

生态环境大数据研究与应用进展

蒋洪强^{1*}, 卢亚灵^{1,3}, 周 思^{1,2}, 杨 勇¹

(1. 生态环境部环境规划院国家环境规划与政策模拟重点实验室, 北京 100012; 2. 荷兰瓦赫宁根大学及研究中心, 瓦赫宁根 6706KN; 3. 天津大学环境科学与工程学院, 天津 300354)

【摘 要】 生态环境大数据是为生态环保决策管理提供服务的大数据集、大数据技术和大数据应用的总称。生态环境大数据除了具有大数据的“6V”特征, 还具有高维、高复杂性、高不确定性的“三高”特性。本文从生态环境大数据的定义和特征出发, 阐述了国内外生态环境大数据的“萌芽—探索—应用—战略”的发展历程及其在科学研究、商业应用和政府决策等领域的主要应用, 总结了基于第四研究范式的生态信息学理论基础和依托于信息技术和大数据挖掘的技术方法, 分析了当前形势下我国生态环境大数据应用面临的三大挑战, 提出了未来生态环境大数据五大热点方向, 生态环境大数据必将在推进生态环境治理体系和治理能力现代化中发挥重要作用。

【关键词】 生态环境; 大数据; 环境管理; 研究进展

【中图分类号】 TP311.131; X3

【文献标识码】 A

【文章编号】 1674-6252 (2019) 06.0011.05

【DOI】 10.16868/j.cnki.1674-6252.2019.06.011

引言

随着现代信息技术的发展, 当今世界已经进入由数据主导的“大时代”。2012年5月, 联合国发布大数据政务白皮书“*Big Data for Development: Challenges & Opportunities*”, 标志着大数据领域的研究已提升为世界战略^[1]。2013年7月习近平总书记在中国科学院考察时指出, 大数据是工业社会的“自由”资源, 谁掌握了数据, 谁就掌握了主动权。2015年8月, 国务院发布《促进大数据发展行动纲要》, 大数据上升为我国国家战略。与此同时, 生态环境也进入大数据时代。习近平总书记指出, 要推进全国生态环境监测数据联网共享, 开展生态环境大数据分析。李克强总理强调, 要在环保等重点领域引入大数据监管, 主动查究违法违规行为。2016年3月, 原环境保护部办公厅印发《生态环境大数据建设总体方案》, 提出在未来五年内通过生态环境大数据建设和应用, 实现生态环境综合决策科学化、生态环境监管精准化、生态环境公共服务便民化。由于我国生态环境保护工作起步晚, 在数据知识更新、数据挖掘分析、大数据共享机制与能力建设等方面还存在不足。

1 生态环境大数据概念及特征

大数据在生态环境领域的应用、积累, 形成了生态环境相关的海量观测数据。这些数据来源于与生态环境相关的不同部门和领域, 来源多样、结构各异。一般认为, 生态环境大数据是为生态环保决策问题提供服务的大数据集、大数据技术和大数据应用的总称。生态环境大数据除

了具有大数据的6“V”特征, 即海量规模 (Volume)、形式种类繁多 (Variety)、处理速度快 (Velocity)、高价值性 (Value)、真实性 (Veracity)、易受攻击性 (Vulnerable), 还更加复杂多变, 具有高维、高复杂性、高不确定性的“三高”特性^[2]。

(1) 高维。数据来源包含反映自然和社会现象之间的多维数据^[3]。生态环境大数据来源于生态环境、气象、水利、国土、农业、林业、交通、社会经济等不同部门, 可以通过空气质量、水环境质量、噪声环境质量监测设备来感知, 还可以通过生物传感器、化学传感器、射频识别技术 (RFID)、卫星遥感、视频感知、光学传感器、人工检查等感知。通过全国“12369 环保举报联网管理平台”的电话、微信、微博等途径, 以及基于互联网技术产生的搜索数据、社交媒体数据、在线新闻等产生的生态环境舆情大数据, 包括半结构化和非结构化多维数据 (文本、项目报告、照片、影像、声音、视频等)。

(2) 高复杂性。生态环境大数据内在的复杂性 (包括类型的复杂、结构的复杂和模式的复杂) 使得生态环境数据的感知、表达、理解和计算等多个环节面临巨大的挑战。生态环境大数据本身价值较低, 只有通过大数据的数据清洗、数据集成、数据建模、结果导出与可视化等, 才能将这种不完善、复杂、非结构化的数据转化为有用的信息。生态环境领域的排放清单建立、环境质量模拟、最优化减排方案制定等, 如果借助大数据预处理技术、数据挖掘技术和云计算平台等, 速度和准确性会显著提高。

(3) 高不确定性。数据采集涉及高度不确定性, 数据

资助项目: 国家自然科学基金重大研究计划“大数据驱动的公共管理决策创新模式与集成示范平台”(91846301), 生态环境部环境规划院青年基金项目“基于大数据的大气污染高分辨率预测方法及研究”(40050701)。

作者简介: 蒋洪强 (1975—), 男, 博士, 研究员, 研究方向为环境规划与生态环境大数据治理, E-mail:jianghq@caep.org.cn。

可能存在错误或不完整。生态环境相关的大数据跟其他大数据一样,最典型的特征是数据量巨大,而且在以每年数百TB的速度增加。由于这些数据来源于不同部门,部门之间缺乏统一的标准规范;而且通过传感器、智能手机或社交网络等不同感知工具采集,即使来源于同一部门的数据格式也多样化;加之目前各部门数据共享程度较低,同一指标数据存在不一致性。

2 发展历程及应用

2.1 发展历程

(1) 1980年及以前的萌芽阶段。1980年,未来学家托夫勒在《第三次浪潮》中首次提出“大数据”的概念,大数据时代正式到来。在此之前,“大数据”已经在萌芽状态。以生态环境大数据为例,国际地球物理年(IGY)(1957—1958年)和国际生物学计划(IBP)(1964—1974年)是现今生态环境大数据研究的雏形,当时被称为“大科学研究”^[4],目的是获得较为可靠的大量观测数据,以研究地球各圈层和生态环境问题。这些研究最后演变成如今以长期定位观测为基础的生态系统研究网络,从而全面获取有关生态环境的观测数据。

(2) 1980—2008年的探索研究阶段。2008年,《Nature》、《Science》等学术刊物相继出版专刊探讨大数据议题,标志着大数据研究得到世界范围内的关注和认可。在此之前,各国学者对大数据及大数据在生态环境领域的研究进行了探索。如Camara等^[5]开发了集成水质数据库和污染数据库,以及面源污染模型和污水处理优化模型,用于西欧TeJo海湾水质管理的决策支持系统“Hypetejo”,利用它解决污水处理厂的选址问题、污染负荷改变对海湾水环境的评价等决策问题。

(3) 2009—2012年的应用研究阶段。在此期间,大数据问题得到越来越广泛的关注,一些商业公司也加入其中。麦肯锡公司发布关于“大数据”的报告,成为“大数据”较早的应用。生态环境领域的研究热点包括区域大气污染防治与污染物减排研究^[6]、环境污染治理及预报^[7]、全球气候变化预测研究等^[8]。由清华大学、华南理工大学、田纳西大学等联合开发的大气污染控制费效评估及空气质量达标规划辅助决策系统“ABaCAS”,集成大气污染控制费效及达标评估、大气污染控制成本分析、空气质量模拟可视化分析、空气质量达标评估等技术,是这一时期的应用研究之一。

(4) 2012年以来的战略化发展阶段。2012年以来,大数据应用问题得到各国政府的高度重视。2012年联合国大数据政务白皮书的发布,标志着大数据领域的研究已提升为世界战略。同年,美国政府启动《大数据研究和发展计划》,标志着美国大数据时代来临。2015年国务院发布《促进大数据发展行动纲要》,大数据发展上升为我国国家战略。2016年原环境保护部办公厅印发《生态环境大数据建设总体方案》,从政策层面对大数据应用于环境管理领域提出了要求。当前生态环境大数据在科学研究、商业应用、

政府决策支持等多个方面得到广泛应用。

2.2 研究与应用

(1) 科学研究方面。数据感知层面,如上所述的“大科学研究”催生了生态系统研究网络的发展。目前国际上已经建立多套全球性和国家/区域性的生态环境监测网络,提供包括环境和生态变量的长期多维观测数据。全球尺度的生态环境观测网络主要包括国际长期生态研究计划(ILTER)、全球环境监测系统(GEMS)、国际生物多样性观测网络(GEO·BON)等。国家尺度的网络包括美国的US-LTER生态环境观测研究网络、英国的ECN生态环境观测研究网络和日本长期生态研究网络(JALTER)^[4,9]。这些观测网络采集了生态环境相关的大数据,能够获得数百万次观测结果^[10]。卫星或飞机的遥感技术的发展使我们通过安装在这些平台上的远程传感器获取大量生态数据。例如,美国宇航局(NASA)部署高光谱传感器以捕获有关叶面营养、树冠结构和性状的详细信息^[11]。2015年6月,NASA联合巨型计算机技术、地球系统模型、工作流管理和遥感数据协作分析平台,发表了从1950年到2100年全世界的气候变化预测数据。中国不同部门和单位陆续建立了一批生态环境监测站,开展了不同区域的环境、资源、污染的调查与研究工作。如中国科学院植物研究所建立了植物介绍网站(iPlant),其数据集形式多样,包括观测数据、实验数据、模拟数据以及其他派生数据^[12]。

研究层面,近年来美国国家科学基金会(National Science Foundation, United States, NSF)投入大量人力、物力研究大数据科学计算应用,目前大数据技术被广泛应用于生态系统研究的各领域。NSF最近公布了BIG DATA方案征集以利于信息共享。一些学科已经开发了海量数据平台并取得了相应的收益。近年来,我国自然科学基金委管理学部也资助了一批大数据在金融、医疗、交通及生态环境领域的管理应用研究。Li Lei^[13]等收集了2005年至2013年我国31个省(区、市)的数据,使用数据包络分析方法研究了经济、社会和生态因素对于我国林业资源效率的影响。He Fuhong^[14]等以沿海矿业城市龙口为例,使用遥感图像、数字高程模型以及降水等数据,建立了包括自然地理、地质条件、采矿强度、生态环境恢复等在内的综合评价指标体系,基于插值等对大数据方法进行了生态环境质量评估。

(2) 商业应用方面。大数据正在从科学研究层面走向商业应用。德克萨斯州立大学和戴尔公司联合研发的巨型计算机Stampede,性能良好且应用广泛,美国南加州地震中心利用该巨型计算机进行加州破坏性地震的频率预测;德州大学奥斯汀分校也应用Stampede通过详细的数据建模,分析从南极洲到海洋的冰川流动^[15]。在生态环境大数据的建设方面,惠普、谷歌、微软等美国企业正在提供最先进的存储设备和搜索等服务,帮助政府和研究机构对环境现状及未来趋势作出判断。

2012年,惠普公司联合环保组织——“保护国际”(Conservation International)启动惠普地球观察(HP Earth

Insights)项目,对全球生物多样性和气候数据进行了系统分析。在中国,生态环境大数据得以进入商业应用的基础是Apache基金会开发的Hadoop平台。在Hadoop平台基础上,IBM公司与北京市政府联合开发了“绿色地平线”大数据平台系统,该系统结合当时的气象卫星和地面监测数据,结合企业排放数据预测未来72小时的空气质量。微软在中国也已与海南、云南、武汉等多个地区进行交通、能源、环境等领域的合作,已经发布了超过100个全球智慧城市案例。针对雾霾问题,微软还推出了Urban Air系统,通过大数据来监测和预报空气质量,该服务覆盖了中国300多个城市^[16]。而京东智能城市研究院将利用大数据和人工智能进行空气质量和水质预测作为重点研究的业务内容。

(3)政府决策方面。世界各国都将大数据技术研究应用的推进作为重大发展战略。2012年3月,美国公布了“大数据研发计划”,以提高从海量和复杂的数据中获取知识的能力。根据该计划,美国国家科学基金会(NSF)、能源部(DOE)、地质勘探局(USGS)等6个联邦部门和机构共同提高了海量数据应用所需的核心技术。美国环境保护署(EPA)建立了统一的中央数据交换系统,形成了排污设施登记数据库。欧盟方面,过去几年已对科学数据基础设施投资了1亿多欧元,并将数据信息化基础设施作为Horizon 2020计划的优先领域之一。英国自然环境研究理事会(NERC)投资超过1300万英镑,计划2020年建立环境数据创新中心^[17]。亚洲地区,2014年新加坡政府提出了“智慧国家平台”(Smart Nation Platform),这是全球第一个全国范围性质的智慧蓝图。在大数据技术的支持下,新加坡联合其他受影响的国家开发了东南亚国家区域烟霾预警系统(AHMS),为防治重污染天气提供即时的政策建议^[18]。

我国已将生态环境大数据列为国家发展战略中的重要一环。生态环境部成立了生态环境大数据建设领导小组,全面推动落实党中央、国务院关于大数据发展的新要求,同时启动了生态环境大数据和环保云建设项目。目前,生态环境部建立了涵盖大气、水和土壤等领域的生态环境监测网络系统并逐步加大监测密度,其中大气环境监测网络包括1436个城市监测点位,水环境监测网包括1000个降水监测点位、1940个地表水水质断面(点位),906个集中式饮用水水源监测断面(点位),1649个海水环境质量国控监测点位,还包括2583个生态环境质量监测点位,约80000个城市声环境监测点位,1410个环境电离辐射监测点位和44个环境电磁辐射监测点位^[19]。当前的环境监管执法,使用视频监控、无人机的红外相机等设备,借助数据实时高速传输技术,建设了有效的监管系统。福建、内蒙古、山东等地启动了生态环境大数据建设工程。

3 理论基础及技术方法

3.1 理论基础

科学研究的前两个范式是实验和理论,计算和模拟为科学研究的第三范式,数据密集型科学是前三种科学方法

之后的第四种科学类型,被称为第四范式^[20]。第四范式中科学家已不根据已知的规则编程,而是以数据为中心进行信息挖掘。前三种范式下的传统的生态环境研究在获取、管理和分析大量数据方面带来了挑战。新的生态环境监测网络系统及其大数据的应用正在慢慢将环境生态学引入大科学研究领域,进而催生了生态信息学,奠定了生态环境大数据的理论基础。生态信息科学是一门以利用现代计算技术(如人工神经网络、遗传算法等)进行生态学分析、综合和预测为重点,综合研究生态信息提取、生态信息管理和生态信息分析的科学^[21]。生态信息学为数据密集的生态环境研究提供了新的统一理论、模拟和分析的方法,使科学家能够通过创新工具和方法产生新知识,同时管理和应用环境与社会经济数据。

3.2 技术方法

生态环境大数据的应用需要突破多项技术难点,包括智能感知、分布式集群、云计算、机器学习、专业化模型等相关的工具、技术和方法在生态环境领域的应用。例如,使用Hadoop的分布式文件系统(HDFS)和分布式数据库(MapReduce)对生态环境大数据进行批量处理;利用决策树、贝叶斯、回归模型、关联规则、认知算法、机器学习等各种模型和算法对海量数据进行关联分析和深度挖掘,通过各种数据产生有价值的信息^[22]。

建立“天地空一体化”的数据感知系统,是生态环境大数据建设的前提。以现有的生态监测网络(空气和水环境质量监测)为基础,引入物联网、卫星遥感、低空航测(视频监控、无人机的红外摄影)等技术,构建“天地一体化”遥感监测、视频监控等智能监测体系。

建立多源异构大数据集成与存储系统,是生态环境大数据建设的基础。以生态空间数据与生态环境业务数据为中心,针对非结构化大数据的多样性及结构化数据的异构异源特性,实现多源数据空间和时间融合,解决生态环境大数据的高效存储与清洗问题。

建立多维时空生态环境大数据治理体系,是生态环境大数据应用的核心。建立集实时监控数据形势诊断、预警预报和会商决策等于一体的生态环境预警会商体系,建立“督察、执法、应急指挥”一体化的督察执法监管系统,构建基于“一张图”的全景指挥平台,是当前生态环境大数据的主要应用场景。

3.3 研究的关键环节

有效开展生态环境大数据研究应用需要解决以下关键问题:

(1)确定研究的主体和目标。由于大数据本身的复杂性以及数据的海量性,生态环境大数据可研究的范围非常宽泛,可以是最原始的元数据,可以是宏观的理论概念,还可以是多学科交叉的过程,因此必须确定要研究的主体和目标。

(2)明确研究的时空尺度。生态环境大数据的不同研

究领域有不同的尺度。例如,生态学家以年甚至万年为单位,关注物种在不同时间空间的演变;而气象学家则以分钟、小时、天来衡量天气变化情况。不同尺度使得本就复杂的数据更冗余。所以,研究的尺度应该在一开始就规定好,在简化工作量的同时加强大数据分析结果的应用。

(3) 建立科学的标准规范。目前的生态环境大数据,国内外普遍缺少统一的数据采集与存储标准规范体系,妨碍了数据的交换与共享。因此要构建科学、完善的生态环境大数据标准规范体系,为我国生态环境大数据的科学发展提供标准化支撑,为推进生态环境领域的发展提供规范依据。

4 面临的主要挑战

(1) 数据获取的局限性。目前主要有四种途径获取生态环境大数据:①整理现有数据集。这种方法成本昂贵,最后得出的数据集可能具有地理或时间差距。②整合遥感平台数据。此类数据在时空尺度上通常没有太大差距,但是这种途径仍受测量手段和变量的制约。③获取观测站或实地实验。该途径通常需要复杂且昂贵的仪器。④从头启动跨越大陆尺度的观测计划。该途径需要一开始就设计专有模型,并使用标准化方法^[23]。现阶段,随着生态环境监管需求导向明显,仅依托现有监测网络和设备还难以真正地实现“用数据说话、用数据管理、用数据决策”的目标,无法为问题分析预警、督察执法、全景指挥提供有力支撑,成为制约生态环境部门进一步提升大数据智慧监管能力的主要问题。此外,从社交媒体获取的生态环境相关大数据从商业用途转用于科学用途时也导致了信息不完整等问题,限制了这些数据的进一步应用。

(2) 数据共享的局限性。生态环境数据资源分散,资源整合利用程度不高。不同部门之间不愿意共享数据是管理上的瓶颈。同时,生态环境是典型的跨行业多类型的大数据综合应用业务领域,不同来源的生态环境数据的标准不规范,导致数据对话成本很高,数据质量难有说服力。尤其是商业大数据供应商,往往过于追求商业利润,在收集数据时通常不采用科学的抽样方法,基于这些大数据的研究得出的结论可能与研究中声称的目标不匹配^[1]。

(3) 数据应用的局限性。部分研究人员倾向于通盘接受大数据,没有对数据的真实性、可靠性进行审查和深入分析,容易出现系统偏差,对决策管理形成负面影响^[24]。大数据技术在环境与经济综合分析、环境污染监督执法与应急、生态环境评估、生态系统管理及全球气候变化预测等方面已经得到了初步的应用,但目前并没有形成成熟的生态环境大数据产业链,对生态环境管理的精准化决策支撑还需加强。我国还缺乏生态环境大数据和环境管理兼备的复合型人才,现有环境管理人员知识结构需要更新。

5 热点发展方向

(1) 生态环境大数据共享融合研究。没有任何一个人

或机构可以同时容纳和有效分析所有形式的生态环境数据。要使生态环境大数据得到应用与发展,需要加大共享融合研究,优先考虑数据、方法、标准和代码的开放性,包括更快地采用新技术和工具方法,开放数据来源和共享解决方案,构建网络基础架构,改进数据和共享工作流程,以及增强大数据服务的翻译和记录。

(2) 开放架构的生态环境云平台建设。构建生态环境大数据平台就是要将多源异构的数据进行有机地整合^[25]。一是加大感知系统建设。要以改善生态环境质量为核心,以现有的生态监测网络为基础,扩大监测范围,增加监测密度,引入物联网、卫星遥感、低空航测等技术,实现重点领域自动监测全覆盖,构建“天地一体化”遥感监测、视频监控等智能监测、监控体系,能够及时发现流域区域内发生违法行为,为生态环境规划、评估、管理和监管提供进一步的有效支持。二是建立完善大数据治理体系。加强对数据资源整体规划,形成“标准统一、动态更新、共享应用”的信息资源目录和平台,提升大数据资源规划、生态环境云平台建设、主题数据库建设、大数据资源采集、大数据资源治理、大数据安全管理、大数据资源服务等数据的治理能力。

(3) 基于大数据的精细化环境管理决策应用。①生态环境与经济社会大数据形势分析,建立生态环境与经济社会形势分析关联指数,将环境变化特征与经济社会发展情况进行关联对比,识别影响环境质量变化的主要原因。②生态环境质量监测预警分析,通过对环境质量现状和变化趋势进行综合统计分析,全面掌握环境质量变化情况,建立监测预警标准和监测预警系统,形成生态环境质量监测预警体系。③生态环境事件举报与舆情监控分析,包括政府提供电子公共服务平台、互联网服务平台。通过网络舆情采集分析技术和公众环保移动应用,让舆情监控由被动变主动。④生态环保督察执法智慧决策支撑。通过大数据感知等手段完善督察执法管理手段,全面提高督察执法工作效能,并将其应用于企业关改搬转、黑臭水体整治、入河/海排污口排查、农业面源污染整治、非法码头专项整治、河道非法采砂整治等领域。

(4) 基于大数据多样性的科学结论验证。大数据作为新的信息技术,在生态环境领域的未来研究中,应进一步理解和评估大数据的质量,并辅以往传统的科学数据收集方法,以获得更详细和有代表性的数据^[26]。另外,通过科学的传统抽样调查,可以收集更多目标的详细信息,并使收集的数据更具代表性。还可以使用另一个平台提供的大数据来测试从一种类型的大数据得出的结论的稳健性。

(5) 国际合作下的生态环境大数据应用。建立国际资源环境数据知识共享平台(全球资源环境数据库与知识库建设),实现我国与其他国家在水资源、气候变化、能源清洁利用、环境保护、环境金融等方面的数据和知识共享,加强我国与其他国家在环境数据领域的合作与交流,共同应对全球性资源环境问题。

总之,生态环境大数据必将在我国推进环境治理体系和环境治理能力现代化中发挥重要作用,应切实提高我国生态环境大数据建设水平。

参考文献

- [1] UN Global Pulse. Big Data for Development: Challenges & Opportunities[R]. New York:Global Pulse, 2012.
- [2] GUO H D, WANG L Z, CHEN F, et al. Scientific big data and digital earth[J]. Chinese science bulletin, 2014, 59 (35) :5066-5073.
- [3] GUO H D. Space-based observation for sensitive factors of global change[J]. Bulletin of the Chinese academy of sciences, 2009, 23 (4) : 226-228.
- [4] 赵海凤, 李仁强, 赵芬, 等. 生态环境大数据发展现状与趋势 [J]. 生态科学, 2018, 37 (1) : 211-218.
- [5] CAMARA A S, DA SILVA M C, CARMONA RODRIGUES A, et al. Decision support system for estuarine water-quality management[J]. Journal of water resources planning and management, 1990, 116 (3) : 417-432.
- [6] 吴刚. 资源环境领域大数据的研究与资助分析 [J]. 中国环境管理, 2015, 7 (6) : 14-18.
- [7] 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.
- [8] 赵国栋. 大数据时代的历史机遇: 产业变革与数据科学 [J]. 中国科技信息, 2013 (19) : 101-101.
- [9] LINDENMAYER D B, LIKENS G E, ANDERSEN A, et al. Value of long - term ecological studies[J]. Austral ecology, 2012, 37 (7) : 745-757.
- [10] DENGLER J, JANSEN F, GLÖCKLER F, et al. The Global Index of Vegetation - Plot Databases (GIVD) : a new resource for vegetation science[J]. Journal of vegetation science, 2011, 22 (4) : 582-597.
- [11] ASNER G P, MARTIN R E. Spectranomics: emerging science and conservation opportunities at the interface of biodiversity and remote sensing[J]. Global ecology and conservation, 2016, 8: 212-219.
- [12] MERCHANT N, LYONS E, GOFF S, et al. The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences[J]. PLoS biology, 2016, 14 (1) : e1002342.
- [13] XIE H L, HE Y F, XIE X. Exploring the factors influencing ecological land change for China's Beijing-Tianjin-Hebei Region using big data[J]. Journal of cleaner production, 2017, 142: 677-687.
- [14] HE F H, GU L J, WANG T, et al. The synthetic geo-ecological environmental evaluation of a coastal coal-mining city using spatiotemporal big data: a case study in Longkou, China[J]. Journal of cleaner production, 2017, 142: 854-866.
- [15] 郭华东, 王力哲, 陈方, 等. 科学大数据与数字地球 [J]. 科学通报, 2014, 59 (12) : 1047-1054.
- [16] ZHENG Y, CAPRA L, WOLFSON O, et al. Urban computing: concepts, methodologies, and applications[J]. ACM transactions on intelligent systems and technology, 2014, 5 (3) : 38.
- [17] 汪自书, 胡迪. 我国环境管理新进展及环境大数据技术应用展望 [J]. 中国环境管理, 2018, 10 (5) : 90-96.
- [18] 胡昊, 曹媛颖, 张啸天, 等. 生态环境大数据探究 [C]//2018 中国环境科学学会科学技术年会议论文集 (第四卷). 合肥: 中国环境科学学会, 2018: 3264-3269.
- [19] 中华人民共和国生态环境部. 2018 中国生态环境状况公报 [EB/OL]. (2019-05-29). http://www.mee.gov.cn/home/jrtt_1/201905/W020190529619750576186.pdf.
- [20] 陈明. 数据密集型科研第四范式 [J]. 计算机教育, 2013 (9) : 103-106.
- [21] 周才平, 何洪林, 于贵瑞. 生态信息科学的理论框架初探 [J]. 资源科学, 2002, 24 (1) : 77-81.
- [22] 李学龙, 龚海刚. 大数据系统综述 [J]. 中国科学: 信息科学, 2015, 45 (1) : 1-44.
- [23] SHIN D H, CHOI M J. Ecological views of big data: Perspectives and issues[J]. Telematics and informatics, 2015, 32 (2) : 311-320.
- [24] GANDOMI A, HAIDER M. Beyond the hype: big data concepts, methods, and analytics[J]. International journal of information management, 2015, 35 (2) : 137-144.
- [25] WILKINSON M D, DUMONTIER M, AALBERSBERG IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship[J]. Scientific data, 2016, 3: 160018.
- [26] VITOLO C, ELKHATIB Y, REUSSER D, et al. Web technologies for environmental big data[J]. Environmental modelling & software, 2015, 63: 185-198.

Progress in Research and Application of Ecological Environment Big Data

JIANG Hongqiang^{1*}, LU Yaling^{1,3}, ZHOU Si^{1,2}, YANG Yong¹

(1. Key Laboratory of Environmental Planning and Policy Simulation, Chinese Academy of Environmental Planning, Ministry of Ecology and Environment, Beijing 100012, China; 2. Wageningen University and Research, Wageningen, 6706KN, Holland; 3. School of Environmental Science and Engineering, Tianjin University, Tianjin 300354, China)

Abstract: Ecological environment big data is the general term for big data sets, big data technologies and big data applications, which provide services for eco-environmental decision management. In addition to the 6 “V” feature of big data, the ecological environment big data also has the “three high” characteristics of high dimensionality, high complexity and high uncertainty. Based on the definition and characteristics of big data in ecological environment, this paper expounded the development, exploration research, application research, strategic development and application of ecological data big data research in China and abroad. Furthermore, the paper explains the big data application in the scientific research, business applications and government decision-making procedures. It summed up the theoretical basis of ecological informatics based on the fourth research paradigm and the technical methods based on information technology and data mining, and then analyzed the three challenges faced by China's ecological environment big data application under the current situation. The five hotspots research and application directions were put forward in the end. The ecological environment big data will play an important role in promoting the environmental governance system and the modernization of environmental governance capabilities.

Keywords: ecological environment; big data; environment management; research progress