

基于偏最小二乘的絮体图像信息与水质的相关性

和梦雪¹, 刘健¹, 范文玉¹, 赵立杰², 左越²

(1. 沈阳化工大学环境与安全工程学院, 辽宁 沈阳 110142;

2. 沈阳化工大学信息工程学院, 辽宁 沈阳 110142)

摘要:针对污水处理运行过程 COD、BOD₅、TN 和 TP 实时测量的问题,提出一种基于絮体形态参数的偏最小二乘法模型。采用图像分析技术对污泥絮体形态特征进行提取,基于相关性从提取的形态参数和试验期间的运行参数中选择模型的输入变量,通过偏最小二乘法建立输入变量与 4 个水质指标的预测关系模型。结果表明,COD、BOD₅、TN、TP 的交叉有效性系数分别为 0.736、0.682、0.839、0.618,絮体形态参数与水质指标有明显的相关性,可用于预测出水水质。

关键词: 图像处理技术;偏最小二乘法;出水指示;活性污泥

中图分类号:X703

文献标志码:B

文章编号:1006-2009(2021)06-0048-04

Correlation Between Floc Morphology and Water Quality Based on Partial Least Squares

HE Meng-xue¹, LIU Jian¹, FAN Wen-yu¹, ZHAO Li-jie², ZUO Yue²

(1. College of Environmental and Safety Engineering, Shenyang University of Chemical Technology,

Shenyang, Liaoning 110142, China; 2. College of Information Engineering, Shenyang University of

Chemical Technology, Shenyang, Liaoning 110142, China)

Abstract: Based on floc morphology parameters, a partial least squares model was established to solve the problem of real-time measurement of COD, BOD₅, TN and TP during the operation of sewage treatment. The floc morphological characteristics of sludge were extracted by image analysis technology. Based on the correlation, the input variables of the model were selected from the extracted morphological parameters and the operation parameters during the experiment. The prediction model of the relationship between input variables and four water quality indexes was established by partial least squares. The results indicated that the cross validity coefficients of COD, BOD₅, TN and TP were 0.736, 0.682, 0.839 and 0.618, respectively. There was a significant correlation between floc morphology parameters and water quality indexes. This model could be used to predict effluent quality.

Key words: Image processing technology; Partial least squares(PLS); Effluent indication; Active sludge

活性污泥絮体是活性污泥法处理污水的关键,污泥絮体按大小被分成 3 个层次:单个细菌、小菌落和活性污泥絮体^[1]。细菌组成失去平衡,会造成污泥絮体絮凝和沉降能力失调,污水处理效率降低。为了保证水厂的正常运行,需要对水质进行监测,近些年,水质的在线监测成为主要研究方向,主要有电化学分析法、遥感监测、生物传感技术分析法等^[2-4]。虽然这些方法具有实时、快速、连续等优点,但不能提供有关过程问题背后的必要信息,也不能对水质的变化做出提前响应。

随着图像分析技术的发展,可以对系统的生物相构成进行定性分析,也可以给出定量的信息来判断污水处理系统的运行状况等^[5-6]。然而,目前该技术的应用忽略了水温、好氧池溶解氧(DO)含量等运行参数。今对沈阳某 A²O 污水处理厂好氧池

收稿日期:2020-10-21;修订日期:2021-10-25

基金项目:国家重点研发计划“工业互联网边缘计算节点设计方法与运行关键技术”基金资助项目(2018YFB1700200)

作者简介:和梦雪(1998—),女,山西晋城人,在读研究生,研究方向为水处理技术。

活性污泥进行为期6个月的研究,收集这段时间的运行参数,通过偏最小二乘法(PLS)建立絮体结构特征和运行参数与好氧池出水水质参数(COD、BOD₅、TN、TP)间的关系模型,分析两者之间的相关性,并利用该模型对出水水质进行预测。

1 试验

1.1 主要仪器与试剂

UV-2600A型紫外可见分光光度计,上海仪器有限公司;Nikon ECLIPSE TS100/100-F型倒置生物显微镜;Toup TekTou Camucoms 03100kpa型显微镜接工业数码相机。

氢氧化钠(分析纯),西陇化工股份有限公司;磷酸二氢钾(分析纯),北京市化学试剂研究所;七水合硫酸镁、硫酸银、过硫酸钾(分析纯),天津博迪化工股份有限公司。

1.2 污泥样品及出水水样采集

用于分析的活性污泥来自沈阳某A²O污水处理厂好氧池。试验在好氧池固定点液面50 cm处采集污泥样本,此处的污泥样本有良好的絮凝和沉降性能。在采集泥样的同时,在好氧池出水口采集瞬时水样,测定水质指标,以便与好氧池污泥建立直接联系。每次污泥样品和出水样品均取500 mL,样品采集与图像采集、水质指标测定之间的间隔时间不超过3 h^[7]。试验持续6个月,平均每2 d取1次样品,共采集91组样本。

1.3 水质指标测定

好氧池出水为泥水混合液,在测定前需要对水样做预处理。先将水样过滤,取滤液调节其pH值至中性,再将水样稀释至合适范围,待测。BOD₅、COD、TN和TP的测定分别采用《水质 五日生化需氧量(BOD₅)的测定 稀释接种法》(HJ 505—2009)《水质 化学需氧量的测定 重铬酸钾法》(GB 11914—89)《水质 总氮的测定 过硫酸钾紫外分光光度法》(GB 11894—89)和《水质 总磷的测定 钼酸铵分光光度法》(GB 11893—89)。

1.4 图像采集与处理

污水处理厂通常以高生物含量运行,在图像采集过程中,用蒸馏水稀释样本以获得足够的污泥特性^[8]。为确定合适的稀释倍数,考察了3种不同的稀释比(1:10、1:50和1:100)。使用不同稀释比时,未发现絮体形态有显著差异,最终选择1:10的比例用于图像采集,在该比例下能获得最多的图像

信息。图像采集前,将稀释好的样品混合均匀,用移液枪移取10 μL置于载玻片上,覆上18 mm × 18 mm的盖玻片。将载玻片置于倒置光学相差显微镜载物台上。显微镜接工业数码相机,并采用配套的Toupview软件进行显微图像采集。为了最大限度地减少取样误差,每个样品做3次平行试验,在总100 ×放大倍数下对载玻片上、中、下3个不同位置共采集36张图像,保存为2 048 × 1 536像素的TIF格式图片。每个样品总共获得108张图片(36 × 3个重复),以获得污泥的代表性信息。试验期间,91组样品共9 828张图片(108 × 91),1张图片大概包含20 ~ 50个絮体,可认为污泥样本的所得结果在统计上是可靠的。图像处理模型采用基于DCNN改进的DeepLabv3+模型,并利用MATLAB的默认方法计算得到的阈值进行二值化。将絮体、丝状菌和背景分割,计算并提取形态学特征参数。

1.5 变量选择

影响出水水质的因素有很多,比如进水流量、进水负荷、进水SS、曝气池溶解氧、水温、pH值、曝气池氧化还原电位(ORP)、曝气池混合液悬浮固体浓度(MLSS)等,这些因素都直接或间接影响着出水水质。变量选择是数据分析和模型开发中最重要的步骤之一,基于之前Amarala和Ferreira^[7]对污泥絮体的研究,试验选取25个形态参数。此外,还收集了该污水厂试验期间6个月的运行参数。基于相关性的选择,将多重相关系数的绝对值>0.90的变量和具有较低相关系数的变量移除,最终确定了21个自变量(见表1),在絮体形态参数中,除了絮体数目外,其他参数均取平均值。

1.6 数据分析

收集的数据通过偏最小二乘法进行关联,通过拟合系数 R^2 和交叉有效性系数 Q^2 来判断模型的适用性,并根据 Q^2 来确定最佳拟合模型。 R^2 指因变量与所有自变量之间的总体关系,可以体现对模型的解释能力,见公式(1), Q^2 表示对因变量的预测能力,见公式(2)。一般认为,当 $R^2 > 0.5$ 且 $Q^2 > 0.5$ 时,模型拟合效果较好。这种方法最重要的优点是依靠迭代算法可对多重共线性数据进行处理,特别是当两组变量的个数很多,且都存在多重相关性,而观测数据的数量(样本量)又较少时^[9]。在该研究中,输入矩阵 X 包含21个输入变量,输出矩阵 Y 由输出变量COD_{out}、BOD_{out}、TN_{out}和TP_{out}组

表 1 污水处理厂参数

Table 1 Parameters from sewage plants

自变量性质	名称	符号	定义	自变量性质	名称	符号	定义
数目	絮体数目	N	絮体在图像中的个数	规则性	形状因子	FF	絮体形状与圆的偏差
大小	絮体平均面积	A	每个絮体所占平面图形的平均大小		分形维数	FD	基于分形理论絮体不规则性的量度
	轮廓线周长	P_{conv}	絮体凸形包络线长度		凸率	C_{ov}	絮体凸包络周长与轮廓线总长度之比
	周长	P	絮体轮廓线总长度	运行参数	进水 COD	COD_{in}	
	最大 Feret 直径	F_{max}	絮体轮廓两边平行线间的最大距离		进水 BOD	BOD_{in}	
	当量直径	D_{eq}	与絮体面积相同的圆的直径		进水 TN	TN_{in}	
密实度	孔率	HR	去除内孔面积与总絮体面积之比		进水 TP	TP_{in}	
规则性	圆度	RD	絮体面积接近理论圆的程度		好氧池水温	T	
	长径比	RR	絮体的延展性		好氧池溶解氧浓度	DO	
	长短轴比	A_{sp}	絮体长轴与短轴之比		好氧池 pH 值	pH 值	
	充实度	E_{st}	絮体面积与外切矩形面积之比				

成。试验共得到 91 个样本,用前 80 个样本进行训练,其余 11 个样本用于测试开发的模型。

$$R^2 = \frac{\text{SSR}}{\text{SST}} \quad (1)$$

$$Q^2 = 1 - \frac{\text{PRESS}_h}{\text{SS}_{h-1}} \quad (2)$$

式中: SSR 表示回归平方和; SST 表示总平方和^[10]; PRESS_h 为每一个维度的预测误差平方和; SS_{h-1} 是前一个维度的残差平方和^[11]。

2 结果与讨论

2.1 PLS 模型的回归拟合及结果分析

基于 PLS 的建模原理,根据累计交叉有效性系数 $Q^2(\text{cum})$ 提取最佳主成分^[12]。在数据标准化后,当提取 4 个主成分时,模型达到最佳预测效果,此时输出矩阵累计拟合系数 $R^2(\text{cum}) = 0.773$,累计交叉有效性系数 $Q^2(\text{cum}) = 0.721$ 。对于单个输出变量来说, COD_{out} 、 BOD_{out} 、 TN_{out} 和 TP_{out} 的交叉有效性系数分别为 0.736、0.682、0.839、0.618,输出变量均达到了较好的拟合效果。

活性污泥絮体、温度、pH 值^[13]直接或间接影响着出水水质,为直观地观察这些变量对输出变量的影响,绘制各个输出变量的回归系数图。从整体来看,输入变量对输出变量的影响程度大体一致,絮体大小及规则参数与输出变量浓度成正比。在活性污泥絮体中,丝状菌为絮体提供网状的骨架结构,同时引导絮体生长为多种形状,当丝状菌大量生长时,絮体面积增加。虽然丝状菌的存在有利于污水的净化效果,但由于絮体呈敞开式结构,絮体沉降缓慢,从而使得污泥流失,出水水质变差。密实度参数 HR 与输出变量浓度成反比,当絮体呈现

致密的状态时,细菌就会相互堆积,导致污泥沉降速度加快,泥水分离效果提高,这样会提高絮体对污染物的去除效率^[14]。对于运行参数来说,除了温度参数,其他参数均与输出变量浓度成正比。

单独来看,对于每个输出变量,絮体数目 N 与絮体平均面积 A 呈现出相反的相关性。由此推断,在絮体总量一定的情况下,絮体间存在“聚集”和“解体”现象,絮体数目增加往往是由于大絮体“解体”为小絮体引起的;当小絮体数目增加时,难以沉降的小絮体随出水排出,小絮体含有大量的有机物质,从而导致输出变量浓度增加,这点与 Wilén 等^[15]的观点相吻合。变量凸率与输出变量浓度成反比,当絮体边缘越粗糙时,絮体表面与污水的接触面积越大,污染物的降解速率越快,出水水质指标浓度越低。变量 FD 与输出变量浓度成正比,现有絮凝动力学模型认为,絮体具有基本粒子-菌胶团-絮体的多层结构,不同层次具有不同的分形维数,絮体的尺寸越大,所含有的基本粒子越多,FD 越小^[16]。上述试验中,变量 FD 与变量 A 成反比,这与此观点相一致。

进水污染物浓度与输出变量浓度成正比,这是由于在相同构筑物尺寸和操作条件的情况下,进水污染物浓度的增加使进水有机负荷增加,将导致污水处理过程出水的污染物浓度相应增大。另外,溶解氧浓度与输出变量成正比,结合絮体的密实度参数 HR 来看,溶解氧浓度与变量 HR 成反比,当好氧池溶解氧含量过高时,会引起活性污泥结构松散问题,导致污泥的沉降性能变差,出水水质变差。

2.2 PLS 模型的拟合效果分析

对 80 个训练样本进行模型训练,若设观测值

为 y , 预测值为 $\hat{y}$, 相对误差为 r , 则 $r = (y - \hat{y})/y$ 。在模型中, 变量 TN_{out} 的预测能力最高, 样本点的预测值与实际值相对误差均不超过0.2%; 其次是 COD_{out} 和 BOD_{out} , 样本点相对误差分别在0.4%和0.6%以内; 拟合效果最差的是 TP_{out} , 部分样本点相对误差超过1%。为检测模型的预测能力, 使用11个测试样本对模型进行测试, 结果见图1(a)(b)(c)(d)。由图1可知, 对于这4个输出变量, 每个测试样本的预测值和观测值均相似, 这意味着通过污泥絮体参数和运行参数预测出水水质的可行性, 表明该模型有着较好的预测能力。

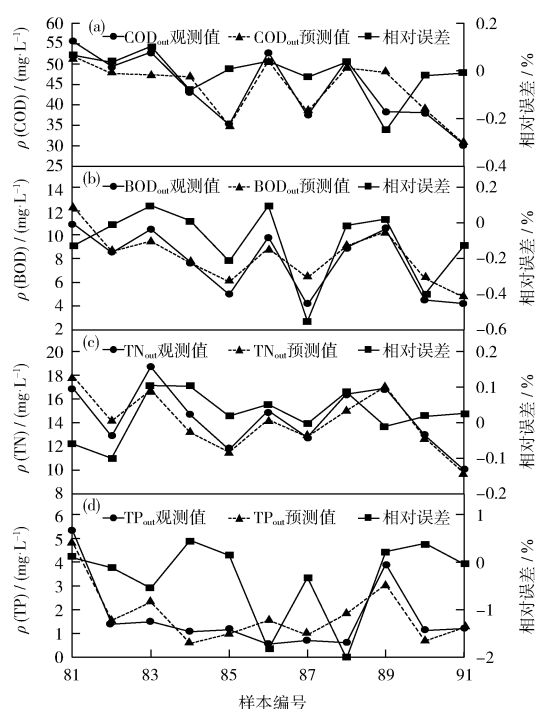


图1 测试样本观测值与预测值比较

Fig.1 Comparison of observed and predicted values of test samples

3 结语

试验研究了运行参数、污泥絮体形态参数与好氧池出水水质之间的相关性, 并通过偏最小二乘法预测了 COD 、 BOD_5 、 TN 和 TP 浓度。在对模型的整体分析中, 发现输入变量对输出变量的影响程度大体一致, 出水指标浓度与大部分输入变量成正比。在絮体形态特征的分析中, 发现当絮体与圆形偏差越大, 或絮体边缘越光滑时, 出水指标浓度越高。

在对模型的评价中, 模型输出与测试目标基本保持一致的趋势, 在测试样本集中, 每个样本的预

测值和观测值均相似, 表明模型有较好的预测能力。

[参考文献]

- [1] 龙向宇, 唐然, 方振东, 等. 活性污泥的絮体结构模型[J]. 城市环境与城市生态, 2010, 23(2): 36-39.
- [2] 翟崇治. 水质监测自动化与实践[M]. 北京: 中国环境出版社, 2015.
- [3] 冯爱萍, 王雪蕾, 吴传庆, 等. 北京市高碑店污水处理厂工况遥感监测[J]. 环境监测管理, 2017, 29(1): 39-42.
- [4] 陈界江, 邹云, 高仪斌, 等. UPLC-MS/MS法同时测定地表水中多种有机污染物[J]. 环境监测管理, 2020, 32(2): 52-55.
- [5] 曹秀芹, 吴蕙蓉. 图像分析在污水生物处理中的研究进展[J]. 环境科学与技术, 2009, 32(2): 60-64.
- [6] TOMPERIJ, KOIVURANTAE, KUOKKANENA, et al. Modelling effluent quality based on a real-time optical monitoring of the wastewater treatment process[J]. Environmental Technology, 2016, 38(1): 1-13.
- [7] AMARALA L, FERREIRA E C. Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression[J]. Analytica Chimica Acta, 2005, 544(1/2): 246-253.
- [8] MESQUITA D P, DIAS O, ELIAS R A V, et al. Dilution and magnification effects on image analysis applications in activated sludge characterization[J]. Microscopy and Microanalysis, 2010, 16(5): 561-568.
- [9] 王惠文, 吴载斌, 孟洁. 偏最小二乘回归的线性和非线性方法[M]. 北京: 国防工业出版社, 2006.
- [10] 张世强. 曲线回归的拟合优度指标的探讨[J]. 中国卫生统计, 2002(1): 9-11.
- [11] MESQUITA D P, DIAS O, DIAS M, et al. Correlation between sludge settling ability and image analysis information using partial least squares[J]. Analytica Chimica Acta, 2009, 642(1/2): 94-101.
- [12] 任东红. 基于集成神经网络的污水处理出水水质软测量模型[D]. 北京: 北京工业大学, 2013.
- [13] 丛秋梅. 活性污泥污水处理过程水质软测量方法的研究[D]. 沈阳: 东北大学, 2011.
- [14] LEAL C, VAL DEL RÍO A, MESQUITA D P, et al. Sludge volume index and suspended solids estimation of mature aerobic granular sludge by quantitative image analysis and chemometric tools[J]. Separation and Purification Technology, 2020, 234: 116049-116055.
- [15] WILÉN B M, LUMLEY D, NORDQVIST A. Dynamics in maximal settling capacity in an activated sludge treatment plant with highly loaded secondary settlers[J]. Water Science and Technology, 2004(7): 187-194.
- [16] 李振亮, 张代钧, 卢培利, 等. 活性污泥絮体粒径分布与分形维数的影响因素[J]. 环境科学, 2013, 34(10): 3975-3980.