

三维荧光光谱技术结合线性支持向量算法 在水体有机污染监测中的应用

戴源¹, 谢继征¹, 袁静¹, 沈薇¹, 郭宏达¹, 孙小平¹, 王志刚^{2*}

1. 江苏省扬州环境监测中心, 江苏 扬州 225100

2. 扬州大学环境科学与工程学院, 江苏 扬州 225009

摘要 针对当前地表水体有机污染的原位快速监测需求, 提出一种基于三维荧光光谱技术的水质指标预测模型和水质等级快速判断方法。以扬州市域内多种地表水体的水质监测数据作为模型训练样本, 充分利用水体三维荧光光谱信息, 结合线性支持向量回归算法(LIBLINEAR), 建立了与化学需氧量(COD_{Cr})、高锰酸盐指数(COD_{Mn})、氨氮(NH₃-N)、总磷(TP)、总氮(TN)和五日生化需氧量(BOD₅) 6项有机污染相关水质指标的预测模型。研究表明, 6项指标预测模型的训练集和测试集决定系数 R^2 均大于 0.73, 预测值与国标及行业标准方法分析结果的相关系数 r 达到 0.9 以上。利用水质指标预测结果进一步判断有机污染指标相关水质等级, 黑臭水体识别率达 86%, 对Ⅲ类~重度黑臭共 6 个水质等级的分类准确率为 60%。结果说明该方法通过水体三维荧光光谱信息预测水质有机污染指标具有较好的准确性和精度, 为广域时空尺度地表水的高效原位监测提供了一种新的解决方案。

关键词 三维荧光光谱; 线性支持向量回归; 水质指标; 水质等级; 原位监测

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)09-2839-07

引言

随着城市化进程的日益加快, 城市及周边地表水的污染源不断增加, 水质不断恶化, 对生态系统和人类健康构成威胁^[1]。城市及周边地表水中的有机污染物主要来自陆地生活源、地表径流、工业、服务业、养殖业和水生生物源污染, 以蛋白质、氨基酸、腐殖酸、脂肪等有机污染物为主。环境监测技术通过化学需氧量(COD_{Cr})、高锰酸盐指数(COD_{Mn})、氨氮(NH₃-N)、总磷(TP)、总氮(TN)和五日生化需氧量(BOD₅)等指标表征水体有机污染, 其中 COD_{Cr}, COD_{Mn} 和 BOD₅ 通常用于表示水体中有机污染物总量; NH₃-N, TN 和 TP 的含量升高会导致水体富营养化, 破坏生物多样性并产生臭味。

城市及周边地表水一直以来都是环境监测工作的重点, 但传统监测手段存在监测周期长、采样缺乏代表性、水样前处理复杂、分析难度高等困难, 往往造成监测数据时空分布不足, 监测数据滞后等问题, 因此开发连续、高效、低耗的

水质原位监测技术具有重要意义^[2]。近年来, 水体荧光光谱技术常被用来快速反演水体中 TP, TN, NH₃-N, BOD₅ 和 COD 等指标^[3-5], 避免了化学试剂的使用和复杂的水样前处理过程。三维荧光光谱技术可以在较宽的激发和发射波长范围内获取水体有机物丰富的光谱信息, 具有快速、可靠、实用的优点, 近年来被广泛应用于化学分析和环境监测领域。Yang 等^[6]使用激发发射矩阵三维荧光平行因子法(EEM-PARAFAC)对污水处理厂水样进行分析, 得到类蛋白等有机污染物的荧光特征峰位置, 利用多元线性回归算法(multiple linear regression, MLR)针对 COD 等有机污染指标建立预测模型, 实现对水处理效果的快速评价。陈方等^[7]使用平行因子算法(PARAFAC)分析苯酚等有机污染因子的三维荧光光谱, 提出针对清洁水和污水的二分类支持向量机(SVM)模型。但是, 现有研究大多针对模拟配制水样或单一类型的少量水质样本, 依赖已有的光谱特征经验选择算法, 从三维荧光光谱中提取若干点状光谱信息用于水质评价。由于地表水中有机污染物种类繁多, 各种物质的荧光峰位置和波段范围不同, 且存在荧光峰重叠现象, 因此这种仅选择少数光谱

收稿日期: 2020-06-09, 修订日期: 2020-10-11

基金项目: 国家重点研发计划项目(2016YFC1400602), 江苏省环境监测科研基金项目(1701)和江苏省重点研发计划项目(BE2016709)资助

作者简介: 戴源, 1986年生, 江苏省扬州环境监测中心工程师 e-mail: david1986213@hotmail.com

* 通讯作者 e-mail: wangzg@yzu.edu.cn

点的计算方法局限性强、泛化性能较差。

支持向量回归算法(support vector regress, SVR)是一种被广泛应用于机器学习和数据挖掘领域的算法模型。常规的 SVR 算法通过不同的核函数来构造非线性模型用以解决复杂的分类和回归问题,但是当样本量较大或特征维度较高时,SVM 算法存在消耗资源多、训练时间长等问题。LIBLINEAR 是一个针对线性分类场景而设计的工具包,支持线性 SVM 和线性逻辑回归等模型,可以对高维度大样本数据进行快速建模。该工具包采用热启动(warm-start)技术实现高效的参数寻优过程,并结合交叉验证方法得到最优惩罚参数 C 和不敏感度 ϵ ,具有建模速度快、计算精度高等特点。

本文对扬州市域内多种类型地表水进行了长期的三维荧光光谱采集和水质分析,形成了具有多样性和代表性的水质样本集合,首次将 LIBLINEAR 技术应用于三维荧光光谱水质监测,充分利用丰富的三维荧光光谱信息,将水体三维荧光光谱的全波段数据作为算法的输入,快速建立了 COD_{Cr} , COD_{Mn} , $\text{NH}_3\text{-N}$, TN , BOD_5 和 TP 等 6 项水质指标的预测模型,并且通过水质指标的预测结果进一步判断水体有机污染指标相关的水质等级,实现对城市及周边地表水水质指标和水质等级的快速原位监测。

1 实验部分

1.1 样品采集

从 2016 年 1 月至 2019 年 8 月,每月对扬州市域内 122 个地表水监测断面进行水样采集,使用直立采样器采集水面下 50 cm 深处的水体 5 L,静置 30 min 后取上层清液,按照水质采样规范平行分装在棕色玻璃瓶中,并于 4 °C 保存。采样现场同时测量水体的温度(T)、溶解氧(DO)含量和 pH 值。

水质监测断面共 122 个,涉及长江和淮河两大流域,覆盖了扬州市域内大部分的主要河流和湖泊,分布如图 1 所示。城市建成区内设有 87 个监测断面,囊括了 55 条城市河流和 4 个小型湖泊,其余 35 个监测断面分布在市域城郊及农村区域。根据 2016 年—2019 年扬州市水环境监测数据,监测断面水质等级包含 II 类~劣 V 类,此外还存在少量轻度

和重度黑臭断面。样本的采集时间涵盖了多个季节和枯丰水期,涉及水温、水位、水流和周边生态系统等多种环境因素变化对水质的影响,由此形成一个覆盖区域广、时间跨度长、水质变化多的样本集合。

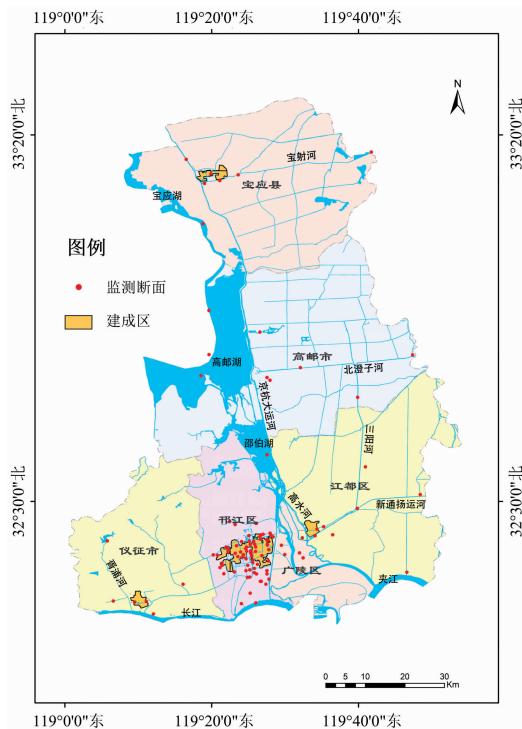


图 1 水质监测断面分布

Fig. 1 Distribution of water quality monitoring sections

1.2 化学分析

为保证化学分析与光谱分析的样品一致,将样品摇匀并静置 30 min 后取上层清液进行检测,检测方法参照相关国标和行业标准,使用仪器和具体分析方法见表 1。其中 pH 值、DO 和 T 在采样时现场测定,所有样品在采样 1 周内完成分析测试,测试结果见表 2。每项指标的测试结果中,最大值与最小值差异大,样本包括不同污染程度的多种水体。此外,本实验涉及的样本数量大、水质指标多,为建立水质指标预测模型提供有利条件。

表 1 仪器及分析方法

Table 1 Instruments and analysis methods

水质指标	仪器(型号)	分析方法
COD_{Cr}	—	HJ 828—2017《水质 化学需氧量的测定 重铬酸盐法》
COD_{Mn}	—	GB 11892—89《水质 高锰酸盐指数的测定》
$\text{NH}_3\text{-N}$	分光光度计(HACHDR6000)	HJ 535—2009《水质 氨氮的测定 纳氏试剂分光光度法》
TP	流动分析仪(SkalarSan++)	HJ 670—2013《水质 磷酸盐和总磷的测定 连续流动-钼酸铵分光光度法》
TN	流动分析仪(SkalarSan++)	HJ 636—2012《水质 总氮的测定 碱性过硫酸钾消解紫外分光光度法》
BOD_5	溶氧仪(YSI5000)	HJ 505—2009《水质 五日生化需氧量(BOD_5)的测定 稀释与接种法》
pH	便携式 pH 计 YSI pH100A	GB 6920—86《水质 pH 值的测定 玻璃电极法》
DO/T	便携式溶氧仪 YSI PRO20	HJ 506—2009《水质 溶解氧的测定 电化学探头法》

表 2 化学分析结果
Table 2 Results of chemical analysis

水质指标	最大值/ (mg · L ⁻¹)	最小值/ (mg · L ⁻¹)	平均值/ (mg · L ⁻¹)	样本数
COD _{Cr}	255	4	22.3	1 266
COD _{Mn}	35.9	1.8	6.4	1 354
NH ₃ -N	69.2	0.04	3.47	1 266
TP	6.64	0.01	0.39	1 266
TN	84.2	0.04	4.79	819
BOD ₅	122	0.11	6.41	579

1.3 三维荧光光谱测量

采用中国科学院安徽光学精密机械研究所改造的日立 F4600 型荧光分光光度计测量水样三维荧光光谱, 该仪器在保留原有光路设计的基础上添加自动进样和清洗装置, 在底部加装避震装置, 可实现在水质自动站或监测车中的连续快速原位监测。

每次对空白样品 (Milli-Q 超纯水) 进行扫描后再进行水样测量。三维荧光光谱测量前, 先将水样摇匀后静置至室温。若水样的荧光强度超出仪器测量范围, 须用超纯水稀释。样品光谱平行测试的相对精度偏差应小于 2%^[4], 同批水样的光谱分析与化学分析时间间隔不超过 24 h。光谱测量参数设置如下: 激发波长 E_x 为 220~400 nm, 采样间隔 5 nm; 发射波长 E_m 为 260~520 nm, 采样间隔 1 nm; 狭缝宽度为 10 nm, 扫描速度为 12 000 nm · min⁻¹。

1.4 模型建立

本文采用 MATLAB2019 (Mathworks, Natick, MA, USA) 软件构建水质指标预测模型。训练集与测试集样本的划分采用随机抽样法, 抽取 20% 的样本作为测试集, 用于评价模型的泛化能力和预测效果, 剩余样本作为训练集用于建立预测模型。

1.4.1 数据预处理

使用 Delaunay 三角形内插值法对原始光谱中包含的瑞利散射和拉曼散射进行修正。为消除实验环境变化和光谱仪光源波动的影响, 从样品光谱中扣除空白样品光谱, 并用空白样品在 E_x = 348 nm 和 E_m = 397 nm 处的拉曼峰强度值对去散射处理后的样品光谱强度值进行拉曼归一化处理^[8]。

1.4.2 线性支持向量回归模型 (LIBLINEAR)

本文将每个激发-发射波长对应的荧光强度作为水质指标的潜在预测因子, 为降低数据冗余度、提高模型收敛度, 将预处理后的三维荧光光谱去除激发波长大于发射波长的光谱区域, 结合 T, DO 和 pH 值形成 7601 维向量。将该向量作为算法的输入, 以各水质指标的化学分析结果作为算法目标值, 使用 LIBLINEAR 工具包建立 L2 正则 L2 误差支持向量回归模型 (L2-regularized L2-loss support vector regression), 通过调整权重向量 ω, 使 L2 正则项与 L2 误差项之和最小。

$$\min_{\omega} \left(\frac{1}{2} \omega^T \omega + C \sum_{i=1}^N (\max(0, |y_i - \omega^T \omega| - \epsilon))^2 \right) \quad (1)$$

式(1)中, y_i 为各水质指标的化学分析结果, N 为样本数量, C 为惩罚因子, ε 为不敏感度。运用网格点搜索法 (grid

searching technique) 和交叉验证法 (cross validation) 相结合对 C 和 ε 因子进行自动寻优, 即建立双层循环交叉验证, 外层对 ε ∈ ($\frac{19}{20} \epsilon_{\max}, \dots, \frac{1}{20} \epsilon_{\max}, 0$) 进行循环, 针对每一个 ε 建立内层循环对 C ∈ (C_{min}, 2C_{min}, 4C_{min}, ..., C_{max}) 进行交叉验证, C 和 ε 的选择范围由算法根据训练集自动设定, 以每组交叉验证集的均方根误差 (CVMSE) 的最小值为寻优指标, 选择对应的 C 和 ε 用以建立模型对测试集进行预测。

1.4.3 模型评价标准

使用决定系数 R² 和均方根误差 RMSE 作为本文模型效果的评价指标。训练集决定系数 R_t² 和测试集决定系数 R_p² 越接近 1, 说明模型相关性越高、预测效果越好; 训练集均方根误差 RMSEC 和测试集均方根误差 RMSEP 的数值越小说明模型精度越高, 泛化能力越强。训练集和测试集决定系数按照式(2)计算。训练集和测试集均方根误差按式(3)计算。

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (3)$$

式中, \hat{y}_i 为算法预测值, y_i 为各水质指标的化学分析结果, \bar{y} 为各水质指标真实值的均值, N 为样本数量。

1.4.4 水质分类标准及方法

表 3 是根据 GB3838—2002《地表水环境质量标准》和《城市黑臭水体整治工作指南》中的标准限值制定的水质分类标准, 基于该标准使用模型预测结果对有机污染指标相关的水质等级进行判断。针对不同的水质判断需求, 本文设计了如表 4 所示的 4 种水质分级方法。方法的分级数量越多, 对水

表 3 水质分类标准限值

Table 3 Water quality classification standard limits (mg · L⁻¹)

水质指标	Ⅲ类	Ⅳ类	Ⅴ类	轻度黑臭	重度黑臭
COD _{Cr}	20	30	40	—	—
COD _{Mn}	6.0	10	15	—	—
NH ₃ -N	1.0	1.5	2.0	8.0	15
TP	0.2	0.3	0.4	—	—
TN	1.0	1.5	2.0	—	—
DO	5.0	3.0	3.0	2.0	0.2
BOD ₅	4.0	6.0	10	—	—

表 4 四种水质分级方法

Table 4 Four water quality classification methods

等级	方法 1	方法 2	方法 3	方法 4
1	非黑臭	优于Ⅴ类	优于Ⅳ类	优于Ⅲ类
2	轻度黑臭	劣Ⅴ类	Ⅴ类	Ⅳ类
3	重度黑臭	轻度黑臭	劣Ⅴ类	Ⅴ类
4	—	重度黑臭	轻度黑臭	劣Ⅴ类
5	—	—	重度黑臭	轻度黑臭
6	—	—	—	重度黑臭

质状况的区分越细致。其中“劣 V 类”在本文中定义为超过 V 类标准限值但尚未达到轻度黑臭的水体。

2 结果与讨论

2.1 水质指标模型及预测结果

图 2 为水质预测模型中三维荧光光谱权重的分布图,如图 2 所示,各模型权重较大的三维荧光光谱位置主要分布在 7 个荧光区域,荧光区域的范围和对对应组分信息如表 5 所示,可知 A—F 均为水体中常见溶解性有机污染物的特征荧光区域,其中 A 与水体中的分子量较大的类腐殖酸相关;B 区域内的荧光峰常出现在城市废水光谱中,被认为是与微生物相关的类腐殖物质(可溶性微生物副产物)^[8];C 对应类富里酸的荧光特征峰,其来源为陆源前驱染物^[9];D 为游离态类色氨酸的荧光峰,其光谱值与水体中微生物细胞数量紧密相关,可以表征水生态系统的微生物活性^[9];E 和 F 为酪氨酸等芳香族蛋白质的特征光谱范围,主要来自生活源有机污染^[10];G 被定义为类色氨酸的特征光谱区域,其光谱强度同采样断面与污染源排口之间的距离和水中污染物的新鲜程

度有关^[9]。

由图 2 可知,6 项水质指标预测模型中正权重均主要分布于 6 个荧光区域内(A—F),而负权重主要集中在 G 区域中,说明 6 项水质指标预测值均与色氨酸、酪氨酸、类腐殖酸、类富里酸和类蛋白等有机污染物的荧光强度成正比,与 G 区域的荧光值成反比。此外,各预测模型的权重分布略有不同,其中 COD_{Cr} 模型正权重的覆盖范围大于 COD_{Mn} ,说明有更多的有机物荧光信号会对 COD_{Cr} 的预测结果产生正影响;D, E 和 F 所代表的蛋白质和氨基酸荧光区域在 $\text{NH}_3\text{-N}$ 和 TN 模型中具有较大的正权重,并且 TN 模型具有更大的正权重范围;TP 的正权重分布较为集中在类腐殖酸特征范围内;BOD₅ 的正权重集中在 D 区域内,说明 BOD₅ 的预测值与色氨酸光谱强度具有较高相关性,这与 Henderson 等的研究结论一致^[11]。由此可知,基于全波段的预测模型与以往的固定点式光谱模型相比,该模型能够针对不同的水质指标对每一个光谱位置设置相应的权重,并且模型权重分布符合水质指标与有机污染物的逻辑关系,可以更加充分地利用水体三维荧光光谱信息建立光谱与有机污染物之间的定量关系。

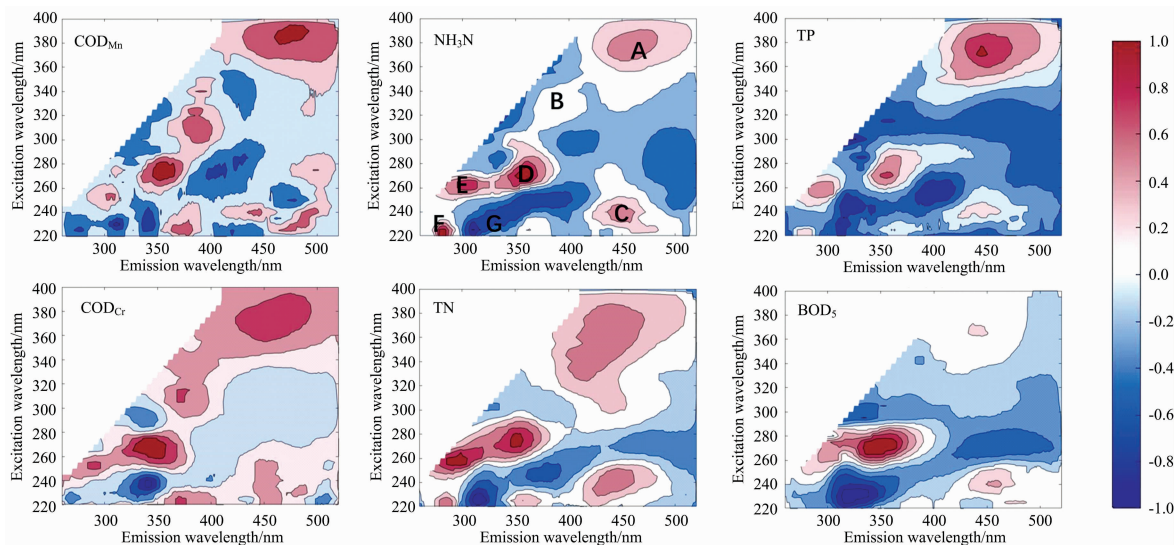


图 2 水质预测模型权重分布

Fig. 2 Weight distribution of water quality prediction model

表 5 荧光区域范围及组分

Table 5 Fluorescence regions and components

荧光区域	成分	荧光峰 $E_x/E_m/\text{nm}$	参考文献
A	类腐殖酸(humic acid-like)	365/470	[12]
B	微生物副产物(microbial byproduct-like)	290~320/380~420	[10]
C	类富里酸(fulvic-like)	230~260/400~480	[8]
D	游离态类色氨酸(free tryptophan-like)	275~285/320~350	[13]
E	游离态类酪氨酸(free tyrosine-like)	270/308	[14]
F	类酪氨酸(tyrosine-like)	220~237/280~300	[8]
G	类色氨酸(tryptophan-like)	215~237/340~381	[8]

基于全波段光谱模型的水质指标预测结果如表 6 所示,各水质指标预测模型的训练集决定系数 R_c^2 均大于 0.8,且

R_p^2 接近于 R_c^2 ,说明 LIBLINEAR 算法的预测值与各水质指标的化学分析值拟合度均较高。此外,各水质指标预测模型

的 RMSEP 与 RMSEC 的差距较小, 说明训练得到的模型精度高、预测能力强、泛化能力好。

图 3 为使用 Matlab2019 软件进行的模型预测值与化学分析值的相关性分析。从图 3 可知, COD_{Cr}, COD_{Mn}, NH₃-N, TN, BOD₅ 和 TP 六项水质指标的预测值和实际测量值之间的相关系数 R 分别为 0.95, 0.92, 0.92, 0.91, 0.94 和 0.90, 并且通过了 $p = 0.05$ 的显著性水平检验, 说明 LIBLINEARSVM 模型预测的各水质指标结果与国标及行业标准分析结果具有较高的拟合度, 证明了三维荧光光谱技术用于监测水质污染状况的可行性, 本方法可为城市及周边地表水的快速、原位、高效监测提供解决方案。

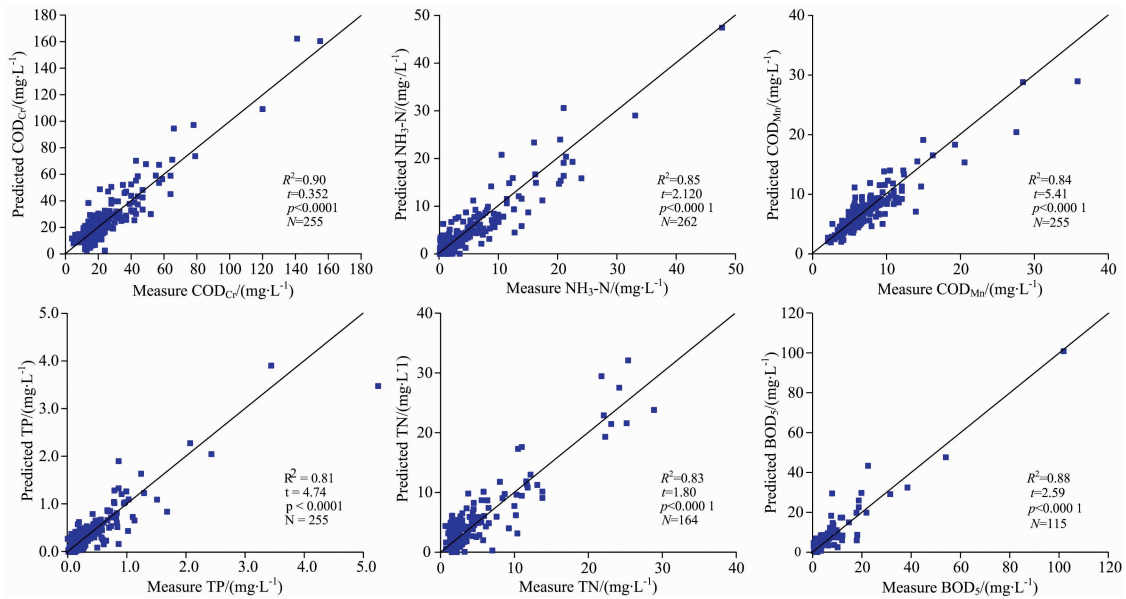


图 3 模型预测值与化学分析值的相关性

Fig. 3 Correlation between predicted value and chemical analysis results

2.2 水质分类结果

为了验证水质分类的预测效果, 使用 LIBLINEAR 模型对 100 个未知水样进行水质指标预测, 并用预测结果按照表 3 中的标准判断其水质类别。分类预测效果通过准确率 A 和 F_1 分数两个指标评价。其中 A 代表正确判断的样本数和总样本数的比值; F_1 分数是查准率 P 和查全率 R_e 的调和平均数, 可以综合评价分类效果。计算公式如式(4)

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R_e = \frac{TP}{TP + FN} \quad (6)$$

$$F_1 = \frac{2 \times P \times R_e}{P + R_e} \quad (7)$$

其中, TP 为真正例样本数, TN 为真反例样本数, FP 为假正例样本数, FN 为假反例样本数。

图 4 为表 4 中 4 种分级方法的水质分类结果。如图 4 所示, 方法 1, 2, 3 和 4 的水质分类准确率分别为 86%, 74%, 67% 和 60%, F_1 分数分别 0.93, 0.88, 0.84 和 0.77, 随着

表 6 模型预测结果

Table 6 Prediction results of models

水质指标	训练集		测试集	
	R_c^2	RMSEC/ ($\text{mg} \cdot \text{L}^{-1}$)	R_p^2	RMSEP/ ($\text{mg} \cdot \text{L}^{-1}$)
COD _{Cr}	0.81	8.53	0.80	8.83
COD _{Mn}	0.87	1.23	0.84	1.65
NH ₃ -N	0.88	2.00	0.82	2.63
TP	0.80	0.25	0.73	0.25
TN	0.84	3.19	0.78	3.29
BOD ₅	0.91	3.00	0.87	3.19

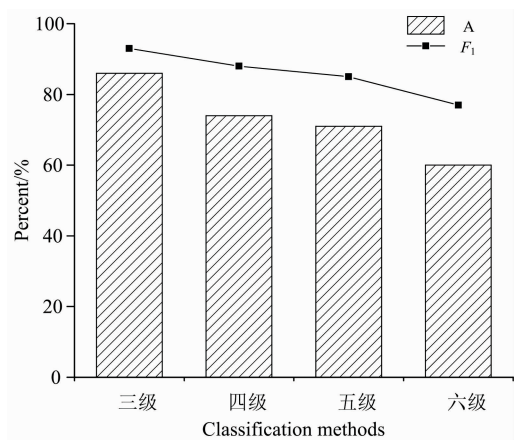


图 4 水质分类结果

Fig. 4 Water quality classification results

分级的细化, 水质分类准确率和 F_1 分数有所下降, 说明水质指标预测结果对清洁水体的细化分类稍有不足, 但对较重污染水体的水质分级具有较高的正确率和识别精度。总之, 本方法可以快速判断水质等级, 并同步显示超标污染物及其

浓度值,实现对地表水水质的高效监测和精准评价。

3 结 论

对扬州市域内 122 个地表水监测断面的三维荧光光谱信息和水质状况进行了长期积累形成了覆盖范围广、时间跨度长、水质变化多的样本集合,基于全波段光谱数据使用 LIB-LINEAR 算法建立了针对 COD_{Cr} , COD_{Mn} , $\text{NH}_3\text{-N}$, TN, BOD_5 和 TP 六项水质指标的预测模型,模型的权重分布与多种溶解性有机物的荧光特征区域重合,说明该模型可以综

合地反应地表水中的有机污染程度。模型预测结果具有较高的决定系数和较低的均方根误差,测试集的预测结果与实际测量值之间的相关系数达到 0.90 以上。此外,使用水质指标的预测结果对水体的水质等级进行判断,其中对黑臭水体判断正确率达 86%,对 III 类~重度黑臭水体的分类准确率达 60%,表明该技术的水质指标预测结果与现行的国标及行业标准方法分析结果一致性较高,可以用于在广域时空尺度中对流域水体的整体水质状况进行全面判识,是一种快速、原位、高效的都市及周边地表水水质监测技术。

References

- [1] Zhao Y, Song K, Li S, et al. *Environmental Science and Pollution Research*, 2016, 23(15): 15381.
- [2] Li P, Hur J. *Critical Reviews in Environmental Science and Technology*, 2017, 47(3): 131.
- [3] Wang X, Zhang F, Kung H T, et al. *CATENA*, 2017, 155: 62.
- [4] Yang L Y, Hur J, Zhuang W. *Environmental Science and Pollution Research*, 2015, 22(9): 6500.
- [5] Chen B S, Wu H N, Li S F Y. *Talanta*, 2014, 120: 325.
- [6] Yang L Y, Shin H S, Hur J. *Sensors*, 2014, 14: 1771.
- [7] CHEN Fang, ZHANG Xiao-yan, HUANG Ping-jie, et al(陈 方, 张晓燕, 黄平捷, 等). *Journal of Zhejiang University • Agriculture and Life Sciences*(浙江大学学报·农业与生命科学版), 2016, 42(3): 368.
- [8] Rodríguez-Vidal F J, García-Valverde María, Ortega-Azabache B, et al. *Journal of Environmental Management*, 2020, 263: 110396.
- [9] Carstea E M, Bridgeman J, Baker A, et al. *Water Research*, 2016, 95: 205.
- [10] Yu H, Song Y, Du E, et al. *Environmental Science and Pollution Research*, 2016, 23: 10644.
- [11] Henderson R K, Baker A, Murphy K R, et al. *Water Research*, 2009, 43(4): 863.
- [12] Mangalgi K P, Timko S A, Gonsior M, et al. *Environmental Science & Technology*, 2017, 51(14): 8036.
- [13] Li J C, Wang L, Geng J J, et al. *Chemosphere*, 2020, 251: 126375.
- [14] Xie M W, Chen M, Wang W X. *Environmental Pollution*, 2018, 243B: 1528.

Application of Excitation-Emission Matrix (EEM) Fluorescence Combined With Linear SVM in Organic Pollution Monitoring of Water

DAI Yuan¹, XIE Ji-zheng¹, YUAN Jing¹, SHEN Wei¹, GUO Hong-da¹, SUN Xiao-ping¹, WANG Zhi-gang^{2*}

1. Jiangsu Province Yangzhou Environmental Monitoring Center, Yangzhou 225100, China

2. College of Environmental Science and Engineering, Yangzhou University, Yangzhou 225009, China

Abstract In view of the increasingly serious organic pollution of urban waterbodies, this paper proposes a water quality indexes prediction model based on excitation-emission matrix (EEM) fluorescence technology and a method for quickly judging the water quality category. In this study, a large number of diversified surface waters around Yangzhoucity were taken as the training sample of the model. Based on the EEM spectrum of water and linear support vector regression (LIBLINEAR), the prediction models of six water quality indexes were established, including chemical oxygen demand (COD_{Cr}) and permanganate index (COD_{Mn}), ammonia nitrogen ($\text{NH}_3\text{-N}$), total phosphorus (TP), total nitrogen (TN) and five-day biochemical oxygen demand (BOD_5). The test results show that the determination coefficient R^2 of the training set and the test set of the six index prediction models are both greater than 73%, while the correlation coefficient between the predicted value and analysis results by the national standard and industry-standard methods is greater than 0.9. Base on the prediction results of the water quality index, the water quality category could be the further judge. The recognition rate of black-odor waterbody reached 86%, and the classification accuracy rate of water bodies above category III was 60%. The results show that the method has good accuracy and precision in predicting the water quality index through the three-dimensional fluorescence spectrum information of the waterbodies, which provides a solution for the efficient in-situ monitoring and rapid classification of water quality of urban and surrounding surface water.

Keywords EEM spectrum; Linear support vector regression; Water quality indexes; Water quality grade; In situ monitoring

* Corresponding author

(Received Jun. 9, 2020; accepted Oct. 11, 2020)

《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的专业学术期刊。国内外公开发行人,从2004年起为月刊,大16开本,2020年仍为月刊,每期332页。《光谱学与光谱分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以8000字(包括图表、参考文献、作者姓名、单位和中文、英文摘要,下同)为宜。
2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过5000字。
3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过10000字。
4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过5000字。
5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以3000~4000字为宜。

稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。
2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以800字为宜,英文摘要(建议经专业英语翻译机构润色)与中文摘要要对照;另附关键词。要求来稿应达到“齐、清、定”,中文、英文文字通顺,方可接受送审。
3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊只接收网上在线投稿,不接收以邮寄方式或e-mail方式的投稿,严禁“一稿多投”,对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。
4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用6号字。电子文档中除实物图外,曲线图要用Matlab, Excel, Visio或Origin等软件制作,稿件中图片的原图并转成相应(可编辑)的文件格式(.fig, .xls, .vsd, .opj),非“.jpg”格式的文档,随电子版修改稿一同发送到本刊的修改稿专用邮箱。
5. 文中出现的单位必须按“中华人民共和国计量标准”及有关GB标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。
6. 名词术语,请参照全国科学技术名词规定缮写。
7. 参考文献,采用顺序编码制,只列主要文献;以15~20条为宜。内部资料、私人通讯、未经公开发表的一律不能引用。日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。
8. 请在投稿第一页左下角写明投稿联系人的电话和两个e-mail,以便及时联系。

稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修改稿在30天内寄回编辑部,并作为作者最终定稿(当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动),若二个月内编辑部没收到修改稿,将视为自行撤稿处理。
2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。
3. 来稿一经发表将酌致稿酬并送样刊2册。
4. 遵照《中华人民共和国著作权法》,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给《光谱学与光谱分析》期刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一律认为已获作者授权认可。
5. 修改稿请寄:100081 北京市海淀区学院南路76号(南院西北门),《光谱学与光谱分析》期刊社(收)

电话:010-62182998 或 62181070

传真:010-62181070

e-mail: chngpxygpfx@vip.sina.com;

修改稿专用邮箱: gp2008@vip.sina.com

网址: http://www.gpxygpfx.com