

DOI:10.19853/j.zgjsps.1000-4602.2022.21.016

# 基于紫外-可见光谱的COD预测模型优化方案研究

杨曼孜, 王晓东, 刘长青, 塔哈·马哈巴  
(青岛理工大学 环境与市政工程学院, 山东 青岛 266000)

**摘要:** 水质在线监测的精确性、实时性和可靠性是制约污水处理厂提高自动化水平和实现智能化管理的瓶颈问题。将光谱仪置于青岛某污水处理厂获取污水的紫外-可见(UV-Vis)全光谱(200~737.5 nm)数据,同时在实验室化验分析对应样品的COD浓度,基于主成分回归法(PCR法)和偏最小二乘回归法(PLSR法)构建COD预测模型,并采用两次回归分析的思想选取特征波长重新建模,深入讨论采用不同方法、不同类型波长建模对提高模型预测能力的影响。结果表明:与PCR法相比,采用PLSR法构建的模型具有更优的预测能力;基于PLSR法,选取COD特征吸收波长以及硝酸盐和浊度特征吸收波段的波长用于校准,构建的COD模型获得了最优的预测能力,模型拟合优度( $R^2$ )为0.992,预测均方根误差(RMESP)为1.979 mg/L。因此,采用COD特征吸收波长与非预测目标特征吸收波长组合建模的方式能够进一步提高模型的预测精度。

**关键词:** 紫外-可见光谱; 全光谱分析; 化学需氧量; 主成分回归; 偏最小二乘回归; 模型优化

**中图分类号:** TU992    **文献标识码:** A    **文章编号:** 1000-4602(2022)21-0113-07

## Optimization of COD Prediction Model Based on Ultraviolet-visible Spectrum

YANG Man-zi, WANG Xiao-dong, LIU Chang-qing, TAHA Marhaba  
(School of Environmental and Municipal Engineering, Qingdao University of Technology,  
Qingdao 266000, China)

**Abstract:** The accuracy, real-time and reliability of on-line water quality monitoring are the bottlenecks that restrict the improvement of automation level and the realization of intelligent management of wastewater treatment plants. The spectrometer was placed in a wastewater treatment plant in Qingdao to obtain the UV-Vis full-spectrum (200–737.5 nm) data of sewage. Meanwhile, the COD of corresponding samples was analyzed in the laboratory. COD prediction models were constructed based on principal component regression (PCR) method and partial least squares regression (PLSR) method, characteristic wavelengths were selected to reconstruct the model by using the idea of twice regression analysis, and the influence of different methods and different types of wavelength on improving the prediction ability of the model was discussed in detail. Compared with PCR method, the model constructed by PLSR method had better predictive ability. Based on PLSR method, the characteristic absorption wavelengths of COD and the characteristic adsorption wavelengths of nitrate and turbidity were selected for calibration, and the COD model obtained the optimal prediction ability. The goodness of fit ( $R^2$ ) of the model was 0.992, and the root mean square error of prediction (RMESP) was 1.979 mg/L. Therefore, the combined modeling

基金项目: 国家自然科学基金资助项目(51908303)

通信作者: 王晓东    E-mail: wangxiaodong@qut.edu.cn

method of COD characteristic absorption wavelength and non-predicted target characteristic wavelength could further improve the prediction accuracy of the model.

**Key words:** ultraviolet-visible spectrum; full-spectrum analysis; chemical oxygen demand; principal component regression; partial least squares regression; model optimization

伴随着近年来自动控制理论和技术装备的进步,工业自动化有了长足的发展。然而,目前国内广泛使用的水质在线检测系统仍按照自动采样、分析测试和数据分析与传输的流程运行<sup>[1]</sup>,污水水质的在线监测受制于仪表成本高、可靠性差、数据获取时间严重滞后等问题,限制了污水处理厂的自动化和智能化发展。化学需氧量(COD)作为污水处理中重要的水质参数,获取其实时浓度以及变化趋势信息有助于改进污水处理厂的运营管理模式,从而减轻污水对水环境的影响<sup>[2]</sup>。实现COD的低成本、高精确性、快速检测将有助于提高污水处理厂的自动化水平,进而提高污水处理运营效率。

使用光学替代参数表征水质的研究在水质监测领域已经得到了一定的发展<sup>[3-4]</sup>。光谱分析法是常用的水质分析方法,通过样品对入射光的特征吸收分析研究出样品中组分及其含量的信息<sup>[5-6]</sup>。采用紫外-可见光(UV-Vis)在200~750 nm全波段扫描水样得到的全波段吸收光谱图中含有大量复杂的水质信息,将全光谱结合相应的数据驱动模型用以提取有效信息,可构建实时检测模型对水质指标进行预测。

在采用UV光测定河流、城市及工业废水中有机污染物浓度的研究中,从简单地基于单波长、双波长建模,已发展到采用多波长、宽波段进一步校准建模<sup>[7]</sup>。随着水环境的不断恶化,水质组分逐渐趋于复杂,对模型预测精度的要求也逐渐提高。主成分回归法(PCR法)和偏最小二乘回归法(PLSR法)是两种广泛应用的光谱建模法,笔者采用这两种方法构建了基于UV-Vis全波段光谱(200~737.5 nm)的COD预测模型,并采用两次回归分析的思想,进一步选择水样在全波段上的特征吸收波长建模,分析比较各模型的预测能力,研究采用不同波长建模对模型预测精确度的提升效果,探寻以全波段紫外-可见光谱为基础,寻求目标水样的最优建模波长的可能性。

## 1 材料与方法

### 1.1 实验样品、仪器和测定方法

水样取自青岛某污水处理厂二沉池出水和三级处理总出水处,共计90个。在实验室采用快速密闭催化消解法对样品的COD进行定量分析,水样的硝酸盐( $\text{NO}_3^-$ -N)和磷酸盐( $\text{PO}_4^{3-}$ -P)含量较低,分别在3 mg/L和0.03 mg/L左右,pH基本为中性,水温为15~21℃;同时测得总悬浮固体(TSS)最高为6 mg/L,远低于设备商规定的可忽略TSS影响的浓度值(15 mg/L)<sup>[8]</sup>,因此,本实验不进行光谱预处理而直接建模。随机选取90个样品中的75个样品数据作为模型训练集,其余15个用于检测模型精度。水样的COD统计值如表1所示,检测集数据覆盖98.18%的COD浓度区间,有助于准确判断模型的预测能力。

表1 90个水样的COD统计值

Tab.1 COD statistics of 90 water samples

数据集	数量/个	均值/ ( $\text{mg}\cdot\text{L}^{-1}$ )	最小值/ ( $\text{mg}\cdot\text{L}^{-1}$ )	最大值/ ( $\text{mg}\cdot\text{L}^{-1}$ )	标准差SD/ ( $\text{mg}\cdot\text{L}^{-1}$ )
训练集	75	30.07	6	60	20.88
检测集	15	32.80	7	61	18.21
All	90	30.57	6	61	20.35

光谱实验仪器采用奥地利是能(s::can)公司的spectro::lyser™全波段光谱传感器,以浸入式(直接进入液体介质)扫描水样,结合配套的控制器和分析软件可在线获取实时全波段(200~750 nm)的UV-Vis吸收光谱图。每个样品使用两次测量计算的平均光谱作为测量光谱,整个测量期需要15 s,每个光谱的波长间隔为2.5 nm,共计216个吸光度。

### 1.2 特征波长的选择

对于已建立好的多元回归模型,通过特定方法筛选出最优特征波长组合再次建模,可以得到预测能力更强、稳定性更好的模型。在多光谱建模中,常见的波长选择方法主要通过相关性分析、显著性检验、消除无信息变量等思想来剔除不相关或非线性变量,从而达到简化模型的目的<sup>[9]</sup>。

统计显著被广泛地应用于评价科学研究数据

及结果的可靠性,能够从统计意义上反映估计结果是否由抽样误差引起<sup>[10]</sup>。然而,统计检验得出的显著结论并不意味着该信息具有实质意义。由于采用PLSR法构建的回归模型无法进行回归方程系数的估计和假设检验<sup>[11]</sup>,因此笔者提出采用经济显著的概念,以回归方程为基础,从构建方程的贡献率角度进行特征波长的筛选。McCloskey等发表了一系列文章讨论机械地使用统计显著带来的问题以及经济显著的重要实质意义<sup>[12-15]</sup>。经济显著并没有明确的定义,其意义为一种因素变化带给另一因素的明显作用,两者之间存在一定的因果关系,且需要根据研究问题具体分析<sup>[10]</sup>;不同学科对其有不同的称呼,包括:临床显著、科学显著、实际显著、政策显著等<sup>[15]</sup>。在回归分析中,经济显著性主要通过估计系数 $k_i$ 的大小来判断:

$$\gamma_i = \frac{\sigma_i \cdot k_i}{\bar{Y}} \quad (1)$$

式中: $\gamma_i$ 为第 $i$ 个波长的经济显著性, $i=1, \dots, 216$ ;  $\sigma_i$ 为训练组样品第 $i$ 个波长吸光度的标准差; $k_i$ 为多元回归模型中第 $i$ 个波长的系数; $\bar{Y}$ 为训练组样品COD的平均值。

回归系数经济显著性表示自变量一个标准差变化引起的因变量变化值相对于因变量均值的变化情况,回归系数经济显著性越大,表示该自变量对因变量的影响越显著。

### 1.3 模型构建方法

主成分分析法(PCA法)通过对多变量数据矩阵的分析和计算,重组出多个相互独立的新变量,使用其中的少数几个变量即可表征原数据矩阵的所有特征,从而实现降维。这几个新组成的变量称为主成分,这些主成分由原始变量转换而成<sup>[16]</sup>,根据其获取的方差递减的顺序依次排列为第一主成分、第二主成分,以此类推<sup>[17]</sup>。PCR法选取前几个满足目标要求的主成分作为自变量进行多元回归建模,由于PCA法的主成分提取过程与因变量无关,故保证了用于回归分析的主成分与因变量的相互独立性。

PLSR法最早由瑞典化学家Wold提出,已在多种学科中广泛应用<sup>[18-19]</sup>,主要优点有可解决多重相关性、需求样本量小且能用于预测分析<sup>[20]</sup>。PLSR法融合了多元线性回归、典型相关分析和PCA法<sup>[11]</sup>,其思路为:首先分别从自变量和因变量中提

取相互独立的主成分 $t_1$ 和 $u_1$ ,同时要求 $t_1$ 对 $u_1$ 有最大的解释能力,然后建立这些主成分与因变量的多元回归方程<sup>[19]</sup>。PLSR法相较于PCR法的优势在于,其采用循环式的信息分解与提取方法,在提取成分的同时能较好地概括因变量的信息<sup>[16]</sup>,并消除光谱噪声的干扰。

本研究的主要内容是探究基于不同建模方法进行两次回归分析的模型优化过程:第一次进行全波长回归建模,得到全波长的回归方程系数,然后利用这些系数筛选特征吸收波长,采用特征吸收波长再次进行回归建模即可得到更优的回归结果<sup>[21]</sup>。PCR法和PLSR法的主要区别在于建模时是否考虑自变量与因变量的内在联系,在本研究中,采用PLSR法经过两次建模将进一步放大COD与特征吸收波长的相关性,因此可获得更优的预测能力。本研究采用R软件进行数据分析。

### 1.4 模型参数选择

根据模型各成分贡献率和内部留一法交叉验证均方根误差的结果来确定建模的主成分个数。成分贡献率是指主成分的方差占数据变化总方差的百分比,在成分贡献率累计达到95%以上、交叉验证均方根误差最小时的成分个数为最优成分数条件下,模型的拟合效果最好。

### 1.5 模型评估指标

模型预测能力的评估根据预测拟合优度( $R^2$ )和预测均方根误差(RMSEP)确定。 $R^2$ 越大, RMSEP越小,模型的拟合效果越好,预测精度越高。

## 2 结果与讨论

图1为90个样品的UV-Vis光谱图,图中每条曲线代表1个样品的吸收光谱,光谱获取有效信号的波段为200~737.5 nm,水样的吸光度从200 nm到737.5 nm整体呈逐渐下降的趋势。水样的光谱吸收主要来自 $\text{NO}_3^-$ -N、COD以及浊度与色度,特征吸收波段分别为200~245、200~380、380~737.5 nm。光谱吸收主要集中在200~450 nm波段,其中在200~250 nm波段有大量吸收,波长为215和375 nm处有两个较大的特征吸收峰,波长为250、262.5和422.5 nm处有3个小型的吸收峰。水中悬浮物引起的散射干扰会在整个波段上影响水样光谱信号的检测精度,因此浊度对采用光谱法检测污水中的溶解性物质至关重要<sup>[22]</sup>;同时 $\text{NO}_3^-$ -N与COD的



光谱吸收也存在相互影响的现象。Rieger等<sup>[8]</sup>基于与本实验同型号的光谱仪,构建了采用5个波长(250~340 nm)建模的COD模型,模型的拟合优度达到了0.905。本研究水样的光谱在 $\text{NO}_3^-$ -N和浊度的特征波段上都有一定的吸收,相较于只选择COD特征吸收波长进行建模,首先基于200~737.5 nm全波长光谱进行初次建模,然后根据该模型的回归系数经济显著性选择特征吸收波长和校准波长进行二次建模,探究进一步提高模型预测精度的建模方案。

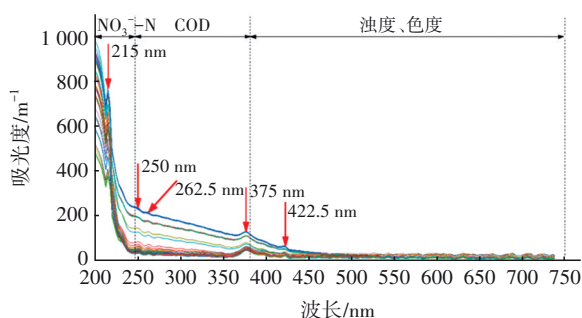


图1 90个水样的紫外-可见光谱图

Fig.1 UV-Vis spectra of 90 water samples

## 2.1 初次建模结果与特征吸收波长讨论

采用PCR法和PLSR法,结合水样的全光谱数据和实验室测定COD含量的化学参考值,基于PCR法和PLSR法进行初次建模,并深入讨论建模过程。采用200~737.5 nm波段建模的COD模型的内部留一法交叉验证均方根误差(RMSE)如图2(a)所示,可知COD模型的RMSE随着主成分数的增加呈下降趋势,PCR模型和PLSR模型的RMSE在主成分数分别到达3个和4个后下降趋势明显减缓,RMSE趋于稳定。各主成分的累计贡献率如图3(a)所示,可知COD的PCR模型第一主成分、第二主成分、第三主成分的方差分别占全部方差的81.72%、11.72%和6.18%,前3个主成分的累计贡献率达到99.62%,超过了95%,足以代替原始变量的信息;同时,主成分数为3个时对应的PCR法的RMSE较小,确定COD的PCR模型主成分数为3个。同理,确定COD的PLSR模型主成分数为4个。

构建基于PCR法与PLSR法的多元回归方程,研究各波长的经济显著性,其值可由波长对应系数来反映。初次建模各波长的经济显著性如图4所示,可知两种方法得到的波长经济显著性总体分布大致相同。正值经济显著性主要对应217.5~405

nm波段,PCR法和PLSR法所得经济显著性的正向最大值分别为250 nm处的2.36%和2.37%,该波长对应吸收光谱图中的一个特征吸收峰,说明该波长为本研究水样中还原性物质的特征吸收波长之一;负向经济显著性的最大绝对值分别为200 nm处的-1.011%和202.5 nm处的-0.601%,而在375~737.5 nm波段内经济显著性的绝对值均低于0.21%(约为最大经济显著性的10%)。

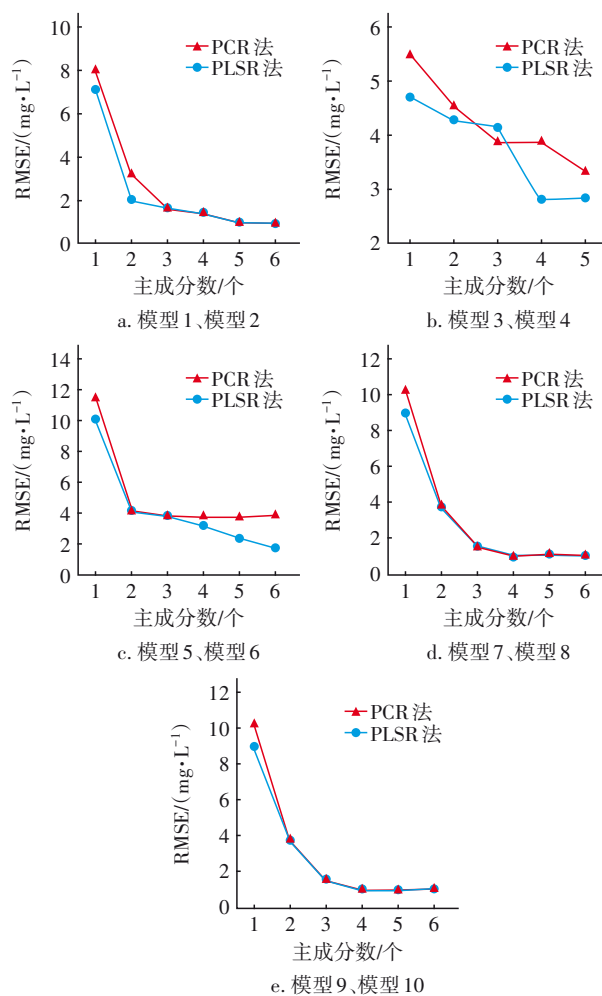
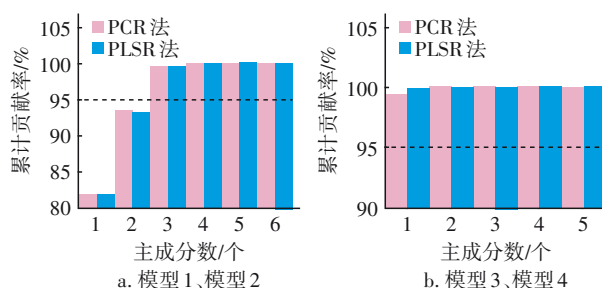


图2 COD模型的RMSE

Fig.2 RMSE of COD models



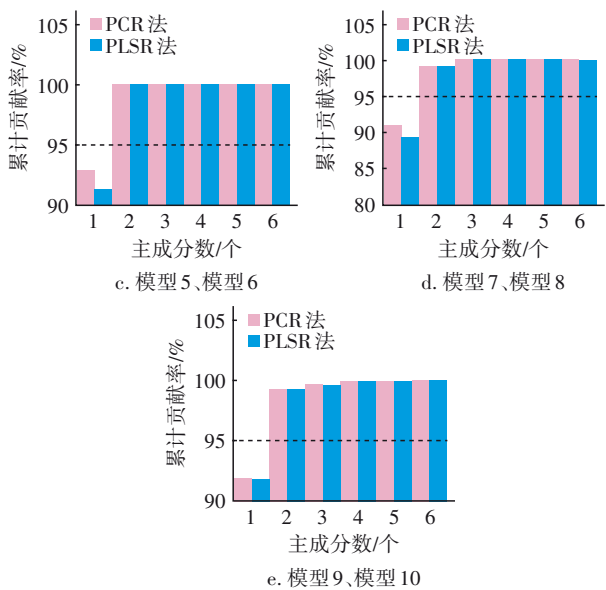


图3 COD模型的主成分累计贡献率  
Fig.3 Cumulative contribution rates of principal components of COD models

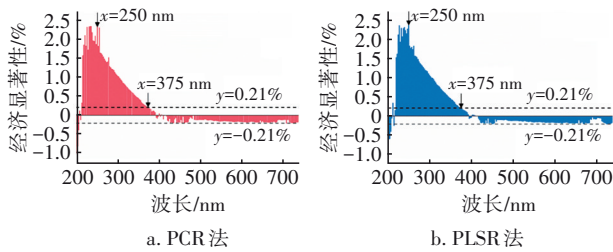


图4 COD模型各波长的经济显著性  
Fig.4 Economic significance of each wavelength of COD models

以优先采用经济显著性绝对值最大的波长为原则,分别在COD特征吸收波段选取正值经济显著性对应波长,在 $\text{NO}_3^-$ -N和浊度特征吸收波段选取负值经济显著性对应波长建模。相较于只选择COD特定波长建模,这种方式的优势是能够结合 $\text{NO}_3^-$ -N和悬浮物的特征吸收波段进一步校正模型的准确性。初次建模分别采用PCR法和PLSR法,得到的回归方程不同,对应各波长的经济显著性存在差异,因此二次建模入选波长也存在差异。建模入选波长和对应模型名称如表2所示,分为4类:220~255 nm波段5个波长;200~215 nm波段5个波长和220~255 nm波段5个波长;200~205 nm波段3个波长、220~255 nm波段5个波长和422.5~727.5 nm波段2个波长;200~372.5 nm全波段共70个波长。其中,220~255 nm波段为COD特征吸收波段,入选波长经济显著性为正值;200~205 nm和422.5~

727.5 nm波段分别为硝酸盐以及浊度与色度的特征吸收波段,入选波长经济显著性为负值。

表2 建模参数及预测模型检验  
Tab.2 Modeling parameters and prediction model testing

入选波长/nm	建模			检验		
	模型名称	建模方法	主成分数/个	RMSE/( $\text{mg}\cdot\text{L}^{-1}$ )	预测 $R^2$	RMSEP/( $\text{mg}\cdot\text{L}^{-1}$ )
200~737.5	模型1	PCR法	3	1.649	0.985	3.298
	模型2	PLSR法	4	1.621	0.986	3.182
225、230、232.5、250、255	模型3	PCR法	3	3.871	0.964	5.750
227.5、232.5、237.5、240、250	模型4	PLSR法	4	2.825	0.987	2.185
200、202.5、205、207.5、210、225、230、232.5、250、255	模型5	PCR法	2	4.159	0.934	6.049
200、202.5、205、207.5、215、227.5、232.5、237.5、240、250	模型6	PLSR法	6	1.756	0.960	2.989
200、202.5、205、225、230、232.5、250、255、707.5、725	模型7	PCR法	4	1.446	0.991	2.084
200、202.5、205、227.5、232.5、237.5、240、250、725、727.5	模型8	PLSR法	4	1.566	0.992	1.979
200~372.5	模型9	PCR法	4	1.158	0.974	4.460
	模型10	PLSR法	3	1.409	0.949	3.413

2.2 二次建模结果与讨论

基于4类波长构建的COD模型的RMSE见图2(b)~图2(e),各主成分累计贡献率见图3(b)~图3(e),可知各模型在主成分数达到2个后,累计贡献率均超过95%,根据上述原则选择最佳主成分数的模型以及使用15个样本数据检验各模型的预测结果见表2。从 $R^2$ 和RMSEP来看,PLSR法明显优于PCR法;并且二次建模后,采用PLSR法的模型4和模型8的预测能力较模型2大幅度提高,RMSEP分别降低了31.3%和37.8%;相较于模型4和模型8,

模型6的预测准确性提升幅度较小(RMSEP仅降低6.1%),说明当采用COD特征吸收波长并仅结合 $\text{NO}_3^-$ -N特征吸收波长建模时, $\text{NO}_3^-$ -N的特征吸收波长对模型进行校准的同时也带来了一定干扰。5组数据对应的10个COD模型中,采用200~737.5 nm和200~372.5 nm波段建模的模型1和模型9的RMSE较小,说明模型拟合效果好,但对应RMSEP较大,超过RMSE的2倍,模型对检测集水样的预测效果不佳。造成这种差异可能是由于全波段和宽波段数据包含大量不相关变量,样本数据有噪声,出现模型过拟合现象。分析采用PLSR法构建的模型,相较于模型4和模型6,模型8采用 $\text{NO}_3^-$ -N和浊度特征吸收波段经济显著性为负值的波长并结合COD特征吸收波长建模,进一步提高了模型的预测精度,取得了最优的预测能力,预测 $R^2$ 为0.992, RMSEP为1.979 mg/L。由此可见,当采用经济显著性为负值的波长并结合COD特征吸收波长建模时,经济显著性为负值的波长对模型可以起到校准作用,减少非预测目标水质参数的光谱吸收对模型的噪声干扰,从而进一步提高模型的预测精确度。

在构建基于光谱法预测水质COD含量模型的研究中,卢福容<sup>[22]</sup>采用200~400 nm波段基于PCR法建模,预测实验室模拟废水的COD, $R^2$ 为0.999 7, RMSEP为0.786 7 mg/L,模型的预测能力较强; Hofstaedter等<sup>[23]</sup>基于与本研究同型号的光谱仪,以COD为20~540 mg/L的SBR工艺过滤水样为研究对象,采用PLSR法,构建了200~737.5 nm全波段的COD模型,模型 $R^2$ 为0.900,拟合效果欠佳,这是由于与实验室模拟废水相比,该实际水样的组分更加复杂,对模型精度的干扰更大;Chen等<sup>[24]</sup>进一步采用基于斜率衍生光谱的PLSR模型,其优点是具有摒弃无关特征变量的能力以及变量间具有更高的协同作用,从而获得了出色的拟合优度( $R^2=0.936$ )。本研究基于全光谱波段选取特征吸收波长构建了COD模型,虽然由于污水中的还原性物质包含多种发色基团,导致其表现在光谱图上的特征吸收不同,但是不同来源的污水主要组分相差不大,因此在实际应用中,可采用本研究中提取特征吸收波长进行建模的方法,对污水处理厂的实际水样建模,从而获得理想的预测模型。

### 3 结论

基于全波段光谱,采用PCR法和PLSR法构建

了污水COD浓度检测模型,并采用两次回归分析的思想,探究从全波段中进一步选取特征波长建模以及非预测目标特征波长进行校准对模型预测能力的提升,详细讨论了基于2种建模方法采用5组建模波长的建模过程以及模型预测效果。与PCR法相比,采用PLSR法构建的模型具有更优的预测能力,并且二次建模后PLSR模型的预测能力得到大幅度提升。基于PLSR法,选取COD特征吸收波长,以及 $\text{NO}_3^-$ -N和浊度特征吸收波段的波长用于校准,构建的COD模型获得了最优的预测能力,模型 $R^2$ 为0.992, RMSEP为1.979 mg/L。这表明采用 $\text{NO}_3^-$ -N和浊度特征波段经济显著性为负值的波长结合COD特征吸收波长建模,对模型可以起到校准作用,减少了非预测目标水质参数的光谱吸收对模型的影响,进一步提高了模型的预测精确度。

### 参考文献:

- [1] 孙海林,李巨峰,朱媛媛.我国水质在线监测系统的发展与展望[J].中国环保产业,2009(3):12-16.  
SUN Hailin, LI Jufeng, ZHU Yuanyuan. Development and prospect of water quality on-line monitoring system in China [J]. China Environmental Protection Industry, 2009(3): 12-16 (in Chinese).
- [2] KORSHIN G V, SGROI M, RATNAWEERA H. Spectroscopic surrogates for real time monitoring of water quality in wastewater treatment and water reuse [J]. Current Opinion in Environmental Science & Health, 2018, 2: 12-19.
- [3] STOREY M V, VAN DER GAAG B, BURNS B P. Advances in on-line drinking water quality monitoring and early warning systems [J]. Water Research, 2011, 45(2):741-747.
- [4] VAILLANT S, POUET M F, THOMAS O. Basic handling of UV spectra for urban water quality monitoring [J]. Urban Water, 2002, 4(3): 273-281.
- [5] MAROSE S, LINDEMANN C, SCHEPER T. Two-dimensional fluorescence spectroscopy: a new tool for on-line bioprocess monitoring [J]. Biotechnology Progress, 1998, 14(1): 63-74.
- [6] 李树珉.烟气排放实时连续监测系统关键技术的研究[D].天津:天津大学,2009:25-26.  
LI Shumin. Study on the Key Technologies of Continuous Emissions Monitoring System [D]. Tianjin: Tianjin University, 2009: 25-26 (in Chinese).

- [7] KHORASSANI HE, TREBUCHON P, BITAR H, *et al.* A simple UV spectrophotometric procedure for the survey of industrial sewage system [J]. *Water Science and Technology*, 1999, 39(10/11): 77-82.
- [8] RIEGER L, LANGERGRABER G, THOMANN M, *et al.* Spectral in-situ analysis of  $\text{NO}_2$ ,  $\text{NO}_3$ , COD, DOC and TSS in the effluent of a WWTP [J]. *Water Science and Technology*, 2004, 50(11): 143-152.
- [9] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, 16(4): 528-542.
- CHU Xiaoli, YUAN Hongfu, LU Wanzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique [J]. *Progress in Chemistry*, 2004, 16(4): 528-542 (in Chinese).
- [10] 吴小康. 关于统计显著与经济显著的若干讨论[J]. *经济学动态*, 2019(1): 145-158.
- WU Xiaokang. Some discussions on statistical significance and economic significance [J]. *Economic Perspectives*, 2019(1): 145-158 (in Chinese).
- [11] 陆洪涛. 偏最小二乘回归数学模型及其算法研究[D]. 北京: 华北电力大学, 2014.
- LU Hongtao. Partial Least Squares Regression Models and Algorithm Research [D]. Beijing: North China Electric Power University, 2014 (in Chinese).
- [12] MCCLOSKEY D N. The loss function has been mislaid: the rhetoric of significance tests [J]. *The American Economic Review*, 1985, 75(2): 201-205.
- [13] MCCLOSKEY D N. *The Secret Sins of Economics* [M]. Chicago: Prickly Paradigm Press, 2002.
- [14] MCCLOSKEY D N, ZILIAK S T. The standard error of regressions [J]. *Journal of Economic Literature*, 1996, 34(1): 97-114.
- [15] MCCLOSKEY D N, ZILIAK S T. Signifying nothing: reply to Hoover and Siegler [J]. *Journal of Economic Methodology*, 2008, 15(1): 39-55.
- [16] 肖雪梦, 张应应. 三种回归方法在消除多重共线性及预测结果的比较[J]. *统计与决策*, 2015(24): 75-78.
- XIAO Xuemeng, ZHANG Yingying. Comparison of three regression methods on eliminating multicollinearity and the prediction results [J]. *Statistics & Decision*, 2015(24): 75-78 (in Chinese).
- [17] 曾九孙, 刘祥官, 罗世华, 等. 主成分回归和偏最小二乘法在高炉冶炼中的应用[J]. *浙江大学学报(理学版)*, 2009, 36(1): 33-36.
- ZENG Jiusun, LIU Xiangguan, LUO Shihua, *et al.* Application of principal component regression and partial least square in blast furnace iron-making [J]. *Journal of Zhejiang University (Science Edition)*, 2009, 36(1): 33-36 (in Chinese).
- [18] WOLD H. Nonlinear iterative partial least squares (NIPALS) modelling: some current developments [C]// KRISHNAIAH P R. *Proceedings of the Third International Symposium on Multivariate Analysis*. Ohio: Academic Press, 1973: 383-407.
- [19] EFRON B. A leisurely look at the bootstrap, the jackknife, and cross-validation [J]. *The American Statistician*, 1983, 37(1): 36-48.
- [20] LIN H M, LEE M H, LIANG J C, *et al.* A review of using partial least square structural equation modeling in e-learning research [J]. *British Journal of Educational Technology*, 2020, 51: 1-19.
- [21] 陈家鼎, 李东风. 多元回归中选择自变量的一种简单方法[J]. *应用概率统计*, 2015, 31(1): 71-88.
- CHEN Jiading, LI Dongfeng. A simple approach in regression variable selection [J]. *Chinese Journal of Applied Probability and Statistics*, 2015, 31(1): 71-88 (in Chinese).
- [22] 卢福容. 基于微型光谱仪的COD快速测定[D]. 重庆: 重庆大学, 2012.
- LU Furong. Rapid Determination of Chemical Oxygen Demand Based on Micro-spectrometer [D]. Chongqing: Chongqing University, 2012 (in Chinese).
- [23] LANGERGRABER G, GUPTA J K, PRESSL A, *et al.* On-line monitoring for control of a pilot-scale sequencing batch reactor using a submersible UV/VIS spectrometer [J]. *Water Science and Technology*, 2004, 50(10): 73-80.
- [24] CHEN B S, WU H N, LI S F Y. Development of variable pathlength UV-Vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring [J]. *Talanta*, 2014, 120: 325-330.

作者简介: 杨曼孜(1995- ), 女, 四川南充人, 硕士, 助理工程师, 主要研究方向为水质光谱建模技术。

E-mail: 947288605@qq.com

收稿日期: 2020-03-18

修回日期: 2020-07-29

(编辑: 沈靖怡)