

· 监管新论 ·

# 地表水环境质量大数据分析技术路线探讨

张仁泉

(苏州市环境监测中心, 江苏 苏州 215004)

**摘要:**阐述了地表水环境质量大数据分析的概念和意义,探讨了基于因素识别、数据选择、数据收集、数据整理、关联分析、模型验证和结果应用流程的地表水环境质量大数据分析技术路线,列举了4种典型的应用示例。提出,应建立数据共享机制,建立大数据分析工作程序,开发大数据分析应用软件,培养大数据分析技术人才。

**关键词:**地表水;环境质量;大数据分析;技术路线;保障措施

中图分类号:X820

文献标志码:B

文章编号:1674-6732(2016)06-0063-05

## Discussion on Technological Approach of Big Data Analysis of Environmental Quality of Surface Water

ZHANG Ren-quan

(Suzhou Environmental Monitoring Center, Suzhou, Jiangsu 215004, China)

**Abstract:** This work elaborates on the concept and importance of big data analysis of environmental quality of surface water, discusses its technological approach based on the procedure consisting of aspect identification, data selection, data collection, data arrangement, relation analysis, model verification and result application. In addition, this paper constructs four typical application examples. It is suggested to establish data sharing systems and big data analysis working procedures, to develop big data analysis application software and train professionals in big data analysis.

**Key words:** Surface water; Environmental quality; Big data analysis; Technological approach; Safeguard measures

随着环境监测网络的逐步完善,水环境、环境空气、声环境、土壤环境、辐射环境、生态环境等环境要素和各类污染源的监测覆盖面日益扩大,监测指标、监测频次、监测方法逐步标准化、规范化,特别是随着环境自动监测系统的建设和完善,产生的环境监测数据量高速增长。然而,目前这些海量环境监测数据被开发利用的程度不高,很多数据沉睡在报表、报告和档案里,没有体现它们应有的价值。环保部门对环境质量状况相关数据掌握不系统,各类数据碎片化,相互关联分析不够,难以说清环境质量状况及变化趋势,说清污染源排放状况、说清潜在的环境风险(简称“三个说清”)。在以改善环境质量为核心的环境保护新形势下,环境质量数据服务于生态环境综合决策科学化、生态环境监管精准化和生态环境公共服务便民化,是环境监测工作面临的巨大挑战。近年来,国家积极推进大数据发展战略,大数据时代应运而生,为提高环境质量数据开发利用水平和效率、实现“三个说清”带来了

难得的发展机遇。地表水环境质量大数据分析是生态环境大数据分析的一个重要组成部分,现对地表水环境质量大数据分析的技术路线进行了初步探讨。

### 1 生态环境大数据分析的背景

2015年8月,国务院印发了《促进大数据发展行动纲要》,将大数据提升为国家发展战略,指出“大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合,正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态<sup>[1]</sup>。”2015年7月,国务院办公厅印发了《生态环境监测网络建设方案》,强调“推进全国生态环境监测数据联网共

收稿日期:2016-08-02

作者简介:张仁泉(1963—),男,研究员级高工,硕士,从事环境监测工作。

享,开展环境监测大数据分析,实现生态环境监测和监管有效联动<sup>[2]</sup>。”2016年3月,环境保护部办公厅印发了《生态环境大数据建设总体方案》,指出“通过生态环境大数据发展和应用,推进环境管理转型,提升生态环境治理能力,为实现生态环境质量总体改善目标提供有力支撑<sup>[3]</sup>。”近年来大数据发展情况表明,它已广泛渗透到国民经济和社会生活的方方面面,通过大数据分析产生了良好的应用效益,成为势不可挡的历史潮流。在生态环境领域开展大数据分析是实现“三个说清”目标的重要手段和必然要求,它将推动生态环境保护工作进入一个崭新的信息化时代。

## 2 地表水环境质量大数据分析的概念和意义

地表水环境质量大数据作为生态环境大数据的一个重要组成部分,是表征地表水环境质量及与其相关的环境因素的种类、数量、质量、时空分布和变化规律的数字、文字、表格和图形等信息的总称。地表水环境质量大数据与其他领域大数据一样,具有“4V”特征,即海量的数据规模(Volume)、快速的数据流动(Velocity)、多样的数据类型(Variety)和巨大的数据价值(Value)<sup>[4]</sup>。

地表水环境质量数据来源广泛,包括环保、水利、水文、气象、农业、国土、统计等政府部门,企业、高等学校、科研院所和社会组织等机构,他们根据各自职能和专项工作需要,组织开展地表水环境质量相关监测和调查工作,获取的数据主要供本单位使用。地表水环境质量数据管理模式是各自为政,某类数据使用者主要依靠自身监测获得,这些数据的所有权、使用权都属于本单位,对其他使用者一般是保密的,不能无偿共享数据资源。这样,各类数据使用者由于自身业务领域的局限性、数据获取成本高等原因,仅能掌握自身监测获得的数据,有时通过协作方式获取少量其他机构产生的数据。与地表水环境质量大数据分析的需求相比较,各类数据使用者掌握的数据都是孤立的、片面的、碎片化的某一类别“小数据”,缺少与水环境质量相关的其他类别数据,只能从一个狭窄的角度了解地表水环境质量的局部情况,其应用价值较低。如果将与地表水环境质量相关的各类数据收集、整理、存储在一起,就形成了“大数据”。进行大数据分析,使各类数据相互关联、融合,就能揭示它们的内在规律、变化趋势及相互影响,通过数据聚合反应形

成新的具有更高价值的成果。地表水环境质量大数据分析可以有多种数据组合、关联模型和应用目的,为水污染控制和水环境保护提供科学依据。

## 3 地表水环境质量大数据分析技术路线

### 3.1 分析目的

地表水环境质量大数据分析采用崭新的思维和技术对与地表水环境质量有关的海量数据进行选择、采集、整理和关联分析,将各类数据进行有机融合,从中发现地表水环境状况的影响因素、内在规律和演变趋势,获得改善水环境质量的优化方案,科学支撑水环境保护决策。

### 3.2 分析流程

地表水环境质量大数据分析流程由因素识别、数据选择、数据收集、数据整理、关联分析、模型验证、结果应用环节构成,其相互关系见图1。图中的实线表示主流程,虚线表示反馈流程,即可能需要根据分析过程信息对前面相关环节进行调整、补充和修正。

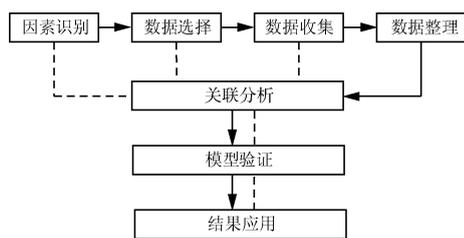


图1 地表水环境质量大数据分析流程

#### 3.2.1 因素识别

因素识别就是通过初步分析研究,明确影响地表水环境质量的主要因素类别,明确大数据收集的主要方向。地表水环境是一个开放的自然生态系统,其内部时时刻刻进行着物理、化学、生物过程,与其他环境要素之间也存在广泛的物质和能量交换。地表水环境系统复杂的内部因素和大量的外部因素影响共同决定了其质量状况。进行地表水环境质量大数据分析,首先需要将这些影响因素识别出来。例如:以某一湖泊水环境为例,影响其水环境质量的主要因素类别包括污染源、水文、气象,湖体物理、化学、生物过程,湖体水产养殖、水资源利用情况,周边社会经济、人口和产业布局状况,湖泊水污染防治和水生态保护措施等。

### 3.2.2 数据选择

影响地表水环境质量主要因素的状态需要由一系列指标来描述,反映这些指标的水平、大小和多少等特性的参数就是“数据”。根据大数据分析需要,在因素识别结果的基础上,根据已有知识和经验,选择反映地表水环境质量有关影响因素状况的数据(指标)。根据数据分析需要和现实可行性,设定数据的空间、时间、频次、准确度、精确度和监测方法等技术要求。数据选择应控制大小适度的范围和数量,太大就会导致大量无关的信息干扰分析,太小则不能为分析提供充分的原始素材。这一过程可以利用现有的反映地表水环境质量状况的物理、化学、生物过程演变模型,例如水污染物迁移、扩散、沉淀、氧化、还原、生物降解等模型,有针对性地选择相关性较强的数据。以某一湖泊水环境为例,通过数据选择确定:

(1) 该湖泊水环境质量数据:反映湖体的主要污染物种类及其浓度的时空变化等情况。

(2) 进入该湖泊的主要污染物数据:反映入湖污染源空间分布及其污染物产生和排放历时变化情况,包括固定污染源、移动污染源、点源、面源、干沉降、湿沉降和主要出入湖河流的污染物通量等信息。

(3) 影响该湖泊水环境质量的水文数据:反映主要出入湖河流及其流量,湖体形状、深度、湖水流场分布等。

(4) 影响该湖泊水环境质量的气象数据:反映温度、风速、太阳辐射强度和降水量等。

(5) 影响该湖泊水环境质量的物理、化学和生物过程数据:反映污染物迁移、扩散、沉淀、氧化、还原、底泥释放和水面逸出,水生动植物、藻类的种群分布及其生长过程与水体交换物质种类、数量和方式等。

(6) 影响该湖泊水环境质量的湖体开发利用数据:湖体水产养殖种类、规模和季节,工业、农业和生活用水情况等。

(7) 影响该湖泊水环境质量的社会经济数据:反映流域范围内产业结构、空间布局、产品产量、原辅材料使用量、用水量、生产工艺、人口数量和空间分布等信息。

(8) 与该湖泊水环境质量改善措施有关的数据:包括污染源控制、水文调控、水生态修复、区域产业结构调整和产业规模变化等信息。

### 3.2.3 数据收集

通过数据选择确定了需要收集数据的指标及相关技术要求,及时准确地收集这些数据才能为大数据分析提供必要的信息输入。地表水环境质量数据往往产生于相关政府部门、企业、高等学校、科研院所和社会组织等机构,其收集途径也多种多样,主要包括:

(1) 互联网信息系统:各类数据产生者建立和使用的互联网信息发布系统;

(2) 内部业务信息管理系统:各类数据产生者建立和使用的行政审批、监督管理、实验室检测等业务信息管理系统;

(3) 自动监测系统:各类数据产生者建立和使用的环境水质自动监测系统、污染源废水自动监测系统、水文自动监测系统和气象自动监测系统等;

(4) 专业领域数据中心:各类数据产生者建立和使用的环境、水利、水文、气象、农业和国土等专业领域数据中心;

(5) 手工收集的数据:对于上述规模化信息系统缺少的关键数据,需要通过人工方式收集予以补充,必要时进行专门监测、统计。

各类数据尽量采用自动接入方式,定期或实时收集,收集方式和频次取决于数据产生的方式、格式和数量等情况。收集到的数据要进行安全、有序存储,确保随时调用。

### 3.2.4 数据整理

从相关政府部门、企业、高等学校、科研院所和社会组织等机构收集到的各类数据,由于数据产生方式、数据格式和数据载体多种多样,必须进行整理,使其转换为统一的数字化结构,才能形成地表水环境质量“大数据”。地表水环境质量“大数据”不仅仅是简单的电子化的水环境资料,而是能被有效存储、查询检索、即时调用、方便管理和高效应用的数字化水环境资料<sup>[5]</sup>。通常收集的纸质报告、数据报表、图片资料,即使通过扫描转化为电子文档,也不能直接应用于“大数据”分析,只有将其转化为软件系统能够检索、统计和分析的数字化信息,才能作为“大数据”分析的原料。

对于地表水环境质量数据来说,数据整理还需要注意相关指标监测方法、监测仪器、监测点位、监测时间和监测质量控制等方面的技术差异,对不同来源的数据进行可比性分析,剔除错误的、不符合技术要求的数据,合理修正存在系统偏差的数据,

才能为“大数据”分析提供真实、准确的输入信息。

### 3.2.5 关联分析

通过数据整理获得的大数据仍然是碎片化的信息,需要对它们进行大数据关联分析。关联分析可以采用人工智能技术领域的技术方法,包括统计分析、数据挖掘、机器学习、自然语言处理、神经网络等技术<sup>[6]</sup>。将地表水环境质量大数据与水污染物迁移、扩散、沉淀、氧化、还原、生物降解等模型结合起来,可以使关联分析结果更加精准。通过对反映地表水环境质量某个方面现象的各类碎片化信息进行关联分析,建立各类数据之间的关联模型,就能把影响地表水环境质量主要因素的影响程度计算出来,实现大数据的预测和判断功能,从中发现趋势、找准问题、把握规律,推动各类水环境问题的有效解决,提高政府管理决策水平<sup>[7]</sup>。

### 3.2.6 模型验证

通过关联分析获得的地表水环境质量数据关联模型是否准确可靠,往往需要通过验证以判断影响因素识别、数据选择是否抓住重点,收集的数据是否系统准确,关联模型是否反映内在规律。可以选取若干组典型实际数据组合,采用所得关联模型进行计算。如果计算结果与实际情况偏差超过一定的限度,就要分析出现偏差的原因,据此反馈至因素识别、数据选择、数据收集、数据整理、关联分析等环节,进行必要的调整、补充和修正,完善关联模型,直至偏差符合设定的要求为止。例如,在模型中输入一组与地表水氨氮浓度有关的实际数据,将计算出的氨氮浓度与实测得到的实际氨氮浓度进行比较,如果偏差超过设定的 $\pm 10\%$ 限度,就需要采取改进模型结构、修正模型参数等方式完善关联模型,直至偏差满足控制要求,该模型才能投入应用。

### 3.2.7 结果应用

应用地表水环境质量大数据分析技术,可以帮助人们高效、准确、科学地发现并解决地表水环境问题,促进环境管理的转型和污染防治效率的提高。地表水环境质量大数据分析技术应用的典型示例如下:

(1) 某一湖泊水体总磷质量浓度为 $0.10\text{ mg/L}$ ,水质目标为达到地表水环境质量标准Ⅲ类,即总磷目标质量浓度为 $0.05\text{ mg/L}$ 。通过大数据分析可以明确:在污染源、水文、气象、湖泊水污染防治和水生态保护措施等诸多因素中,影响显著的因素

有哪些,有哪几种可供比选的达标方案,各因素调节措施如何组合形成最优达标方案。

(2) 某一湖泊水体现设有10个水质监测点,每月1次进行例行监测。通过大数据分析可以获得湖体总磷浓度的空间、时间连续分布数据,从而能够判断10个监测点某一次监测获得的总磷浓度算术平均值能否反映该湖泊总磷浓度总体平均值,据此可分析点位布设方案的合理性,并提出优化调整监测点位的方案;一年12次监测获得的总磷浓度算术平均值能否反映该湖泊全年总磷浓度总体平均值,据此可分析监测频次的合理性,并提出优化调整监测频次的方案。

(3) 某一湖泊水源地水质自动监测站测得的氟化物质量浓度突然从正常水平 $0.3\sim 0.4\text{ mg/L}$ 上升到 $1.2\text{ mg/L}$ 。通过大数据分析可以了解该水源地周边可能排放氟化物的环境风险源及其排放强度,结合其对水源地水质影响的程度,进行污染源解析和追踪,识别肇事污染源,并提出有效的应急措施建议方案。

(4) 某一湖泊水质在夏季(丰水期)可达标;在冬春季节(枯水期)东部湖区总磷超标0.5倍,西部湖区总磷超标1.0倍。通过大数据分析可以获得湖体总磷环境容量的时空特点,采取分季节、分区域的总量精细化管理,实施分区域动态容量总量控制<sup>[8]</sup>,实现湖体水质达标。

由此可见,由于各类相关数据组合方式、关联模型的多样性,地表水环境质量大数据分析可以具有多种应用目的,为地表水环境保护的定量决策和精细管理提供科学依据。由于影响地表水环境质量的因素多、情况复杂,大数据分析技术在应用时仍然存在较大的不确定性,需要根据应用效果不断修正、完善和改进。

## 4 地表水环境质量大数据分析保障措施

大数据分析作为新生事物,虽然发展前景良好,但目前在我国还处于起步阶段,特别是在地表水环境质量大数据分析方面的工作基础非常薄弱。要积极探索研究,做好顶层设计,扎实组织推进,采取相应的保障措施为地表水环境质量大数据分析创造良好的工作环境<sup>[9]</sup>。

### 4.1 建立数据共享机制

地表水环境质量大数据分析的基础是实现各类相关数据的共享。地表水环境质量数据来源广

泛,包括相关政府部门、企业、高等学校、科研院所和社会组织等机构。针对目前条块分割、部门封锁、各自为政的数据管理现状,必须通过立法方式建立数据共享机制,明确各相关方监测、发布和共享数据的种类、数量、频次等法定责任,打破条块职能分割,建立数据共享机制,让碎片化的数据充分融合起来,实现全社会共享。积极推进大数据标准规范建设,实现大数据技术体系标准化,消除业务技术壁垒,促进各类数据互联互通。加快建设大数据共享平台和专业数据中心,实现大数据规模化集中管理,提高大数据集成利用的水平和效率<sup>[10]</sup>。

#### 4.2 开发大数据分析应用软件

地表水环境质量大数据分析在数据采集、整理、分析和结果展示等方面的功能都是通过软件来实现的,而目前我国缺少适用的、成套的大数据分析软件,大数据分析技术难以落地生根。因此,必须大力推进地表水环境质量大数据分析软件开发,尽快开发规范化、智能化水平高,通用性、适用性强,维护使用方便,价格合理的成套软件,特别是要开发处于核心地位的关联分析软件,使大数据分析技术转化为地表水环境保护的实用工具。

#### 4.3 培养大数据分析技术人才

地表水环境质量大数据分析是一门崭新的综合性技术,涉及计算机科学与技术、信息技术、环境科学与技术、水文学和气象学等诸多学科领域。要推进地表水环境质量大数据分析工作,必须培养一

批具备相应学科知识和技术能力的复合型人才。在高等学校建立相关的学科、专业,开展大数据分析专业教育,为大数据分析输送对口的专业人才。在现有环境保护、水资源管理和信息技术人员中选拔专业基础好、业务能力强、相关工作经验丰富的人员,有针对性地开展大数据分析技术培训,尽快造就一批专业化、高素质的大数据分析人才队伍。

#### [参考文献]

- [1] 国务院. 关于印发促进大数据发展行动纲要的通知(国发[2015]50号)[Z]. 2015.
- [2] 国务院办公厅. 关于印发生态环境监测网络建设方案的通知(国办发[2015]56号)[Z]. 2015.
- [3] 环境保护部办公厅. 关于印发生态环境大数据建设总体方案的通知(环办厅[2016]23号)[Z]. 2016.
- [4] 詹志明,尹文君. 环保大数据及其在环境污染防治管理创新中的应用[J]. 环境保护,2016,44(6):44-48.
- [5] 黄波,周愿. 融入“大环境”中的水环境[J]. 水利水电快报,2015,36(7):7-10.
- [6] 常杪,冯雁,郭培坤,等. 环境大数据概念、特征及在环境管理中的应用[J]. 中国环境管理,2015,7(6):26-30.
- [7] 魏斌. 推进环境保护大数据应用和发展的建议[J]. 环境保护,2015,43(19):21-24.
- [8] 常杪,冯雁,解惠婷,等. 大数据驱动环境管理创新[J]. 环境保护,2015,43(19):25-29.
- [9] 解辉. 大数据技术在环境监测中的应用[J]. 环境监控与预警,2016,8(4):62-66.
- [10] 申剑,陈威. 美国地表水环境监测管理体系及对我国的启示[J]. 环境监控与预警,2016,8(4):58-61.
- [8] 环境保护部. 水质 硝基苯类化合物的测定 气相色谱-质谱法:HJ 716—2014 [S]. 北京:中国环境科学出版社,2014.
- [9] 环境保护部. 水质 硝基苯类化合物的测定 液液萃取/固相萃取-气相色谱法:HJ 648—2013 [S]. 北京:中国环境科学出版社,2013.
- [10] 环境保护部. 水质 硝基苯类化合物的测定 气相色谱法:HJ 592—2010 [S]. 北京:中国环境科学出版社,2011.
- [11] 中华人民共和国建设部. 还原-偶氮分光光度法测定城市污水中硝基苯类化合物:CJ T51—2004[S]. 北京:中国标准出版社,2004.
- [12] Nitroaromatics and Nitramines by High Performance Liquid Chromatography (HPLC):EPA 8330A—1998 [S].
- [13] 李利荣,魏恩棋,时庭锐,等. 气相色谱法测定水体中15种硝基苯类化合物的含量[J]. 理化检验(化学分册),2011,47(12):1436-1439.
- [14] 赵辰,王晋宇,陈玲珊,等. 液液萃取-气相色谱法测定地表水中10种硝基苯类化合物[J]. 理化检验(化学分册),2012,48(7):834-835,838.
- [15] 张蓓蓓,章勇,赵永刚. SPE-GC-ECD法测定水体中五种硝基苯类化合物[J]. 环境监控与预警,2010,2(2):18-20.
- [16] 吴伟,余昂,赵珊珊. 固相萃取-气质联用法测定固体废物浸出液中的二硝基苯[J]. 环境监控与预警,2015,7(5):28-30.
- [17] 陈明,阴永光,郜超,等. 顶空固相微萃取-气相色谱-质谱联用快速测定环境水样中的硝基苯、苯和苯胺[J]. 科学通报,2006,51(11):1359-1362.
- [18] 刘鹏,张兰英,焦雁林,等. 高效液相色谱法测定水中硝基苯和苯胺含量. 分析化学,2009,37(5):741-744.
- [19] 黄毅,饶竹,刘艳,等. 超高效液相色谱法直接快速测定环境水样中硝基苯和苯胺[J]. 岩矿测试,2012,31(4):666-671.

(上接第31页)