变化最为平稳且保持在较低水平;MAE为2.55%,分别降低了6.27%,4.49%和1.01%,RMSE为0.303,为四种模型中最小,预测效果明显优于SVM,LS-SVM和PLS等模型。与文献[5—7]中的测量方法进行对比,该混合预测方法可以简单快速的测量水体中硝酸盐氮的浓度,且不需要试剂,无二次污染,与文献[9]中的模型相比,预测精度明显提高,此混合预测模型可为在线监测水体中硝酸盐浓度提供有效的技术参考。

References

- [1] ZHANG Cui-yun, ZHANG Sheng, LI Zheng-hong, et al(张翠云, 张 胜, 李政红, 等). *Advances in Earth Science(地球科学进展)*, 2004, 19(2): 183.
- [2] ZHANG Ya-li, ZHANG Yi-zhang, ZHANG Yuan, et al(张亚丽, 张依章, 张 远, 等). *China Environmental Science(中国环境科学)*, 2014, 34(1): 170.
- [3] HUANG Jian-feng, ZHANG Yi-zhang, ZHANG Yuan, et al(黄剑峰, 张依章, 张 远, 等). *Research of Environmental Sciences(环境科学研究)*, 2012, 25(11): 1229.
- [4] SUN Hao, GUO Hui, ZHAO Hui, et al(孙 浩, 郭 慧, 赵 辉, 等). *Chinese Journal of Environmental Engineering(环境工程学报)*, 2016, 10(4): 2122.
- [5] Horioka Y, Kurata S, Ito K. *Bunseki Kagaku*, 2014, 63(8): 657.
- [6] Lorrana N N Nóbrega, Laiz de O Magalhães, Alexandre Fonseca. *Microchemical Journal*, 2013, 110: 553.
- [7] Manea F, Remes A, Radovan C, et al. *Talanta*, 2010, 83(1): 66.
- [8] 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.
- [9] YANG Peng-cheng, DU Jun-lan, CHENG Chang-kuo(杨鹏程, 杜军兰, 程长阔). *Marine Environmental Science(海洋环境科学)*, 2016, 35(6): 943.
- [10] YUAN Yue, YAO Shuang-quan, WANG Shuang-fei(袁 月, 姚双全, 王双飞). *Modern Chemical Industry(现代化工)*, 2016, (2): 176.
- [11] WANG Shuai, ZHANG Yun-jia, WANG Xue-mei, et al(王 帅, 张云佳, 王雪梅, 等). *Computer Systems and Applications(计算机系统应用)*, 2017, 26(5): 232.
- [12] ZOU Yong-jie, DUANMU Jing-shun, GAO Hai-long(邹永杰, 端木京顺, 高海龙). *Computer Engineering and Design(计算机工程与设计)*, 2010, 31(10): 2290.
- [13] Zhang C, Zhang H. *International Journal of Computer Integrated Manufacturing*, 2016, 29(1): 76.

Study on Mixed Prediction Model of Nitrate Concentration in Water Based on Ultraviolet Spectroscopy

CHEN Ying¹, HE Lei¹, CUI Xing-ning¹, HAN Shuai-tao¹, ZHU Qi-guang², ZHAI Ying-jian³, LI Shao-hua³

1. Hebei Province Key Laboratory of Test/Measurement Technology and Instrument, School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China
2. Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, Yanshan University, Qinhuangdao 066004, China
3. Hebei Sailhero Environmental Protection Hi-tech Co., Ltd., Shijiazhuang 050000, China

Abstract High concentration of nitrates in water will not only cause water environment pollution but also pose a great threat to human health. The traditional methods for detecting nitrates have a long detection time and are complex to operate. In view of the difficulty in rapid on-line detection of nitrate nitrogen in water, a method combined a mixed prediction model with spectral integration was proposed to rapidly detect nitrate concentration in water based on ultraviolet absorption spectroscopy. The mixed prediction model is a model after data fusion of the dual wavelength prediction model established by low-concentration samples and the partial least-squares support vector machine (LS-SVM) prediction model based on the high concentration samples. Ac-

ording to the appropriate concentration gradient, 19 sets of nitrate nitrogen standard solution were equipped, and the spectral data of nitrate nitrogen samples of different concentrations were measured by experiment. First, Regression analysis was performed on all samples based on the dual wavelength method. Absorbance A was calculated for different experimental samples according to $A = A_{220} - 2A_{275}$, where A_{220} and A_{275} were the absorbance of the samples at 220 and 275 nm. The values were linearly regressed to fit the predicted values of the sample concentrations. The results showed that when the sample concentration is small, the correlation is very good, and r is 0.997 4. The two-wavelength method is only suitable for the establishment of low-concentration samples prediction model with a serious nonlinear drift in the rising curve of the experimental samples concentration. For high-concentration samples, spectral overlap is severe and it is suitable for establishing nonlinear prediction models. Both support vector machine (SVM) and partial LS-SVM are suitable for nonlinear data modeling of small samples. The LS-SVM has a slightly higher prediction accuracy and a slightly faster operating speed. By performing full-wavelength spectral integration on all experimental samples and comparing the rate of change of the spectral integrals of adjacent samples, the critical concentration of the sample can be selected. The $4 \text{ mg} \cdot \text{L}^{-1}$ nitrate sample has the largest change rate before and after the integrated value, so it is appropriate to select $4 \text{ mg} \cdot \text{L}^{-1}$ as the the critical concentration value. The LS-SVM prediction model was established for experimental samples with concentrations higher than $4 \text{ mg} \cdot \text{L}^{-1}$. Cross-validation methods were used to select the appropriate parameters. The regularization parameter was $\gamma=50$, and the Gaussian kernel function width was $\sigma^2=0.36$. The other samples were used to establish the dual-wavelength prediction model, and finally performed the data fusion of the two models, which formed the detection of nitrate from low concentration to high concentration. In order to verify the prediction accuracy of the mixed prediction model, the model of SVM, LS-SVM and PLS was established, and evaluated the model with mean absolute error (MAE), correlation coefficient (r), and root mean squared error (RMSE). The verification results showed that compared with other models, the correlation coefficient of the proposed mixed model regression is 0.999 86, which is increased by 1.8%, 1.6%, and 0.45% respectively, and the average absolute error between the predicted value and the true concentration is 2.55%, which decreased by 6.27%, 4.49%, and 1.01% respectively, and the root-mean-square error is 0.303, which is the smallest of the four prediction models. The relative error of SVM and LS-SVM is relatively high, and PLS model fluctuates up and down relatively. The relative error of mixed forecasting model is the most stable and remains at a low level, and the forecasting effect of mixed forecasting model is obviously better than that of other models. Compared with the measurement method in [5-7], this hybrid prediction method can simply and quickly measure the nitrate nitrogen concentration in water without reagents and no secondary pollution, and the prediction accuracy is significantly improved compared with the model in [9]. Therefore, the proposed mixed model can correctly and quickly predict the concentration of nitrate in water, and provide an effective technical reference for on-line monitoring of nitrate concentration in water.

Keywords Ultraviolet spectrum; Mixed prediction; Dual wavelength; LS-SVM; Data fusion; Online monitoring

(Received Mar. 30, 2018; accepted Jul. 21, 2018)