



移动扫码阅读

陈运,廖振良.突发水污染事件的在线检测技术研究进展[J].能源环境保护,2019,33(6):1-5+31.

CHEN Yun, LIAO Zhenliang. Anomaly detection technology for online water quality monitoring: A Review[J]. Energy Environmental Protection, 2019, 33(6): 1-5+31.

突发水污染事件的在线检测技术研究进展

陈 运^{1,2}, 廖振良^{1,2}

(1. 同济大学 环境科学与工程学院, 上海 200092;

2. 上海市污染控制与生态安全研究所, 上海 200092)

摘要:为构建机构化的水质异常检测方法,对国内外水质异常检测技术进行了综述分析。根据研究方向的不同将水质异常检测技术分为机理模型和算法模型,着重探讨了基于算法模型的水质异常检测技术评价方法,对比了统计分析法、机器学习法和模态分析法的特点。对发展趋势进行了展望,认为复合化水质异常检测技术和新型在线监测设备开发是未来的重要研究方向。

关键词:在线监测;水质异常;检测技术;机理模型;算法模型

中图分类号:X830.7

文献标识码:A

文章编号:1006-8759(2019)06-0001-05

Anomaly detection technology for online water quality monitoring: A Review

CHEN Yun^{1,2}, LIAO Zhenliang^{1,2}

(1. College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China;

2. State Key Laboratory of Pollution Control and Resource Reuse, Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092, China)

Abstract: This review of domestic and international water quality anomaly detection methods attempts to construct a framework for institutionalized water quality anomaly detection methods. In this paper, the water quality anomaly detection technologies were divided into two types, the mechanism model and algorithm model, according to the research direction. The evaluation method of water quality anomaly detection technologies based on the algorithm model was discussed. The characteristics of statistical analysis method, machine learning method and modal analysis method were compared. At last, the development trend was prospected and it was considered that the future research will focus on combining mechanism-model and algorithm-model technologies and developing a new online monitoring device.

Key words: Online monitoring; Water quality anomaly; Detection technology; Mechanism model; Algorithm model

0 引言

水质在线监测系统避免了人工采样实验室分析过程的滞后性、复杂性^[1],而且能实时监测水质变化,是城市水体管理系统的重要组成部分。欧美、日本等国在上世纪六七十年代就设立了水质在线监测设备,如今水质监测已经完全实现自动化,并朝着移动式、便携式发展。国内的水质在线

监测系统虽然起步较晚,但自2006年后,我国水质在线监测系统迅猛发展,初步形成了市、省、国家的三级网络,现在我国水质在线监测系统共建成471个市的约1000个降水监测点位,978条河流和112座湖泊(水库)的1940个地表水水质评价、考核、排名断面,338个地级及以上城市的906个集中式饮用水水源监测断面,能实现水质实时监测,并按《地表水水质标准》(GB 3838-2002)进

收稿日期:2019-07-25

基金项目:国家重点研发计划政府间国际科技创新合作重点专项(2016YFE0123300);自然科学基金项目(No.51578396;51778451)

第一作者简介:陈运(1993-),男,湖北黄石人,硕士,主要研究方向为水质污染在线检测技术。E-mail:1832815@tongji.edu.cn

行水质分类^[2]。同时,快速发展的在线监测网络产生了海量的水质数据,从这些数据中及时准确地检测出水质异常,对污染物预警系统正确启动响应至关重要^[3]。水质异常检测技术作为污染预警系统的核心部分,一直受到关注,也是研究的重点。Jeffrey Yang 等^[4,5]通过实验模拟污染物在供水管网中的扩散过程,对杀虫剂、除草剂、生物碱、大肠杆菌、生物生长介质以及无机化合物等 11 种化学及生物污染物质进行管道流动实验,探索了一种实时事件自适应检测、识别和预警(READiw)方法;基于物质质量守恒与能量守恒原理的综合水质模型的异常检测研究比较深入,其中 Streeter-Phelps 模型^[6]是最早提出的模型,美国环保局(USEPA)随后推出 QUAL 系列模型^[7]、WASP 系列模型^[8]及 BASINS 系列模型^[9]等,然而这类研究方法更多的是基于水质的物理化学生物机理,随着时间空间的变化,水质模型难以适用。

然而,直接以监测数据为研究对象,利用数据统计、模态分析、人工智能以及机器学习等算法进行分析,可以避免上述问题。Byer 和 Carlsson^[10]将测量值与背景数据均值的差超过背景数据 3 倍标准差时,就认为有异常发生,实现了一种简单的水质异常检测方法;Klise K A 等^[3]研究比较了不同水质异常检测方法(事件序列递增、多维欧式距离、聚类方法)的检出率和误报率;Feng Shang 等^[11]基于流体动力学模型研究了在线水质监测指标的污染物事件实时检测方法,结合水体动力学,分析了污染物浓度与余氯之间的关联特性;方海泉等^[12]利用经验模态分解方法与中位数方法结合对异常值进行检测,再分段曲线拟合进行异常值校正,对于时间跨度大且变化趋势复杂的非线性非平稳水资源在线监测数据效果较明显。新型在线监测仪器的发展,如生物检测技术、重金属在线监测技术等,也为水质异常检测技术带来了新的思考。李志良^[13]通过实验得出鱼类的行为强度变化是农药的敏感指标,可利用在线监测设备实时监测鱼类的行为变化,达到监测及预警突发污染事故的目的;刘洁等^[14]采用鱼类作为水污染指示物,设计基于计算机视觉与支持向量机的水质监测方案,异常水质识别率可达 90% 以上。

水质异常检测方法涉及领域广泛,目前的综述性文章大多探讨的是饮用水供水系统的水质异常检测。S.Dejus 等^[15]对饮用水供水在线监测系统相关的生物污染检测技术进行了综述;Michael V.Storey 等^[16]探讨了饮用水在线水质监测及早期

预警系统的研究进展。因此,对整个水环境领域的水质异常检测技术的综述是必要的。

1 水质参数

1.1 水质参数的选择

按我国地表水水质标准^[17],所有的水质参数难以一一监测。在实际选择中,监测越多的水质参数,虽然能更全面更准确地检测出污染事件,但是会造成高昂的经济成本,而监测水质参数太少,又会导致污染事件的误判和漏检。因此,正确、准确地选择水质监测参数是非常重要的。关于选择合适、有用的水质指标已经进行了大量研究。世界卫生组织^[18]及欧盟法规^[19]建议监测游离氯、氧化还原电位(ORP)、硝酸盐浓度、温度、pH、电导率(EC)、溶解氧(DO)、浊度;John Hall 等^[20]针对饮用水投毒事件,通过实验分析某些常见污染物对常规水质指标的影响,得到铁氰化钾、草甘膦、马拉松、涕灭威、大肠杆菌、三氧化二砷和尼古丁等化学污染物与余氯、总氯、浊度、pH、ORP、EC、总有机碳(TOC)、氨氮和硝氮等水质指标的关系;较多研究提到,对各种类型污染物反应最好的参数是 TOC、游离氯、氯化物、ORP、pH、温度、EC 和浊度^[21,22];J.I.Pineiro Di Blasi 等^[23]选用浊度、EC、铵离子作为水质参数用于河流的水质异常检测,取得较好的效果;Lee^[24]及 S.Dejus^[15]等对水质指标的选择进行了全面的研究和总结。从文献中发现,针对饮用水突发污染事件,其水质监测参数大多选用 pH、EC、游离氯、浊度等常用指标,其中 pH 和游离氯是分析的重点;而对于河流、湖泊等水体,温度、浊度、pH、DO、EC、TOC、氨氮、总磷(TP)、叶绿素等是常规水质监测指标,其中 pH、DO、TOC、氨氮、总磷(TP)是分析的重点。

1.2 水质数据的特点

在线监测设备采集到的数据是进行水质异常检测的对象,了解水质数据的特点对异常检测尤为重要。水质在线监测系统利用现代数据传输技术将传感器监测的数据传送到基站,再以一定的时间间隔传输到监测总站^[25]。即使没有污染事件发生,水的 pH 值、温度、氯含量等参数仍可能在很大范围内变化,参数值的周期性变化可能是由于白天和晚上的需水量变化造成的,也可能是所谓的基线变化,例如使用另一个供水水库造成的^[15,28]。Yang 等^[4]提出传感器的输出值应该等于仪器对目标物的反应值、自然背景变化的影响值、仪器噪音、仪器漂移以及与操作相关的随机方

差之和。何慧梅^[26]及魏媛^[27]将水质数据的波动分为背景信息波动、工艺操作引起的变化、尖峰噪音以及污染事件。因此,初步形成了对水质数据分析的重点,即是从水质数据中分离出背景波动、仪器的尖峰噪音以及工艺操作等引起的数据变化,将污染事件引起的异常检测出来的。

2 水质异常检测技术

本文针对目前众多的水质异常检测技术,按研究方向分为机理模型和算法模型两大类。机理模型的异常检测技术指的是研究污染物在水体中发生的迁移、转化、沉淀、吸附等复杂的物理化学生物过程,运用大量实验数据来准确地构建水质模型,从历史经验中推演未来污染事件的发生情况。而算法模型利用各种算法从海量的数据中挖掘出潜在的、未知的、与历史经验不相符的异常检测方法已经成为了当下的研究热点,尤其是在数据规模庞大的网络流量监控、光谱图像处理以及金融数据分析领域,异常检测算法模型研究得到了广泛的应用。

2.1 机理模型的异常检测技术

根据不同的应用场景,主要分为管网综合水质模型和河流综合水质模型两个方向。

管网综合水质异常监测研究是对发生在供水管网中的污染物扩散过程进行模拟。John Hall^[20]等研究了包括草甘膦、涕灭威、尼古丁以及三氧化二砷、铁氰化钾等常见饮用水污染物对于氨氮、硝态氮、余氯、总氯、浊度、pH 值、电导率等常规水质监测指标的影响,并定量地给出了污染物浓度与水质指标变化幅度之间的对应关系。Jeffrey Yang 等^[4,5]则对除草剂、杀虫剂、大肠杆菌、生物培养基等常见饮用水污染物在输水管道中的扩散进行建模,研究污染物随管网水体流动传输的过程中对各个常规水质指标所产生影响的变化,并通过污染物质与各指标变化模式之间的对应关系实现了初步的污染物分类。John Cook 等^[29]利用实验室的模拟管网系统,对比分析了各常规水质参数对于不同污染物质的响应敏感程度,并结合实验数据给出余氯和总有机碳两个参数更适合作为污染检测指标的结论。魏媛等^[30]设计不同浓度的硫酸铜和铁氰化钾污染物的注入实验,有效地挖掘污染物注入前后 pH 值、浊度、余氯、溶解氧等 8 种水质指标间的内在联系,利用关联信息对因污染物引起的水质异常波动进行检测和识别,在保证检出率的情况下尽可能降低误报率。

河流综合水质模型的异常检测研究是在物质质量守恒与能量守恒原理的基础之上,通过结合流体力学的相关理论,构建出描述河道水体中污染物质随空间与时间迁移的河流水质模型,然后依据模型对污染物进入河流水体后的一系列物理、化学以及生物过程进行推演,从而对水质是否存在异常做出判断。在这一类研究方法中,Streeter-Phelps 模型^[6]是最早提出并且应用最广泛的模型,然而这种方法只能模拟水体中的溶解氧含量和生化需氧量的变化。美国环保局 (USEPA) 随后推出了一系列的水质模型体系,包括用于河流推演,对于河流河道形状以及污染排放模式有了更好适应性的 QUAL 系列模型^[7];用于模拟城市单一降雨事件 SWMM 模型^[31],可以跟踪模拟不同时间步长任意时刻每个子流域所产生径流的水质和水量,以及每个管道和河道中水的流量、水深及水质等情况;以及基于 GIS 环境,用于多目标环境分析的 BASINS 系列模型^[9]等。此外,丹麦水动力研究所开发的 MIKE 系列模型^[32]能够用于河网、河口以及滩涂等特殊地区的水质模拟,谢家强等^[33]基于 MIKE 模型用于中心城区的排水管网的预测评估。

机理模型的水质异常检测方法依托于污染物的理化性质以及水体动力学的基础,建立各种水质模型,在过去的水污染控制、环境规划以及水质管理中起到了重要作用。然而,水质模型的研究目前还存在模型复杂度过高、模型参数率定依赖经验假设等问题,这导致了计算结果的随机性较高,且准确性和适用性难以得到保障。

2.2 算法模型的异常检测技术

近几年来,随着传感器网络的快速普及和大数据处理技术的提升,在线监测识别算法在水环境领域逐渐成为研究的主流方向。突发水污染事件的在线监测识别算法模型研究主要包含统计分析、机器学习和模态分析三个大类。

2.2.1 基于统计分析的异常检测算法

统计分析型算法是假设正常的监测数据以一定的概率分布服从某种随机模型,再基于历史数据进行模型训练,用于预测未来一个时间点或多个时间点的状态,然后依据预测残差是否超出报警阈值来衡量当前时刻的水质是否正常。在众多的统计模型中,时间序列递增算法 (Time Series Increment, TSI) 与自回归模型 (Auto Regressive, AR) 的应用最为广泛。

TSI 是一种非常简单的预测算法,其基本思想

可描述为:用前一时刻经过标准化后的测量值代替当前时间点的预测值,与当前时刻标准化后的测量值比较,得出预测残差。而 AR 则是利用过去多个时刻观测值的线性组合来对当前时刻进行预测的,但是模型阶数的确立对 AR 模型影响较大。相比而言,AR 模型由于结合了更多最近时间的信息,可以更好地跟踪水质背景波动,得到的预测残差大部分时间内处于一个较小的水平;而 TSI 算法只考虑最近一个时间的信息,无法获得信号的历史特性,很多时候并不具有预测能力。总体而言,依据统计分析的在线检测方法对于平稳序列有着较好的检测效果,然而水质时间序列并非完全平稳,其整体基线水平存在着周期性、季节性的自然波动,基于统计分析的异常检测对于这种非平稳变化的捕捉效果并不理想。

2.2.2 基于机器学习的异常监测算法

基于机器学习方法的异常检测是指利用训练样本归纳得到的一般普适学习机对新样本进行异常判别的过程,机器学习相关算法的引入能够增强对水质波动情况的分析和识别能力,学习过程依据训练样本是否带有先验分类标签而分为有监督学习和无监督的学习。有监督学习的代表方法包括贝叶斯网络 (Bayesian Network, BN)、决策树 (Decision Tree, DT)、K 阶最近邻 (K-Nearest Neighbor, KNN)、人工神经网络 (Artificial Neural Networks, ANN)、支持向量机 (Support Vector Machine, SVM)、多元方法等。齐培培^[34]利用贝叶斯网络的方法,基于长江 2004 年的水质数据,构建贝叶斯网络预测模型,对 2005 年的水质进行了精准的评价和预测;Towler 等^[35]利用 K 阶最近邻算法对水质异常检测的效果进行了提升优化,首先构建一个包含了水质监测信息和地理位置信息的特征向量,然后利用 KNN 算法对各个相邻节点进行加权计算,得到融合了空间信息的水质异常概率;Wei Yang 等^[25]利用监测的 UV_{254} 和 pH 作为输入值,以 COD 作为输出值,建立 ANN 模型,经过迭代训练,预测 COD 值的效果较好;Modaresi 等^[36]研究比较了 SVM、概率神经网络 (Probabilistic Neural Networks, PNN) 和 KNN 三种水质异常检测方法的性能和效率,结果表明 SVM 在标定和验证阶段均无误差,性能最佳,KNN 算法效果最差,存在较高的误报率;C. Díaz Muñoz 等^[37]以 DO 和浊度为指标变量,把水质数据集看作是一个与时间有关的函数,运用功能深度概念的多元分析方法应用于圣埃斯特班河口水质监测

样品中异常值的检测,可以成功地评价河口和河流的水质。从文献中可知,SVM 算法的检测误报率最低,但模型参数的确定比较困难,使得模型训练的过程略为复杂;贝叶斯网络、KNN、ANN 等算法的训练相对简单,但是误报率相对较高。

无监督学习的典型代表方法为聚类算法。聚类算法的特点是不需要水质异常的先验知识,而是可以直接根据水质指标之间的相似性或距离对观测数据进行聚类,有 K-means、模糊聚类、非参数聚类、系统聚类等。Klise 等^[3]提出了一种利用 K-means 聚类与多元算法相结合的水质异常检测方法,将 4 个水质参数的观测值分别聚成 1 类、4 类和 8 类,然后分别计算当前时刻的检测值与各个聚类中心的最大欧氏距离,从而做出水质异常判断;何慧梅等^[38]提出基于 AR 模型和模糊 C-均值聚类 (FCM) 的多因子融合水质异常检测算法,即通过 AR 模型实现水质背景信号的高精度跟踪,采用 FCM 算法融合多种水质指标的 AR 预测残差,与设定阈值作比较并判断异常,结果表明该算法比传统的单因子法和利用多维欧氏距离融合 AR 预测残差的算法具有更高的异常检出率和较低的误报率;Vugrin 等^[39]为了更进一步降低异常检测的误报率,提出了一种基于轨迹聚类的异常检测方法,通过对水质监测数据的拟合多项式进行特征提取与聚类,来更好地区分工况操作与污染事件所引起的水质参数波动。可以发现,聚类算法主要是以两种方式进行水质异常检测:一是通过模型训练将历史数据模式归类,通过比较当前测量值与各个类别的相似度来判断当前的水质状态;二是用于预测残差归类,得到 N 个预测残差类别,判断当前水质是否正常。

2.2.3 基于模态分析的异常检测算法

上述提及的水质分析方法大多是在时域进行的异常诊断,对于水质指标幅值上的变化展开较为深入的挖掘。此外,Liu Y 等^[40]提出了时频域的水质异常检测方法,利用希尔伯特黄变换 (Hilbert-Huang Transform, HHT) 的方法将水质时间序列分解为一系列的固有模态函数 (Intrinsic-Mode Function, IMF),通过分析各个模态瞬时频率和瞬时幅值的变化来进行异常检测,该方法能够较好地抑制非平稳非线性的水质背景波动对检测结果的影响;方海泉等^[12]利用经验模态分解方法与中位数方法结合对异常值进行检测,再分段曲线拟合进行异常值校正,对于时间跨度大且变化趋势复杂的非线性非平稳水资源在线监测数据效

果比较明显。

3 结论与展望

本文对水质异常检测技术做了一个综述性回顾,分析了水质数据的特点和水质参数的选择,并将突发水污染事件的在线检测方法分为机理模型、算法模型,着重探讨了算法模型的水质异常检测技术,按算法类型又分为统计分析方法、机器学习方法以及模态分析方法,比较了各自的优缺点。

目前机理模型的在线检测技术因模型复杂度过高、模型参数率依赖经验假设等问题存在都会导致模型计算结果的随机性较高,准确性和适用性难以得到保障,算法型的在线检测技术又因检测结果无法进行合理解释而遭诟病,因此寻求两者结合的水质异常检测技术是未来研究的方向之一。其次现在的在线监测仪器大多只能监测常规水质指标,检测到的异常事件难以确定具体的污染物及来源,而新型在线监测设备的出现,如重金属在线监测、水质毒性在线监测、生物传感器等能直接监测水体的污染物质和状况,将是突发水污染事件的在线监测技术新的研究方向。

参考文献

- [1] Robert Janke, Regan Murray, James Uber, et al. Comparison of physical sampling and real-time monitoring strategies for designing a contamination warning system in a drinking water distribution system [J]. *Journal of Water Resources Planning and Management*, 2006, 132(4): 310-313.
- [2] 2018 年中国生态环境状况公报[R]. 2019: 61.
- [3] Katherine A. Klise, Sean A. McKenna. Water quality change detection: multivariate algorithms [J]. *Proceedings of Spie*, 2006: 10.
- [4] Yang Y. Jeffrey, Haught Roy C., Goodrich James A. Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results [J]. *Journal of Environmental Management*, 2009, 90(8): 2494-2506.
- [5] Y. Jeffrey Yanga, James A. Goodrich, Robert M. Clark, et al. Modeling and testing of reactive contaminant transport in drinking water pipes: chlorine response and implications for online contaminant detection [J]. *Water Research*, 2008, 42(6-7): 1397-1412.
- [6] Rinaldi S, Soncini-Sessa R, Stehfest H, et al. Modeling and control of river quality [J]. *Water Research*, 1979, 15(9): 1131.
- [7] Brown L C, Bamwell T O. The enhanced stream water quality models QUAL2E and QUAL2E - UNCAS: documentation and user manual [R]. US Environmental Protection Agency, 1987.
- [8] Di Toro D M, Fitzpatrick J J, Thomann R V. Water quality analysis simulation program (WASP) DRAFT: User's Manual (Version 6) [M]. U.S. Environmental Protection Agency, 2008: 268.
- [9] Antonio Marcomini, Glenn Walter Suter II, Andrea Critto, et al. Decision Support Systems for Risk-Based Management of Con-

- taminated Sites [J]. *Nato Security Through Science*, 2009: 267-273.
- [10] D. Byer, K. Carlson. Real-time detection of intentional chemical contamination in the distribution system [J]. *American Water Works Association*, 2005, 97(7): 130-133.
- [11] Feng Shang, James Uber, Regan Murray, et al. Model-based real-time detection of contamination events [J]. *Proceedings of the 10th Annual Water Distribution Systems Analysis Conference*, 2008: 1102-1106.
- [12] 方海泉, 薛惠锋, 蒋云钟, 等. 基于 EEMD 的水资源监测数据异常值检测与校正 [J]. *农业机械学报*, 2017, 48(9): 257-263.
- [13] 李志良. 鱼类行为学在水质在线监测与预警中的应用研究 [D]. 济南: 山东师范大学, 2008.
- [14] 刘洁, 程淑红, 王国兵. 基于计算机视觉的水质异常评价因子研究 [D]. 秦皇岛: 燕山大学, 2016.
- [15] S. Dejus, A. Nescerecka, S. Nazarovs, et al. Review of existing and emerging biological contamination detection tools for drinking water distribution systems (DWDS) online monitoring [Z]. 2015: 320-332.
- [16] Michael V. Storey, Bram van der Gaag, Brendan P. Burns. Advances in on-line drinking water quality monitoring and early warning systems [J]. *Water Research*, 2011, 45: 741-747.
- [17] 国家环境保护总局, 国家质量监督检验检疫总局. 地表水环境质量标准 [J]. 政策法规, 2002: 8-9.
- [18] European Communities. Drinking water regulations (No. 2) [S]. 2007: 37.
- [19] World Health Organization. Guidelines for drinking-water quality, fourth edition [S]. 2011: 564.
- [20] Hall J, Szabo J. Watersentinel online water quality monitoring as all indicator of drinking water contamination [R]. U.S. Environmental Protection Agency, 2005.
- [21] Hall J., Zaffiro A. D., Marx R. B., et al. On-line water quality parameters as indicators of distribution system contamination [J]. *American Water Works Association*, 2007, 99: 66-77.
- [22] Szabo J., Hall J. On-line water quality monitoring for drinking water contamination [J]. *Comprehensive Water Quality and Purification*, 2014: 266-282.
- [23] J. I. Pineiro Di Blasi, J. Martinez Torres, P. J. Garcia Nieto, et al. Analysis and detection of functional outliers in water quality parameters from different automated monitoring stations in the Nalon river basin (Northern Spain) [J]. *Environmental Science and Pollution Research*, 2015, 22(1): 387-396.
- [24] Adam Lee, Alex Francisque, Homayoun Najjaran, et al. Online monitoring of drinking water quality in a distribution network: a selection procedure for suitable water quality parameters and sensor devices [J]. *International Journal of System Assurance Engineering and Management*, 2012, 3(4): 323-337.
- [25] Wei Yang, Jun Nan, Dezhi Sun. An online water quality monitoring and management system developed for the Liming River basin in Daqing, China [J]. *Journal of Environmental Management*, 2008, 88(2): 318-325.
- [26] 何慧梅. 基于多传感器数据融合的水质异常检测方法研究 [D]. 杭州: 浙江大学, 2013.
- [27] 魏媛. 基于时间与空间关联分析的城市供水管网水质异常检测方法研究 [D]. 杭州: 浙江大学, 2016.
- [28] U. S. EPA. Water quality event detection systems for drinking water contamination warning systems: development, testing, and