

•综述•

大数据时代的生物多样性科学与宏生态学

张 健*

(华东师范大学生态与环境科学学院, 上海 200241)

摘要: 高质量的生物多样性数据是认知生物多样性的起源和维持机制及应对其丧失风险的科学基础。当前, 在新物种发现、已知物种的地理分布、种群数量与时空动态、物种进化史、功能性状、物种与环境之间以及物种与物种之间的相互作用等7个方面都存在着知识上的空缺。大数据时代的到来为弥补这些知识空缺提供了可能, 大数据的挖掘及其应用最近已成为国际生物多样性与宏生态学研究的前沿内容。如何有效地利用和分析不断增长的生物多样性大数据是生物多样性研究面临的一个极大挑战。本文通过全球、大陆和区域尺度上的研究案例展示了大数据在生物多样性研究中的应用的新进展, 内容涉及森林覆盖变化、保护生态学、生物多样性与生态系统功能、气候变化对生物多样性的影响等。最后, 对大数据在生物多样性研究中存在的数据采集、处理和分析等方面的问题进行了总结, 并对其潜在应用前景进行了探讨。

关键词: 大数据科学; 保护生物学; 生物多样性信息学; 宏系统生态学; 公众科学

Biodiversity science and macroecology in the era of big data

Jian Zhang*

School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241

Abstract: High-quality biodiversity data are the scientific basis for understanding the origin and maintenance of biodiversity and dealing with its extinction risk. Currently, we identify at least seven knowledge shortfalls or gaps in biodiversity science, including the lack of knowledge on species descriptions, species geographic distributions, species abundance and population dynamics, evolutionary history, functional traits, interactions between species and the abiotic environment, and biotic interactions. The arrival of the current era of big data offers a potential solution to address these shortfalls. Big data mining and its applications have recently become the frontier of biodiversity science and macroecology. It is a challenge for ecologists to utilize and effectively analyze the ever-growing quantity of biodiversity data. In this paper, I review several biodiversity-related studies over global, continental, and regional scales, and demonstrate how big data approaches are used to address biodiversity questions. These examples include forest cover changes, conservation ecology, biodiversity and ecosystem functioning, and the effect of climate change on biodiversity. Furthermore, I summarize the current challenges facing biodiversity data collection, data processing and data analysis, and discuss potential applications of big data approaches in the fields of biodiversity science and macroecology.

Key words: big data science; conservation biology; biodiversity informatics; macrosystems ecology; citizen science

进入21世纪以来, 随着信息和网络技术的飞速发展, 全球数据呈爆发式增长, 大数据时代也随之而来(Mayer-Schönberger & Cukier, 2013)。与传统数据相比, 大数据具有数据量大(Volume)、类型多(Variety)、处理时效快(Velocity)、数据价值高(Value)

这四“V”特征, 国际数据公司(International Data Corporation, IDC)也以此来定义大数据(Gantz & Reinsel, 2012)。目前, 大数据正影响着全球人类的生活、科技、经济、文化和政治发展等诸多方面, 也在逐渐改变着我们的思维方式。2007年, 图灵奖得

收稿日期: 2017-02-15; 接受日期: 2017-04-07

基金项目: 中组部千人计划青年人才项目和华东师范大学紫江优秀青年项目

* 通讯作者 Author for correspondence. E-mail: jzhang@des.ecnu.edu.cn

主Jim Gray在美国国家科学研究委员会的演讲中描绘了“数据密集型”科研的愿景,将大数据科研从3个科学研究范式(实验归纳、模型推演、仿真模拟)中区别开来,称之为第四范式(Hey et al, 2012)。近年来,大数据科研日益应用于基础数据采集、信息存储、数据挖掘以及科学假设的形成与验证等各个方面。*Nature*和*Science*杂志分别于2008年和2011年出版了大数据专辑(<http://www.nature.com/news/specials/bigdata>; <http://www.sciencemag.org/site/special/data>),表明大数据的影响已涉及自然科学、社会科学、人文科学和工程学等各个领域。基于Web of Science核心数据库检索发现,以“big data”为主题的论文数量从2008年的14篇增加到2016年的3,927篇(图1a)。同时,各国政府也很快制定了大数据发展的相关战略。例如,2012年美国政府启动了“大数据研究和发展倡议”;自2014年起欧盟委员会依托2020地平线、云计算行动等计划主导大数据领域的研发和创新。我国自2012年开始就将大数据科学提上日程,科技部、国家自然科学基金委员会等先后设立了与大数据相关的重点与重大研究项目。2015年,国务院正式印发了《促进大数据发展行动纲要》,并在十八届五中全会上提出实施“国家大数据战略”,标志着大数据发展正式成为国家战略。

生物多样性的起源、维持和丧失机制是生态学研究的基础和前沿问题(Pennisi, 2005; 马克平,

2016),也是生物多样性科学、宏生态学(macroecology)等研究的核心问题(Brown, 1995; 胡慧建等, 2003; Sutherland, 2009)。收集、整理和分析高质量的生物多样性数据(包括物种分类、分布、起源与进化、时空动态等)是认知和应对这些问题的科学基础。生物多样性的数据收集历史至少可以追溯到林奈、达尔文、华莱士、洪堡等时期。尽管已经有了二百多年的数据积累,但生态学家始终苦于缺乏高质量的生物多样性数据(Hortal et al, 2015)。究其原因,主要是我们赖以生存的生态系统本身的复杂性(郭建国和申卫军, 2002)。生态系统的组成单元极其多样且常常存在复杂的相互作用,生态系统中的反馈与调节方式也非常多样且不断变化,系统组分与相关过程也往往表现出高度的时空异质性。Hortal等(2015)对生物多样性研究存在的知识空缺(shortfalls or gaps)进行了归纳和总结,指出我们在新物种发现、已知物种的地理分布、种群数量与时空动态、物种进化史、功能性状、物种与环境之间以及物种与物种之间的相互作用这7个方面都存在着知识上的空缺(详细描述见表1)。这些空缺导致我们当前的生物多样性研究只能限制在一些特定的区域、特定的生物有机体、特定的环境条件等,基于某一个区域或类群提出的理论或假说难于在其他的区域或类群来进行验证,这极大地限制了生态学研究的预测能力(Houlahan et al, 2017)。大数据时代的

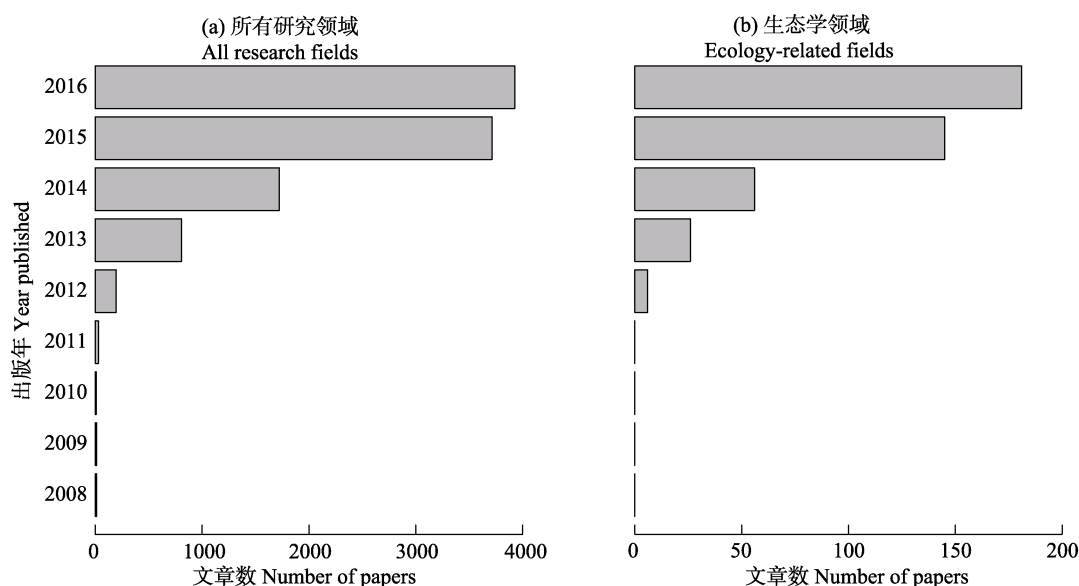


图1 基于Web of Science核心数据库以“big data”为主题的年度论文数量变化(2008–2016)

Fig. 1 The number of papers between 2008 and 2016 using “big data” as the topic in the Web of Science Core database

表1 生物多样性研究中存在的7个主要知识空缺(基于Hortal et al, 2015修改)

Table 1 Definitions for seven main shortfalls (or gaps) of current biodiversity knowledge (Adopted from Hortal et al, 2015)

知识空缺 Knowledge shortfalls	生物多样性的不同方面 Aspect of biodiversity	定义 Definition	相关文献 References
Linnean shortfall	物种 Species	缺乏对世界上很多现存和已灭绝物种描述的知识。Lack of knowledge about the description of most of living and extinct species on Earth.	Brown & Lomolino, 1998; Brito, 2010
Wallacean shortfall	地理分布 Geographic distribution	缺乏有关大多数物种在各个时间尺度上的地理分布的知识。Lack of knowledge about the geographic distribution of most species at all scales most of the time.	Lomolino, 2004
Prestonian shortfall	种群 Populations	缺乏在时空尺度上的物种多度和种群动态的知识。Lack of knowledge about species abundance and population dynamics in space and time.	Cardoso et al, 2011
Darwinian shortfall	进化 Evolution	缺乏关于生命之树以及物种和它们的性状进化的信息。Lack of knowledge about the tree of life and the evolution of species and their traits.	Diniz-Filho et al, 2013
Raunkiaeran shortfall	功能性状和生态功能 Functional traits and ecological functions	缺乏物种的性状及其生态功能的知识。Lack of knowledge about species' traits and their ecological functions.	Hortal et al, 2015
Hutchinsonian shortfall	非生物耐性因子 Abiotic tolerances	缺乏物种如何应对和忍耐非生物因子的知识。Lack of knowledge about the responses and tolerances of species to abiotic conditions.	Cardoso et al, 2011
Eltonian shortfall	生态的相互作用 Ecological interactions	缺乏物种相互作用及其影响个体存活和适合度的知识。Lack of knowledge on species' interactions and these interactions' effects on individual survival and fitness.	Hortal et al, 2015

到来为弥补这些知识空缺提供了可能(Hampton et al, 2013)。

目前, 大数据的挖掘及其应用已经成为国际生物多样性与宏生态学研究的前沿内容。国际上的多个合作研究项目产生了规模空前的生物多样性数据, 如: 全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF, <http://www.gbif.org/>)、美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI, <https://www.ncbi.nlm.nih.gov/>)、全球30 m空间分辨率的森林覆盖变化数据(<https://earthenginepartners.appspot.com/>)等。这些数据既包括基于传统生态学方法或遥感监测所收集的结构化的数据, 也包括由公众通过社交平台、互联网交易平台等产生的非结构化数据(如图像、声音等多媒体信息)。基于这些海量的、高异构性的数据, 生态学家已经开展了一些突破性的研究工作(图1b), 内容涉及生物多样性研究的各个方面。这些工作的开展也极大地推动了生物多样性信息学(Biodiversity Informatics)这一分支学科的发展(Soberon & Peterson, 2004; 王利松等, 2010; 马克平等, 2014)。本文通过全球、大陆和区域上的研究案例展示了大数据在生物多样性研究中的当前进展和巨大潜力, 以期为我国相关领域的发展提供参考和借鉴。

1 全球尺度的大数据研究

在全球尺度下, 生态学家已经通过整合全球范围内的海量遥感数据、不同来源的地面调查数据、由非专业人士提供的公众科学(citizen science)数据(张健等, 2013)等开展了一些重要的研究工作, 内容涉及人类活动(土地利用和覆盖变化、气候变化、城市化等)对生物多样性的影响、物种进化历史、物种灭绝风险评估与预测、生物多样性与生态系统功能的关系等。本文选取了最近发表的4篇文章予以介绍。

(1)森林覆盖变化。森林覆盖率是反映森林资源的丰富程度、生态平衡状况、生物多样性保护状态等的重要指标。2013年由美国马里兰大学Matthew Hansen领导的研究团队利用Landsat-7对地卫星观测数据和Google Earth Engine工具绘制了2000–2012年全球30 m空间分辨率的森林覆盖与变化图(Hansen et al, 2013)。结果发现, 在2000–2012年间, 全球森林覆盖减少了230万km², 同时有80万km²的森林恢复。热带地区既是森林丧失面积最大的区域, 也是森林恢复面积最大的区域。该项工作是在Google的合作下完成的。Hansen等共分析了654,000张遥感图像中的1,430亿个像元(pixel)的森林覆盖变化。Hansen表示, 如果使用一台独立的计算机需要15年才能完成地图绘制, 而使用Google Earth Engine

只要几天就能完成(<https://svs.gsfc.nasa.gov/11393>)。文章发表后引起了科学家和公众的极大关注, Hansen等也将这一系列数据免费共享(<https://earthenginepartners.appspot.com>), 并持续更新(目前已更新到2014年)。这些全球范围的高分辨率的数据已经被应用到生物多样性保护、保护区规划、森林采伐监控、气候变化谈判等诸多领域。在文章发表的3年时间内, 已经获得广泛引用(Web of Science引用930余次、谷歌学术引用1,600余次)。

(2)全球无路区的制图及其保护状态评估。日益密集的道路网络会导致景观碎片化, 进而造成自然生态系统的恶化, 危及生物多样性和生态系统功能。在全球范围内, 没有道路的区域很多都是生物多样性保护和生态系统功能维持的重要区域。2016年国际保护生物学学会(Society for Conservation Biology)的Roadless Initiative团队(<http://www.roadless.online/>)基于完全公开的OpenStreet Map数据(<http://www.openstreetmap.org>), 以1 km为缓冲区, 在全球范围(除南极洲和格陵兰岛)绘制出无路区的地图, 并对其当前的状态进行评估(Ibisch et al, 2016)。他们发现, 尽管地球上80%的区域是无路区, 但是当前的道路建设已经将地球表面划分成了60多万个破碎化斑块, 且半数以上小于1 km², 80%的区域小于5 km², 仅有7%的区域大于100 km²。他们也将无路区的分布与全球保护区的分布进行了比较, 发现仅有9.3%的大而无路区位于当前保护区范围内。这些研究成果是完全共享的, 可在Roadless Initiative的网站免费下载。这些研究成果无疑为全球范围内的生物多样性保护和保护区规划等提供了重要的参考。另外, 值得指出的是, OpenStreet-Map数据是基于公众的力量来不断完善的, 表明公众科学可以在生物多样性相关的研究中发挥积极的作用(张健等, 2013)。

(3)生物多样性与生产力之间的关系。生物多样性与生产力或生态系统功能之间的关系一直是生态学研究热点问题(马克平, 2016)。尽管科学家已经在这个领域开展了很多研究工作, 但还缺少全球尺度上的整合分析。2016年美国西弗吉尼亚大学 Jingjing Liang等组成的研究团队汇集了44个国家的77万多个森林永久调查样地的数据, 包括了三千多万棵树和8,737个树种(Liang et al, 2016)。基于这一海量的数据集, 他们发现生物多样性与生产力之间

存在着全球一致的正相关性, 表明目前极高的生物多样性丧失速率将导致全球森林生产力的加速下降。鉴于这些大数据科学的证据, 他们再次强调了生物多样性保护的重要性, 指出需要在全球范围内重新评估生物多样性、保护优先区域等的价值。

(4)全球光污染地图。随着经济和社会的发展, 城市化进程加快, 人工照明发展迅速, 光污染所引起的生态问题也日益严重。其中, 光污染对动植物的生长、繁殖、迁徙等的影响是生物多样性保护中的新的研究热点(Gaston et al, 2015)。根据高精度卫星成像数据(the VIIRS DNB sensor on the Suomi National Polar-orbiting Partnership), 并结合由科学家和公众共同采集的全球20,865个地面站点数据, 由意大利的高中物理老师Fabio Falchi领导的团队绘制出迄今最精确的全球光污染地图(Falchi et al, 2016)。结果显示, 地球上超过80%的人类生活在受人工光线污染的天空下, 这种趋势还在继续恶化。全球近1/3的人, 包括北美近80%的人与欧洲60%的人, 因光污染而不能看到银河。他们指出这些光污染还将导致一系列的其他问题, 包括对生物多样性的影响。根据论文公众影响力评估系统Altmetric的统计, Falchi等(2016)的Altmetric指数为3,020, 列在2016年论文Altmetric指数百强榜第10位。

2 大陆和区域尺度的大数据研究

与全球尺度的大数据研究相比, 大陆和区域尺度下, 海量的地面调查数据和高精度遥感数据的获取、规范化和数据共享相对容易, 同时这也是政府部门和公众更加关注的空间范围。因此, 大数据在大陆和区域尺度下的生物多样性研究中的应用近年来发展较为迅速。例如, 最近提出了生态学的新分支“宏系统生态学”(Macrosystems Ecology), 旨在更好地整合大陆和区域尺度的研究工作(Heffernan et al, 2014)。根据Heffernan等(2014)的定义, 宏系统生态学是“研究从区域到大陆尺度多样化的生态学现象及其与其他尺度(全球和局域尺度)的相互作用”。生态学期刊*Frontiers in Ecology and the Environment*和*Landscape Ecology*分别于2014年和2016年出版专刊报道了这个领域的研究进展(Soranno & Schimel, 2014; Fei et al, 2016)。本文选取了4个项目来简要介绍。

(1)美国和加拿大森林固定样地调查。固定监测

样地为生态学研究提供了大量的基础数据。在北美地区, 美国和加拿大林业部门已经积累了半个多世纪的森林样地调查数据, 并已将其广泛应用到科学研究和国家政策的制定中。美国的森林调查与分析项目(Forest Inventory and Analysis, FIA; <https://www.fia.fs.fed.us/>)目前已经设置了十多万永久监测样地, 对每个样地超过100项的林地属性进行测量或计算, 并对数百万棵树进行测量与跟踪(Bechtold & Patterson, 2005)。加拿大的森林固定样地数据由各个省的林业部门来负责, 调查规范在各省之间有一些差异, 迄今已经收集了数万个样地的数据。基于这些海量的数据, 近年来已经开展了很多关于生物多样性保护、气候变化等方面的研究工作。例如, Zhang等(2015)通过整合植被调查和历史气候数据分析了加拿大西部过去50年的森林动态变化, 发现树木竞争在驱动森林的动态变化中比气候变化的作用更为重要。这一观点不同于许多气候变化相关的研究结果, 强调了植物竞争等内在过程在理解群落动态过程中的重要性(Zhang et al, 2015)。Zhang等(2017)通过整合美国和加拿大的样地调查、植物标本、谱系树等数据, 用物种分布模型分析了未来气候变化情景下7,465种植物的灭绝风险, 结果发现近1/3的物种在世纪末将面临着极大的灭绝风险, 并存在着极大的区域差异。这一工作为北美区域的植物多样性保护、自然保护区规划等提供了重要参考。

(2)北美繁殖鸟类调查。北美的繁殖鸟类调查(North American Breeding Bird Survey, BBS, <https://www.pwrc.usgs.gov/bbs/>)已经有半个多世纪的历史。该项目组织志愿者在每年鸟类繁殖高峰期开展鸟类调查。每条样线长39.43 km, 每隔805 m设置1个记录点, 每条样线共50个记录点。到目前为止, 该项目在北美大陆已有4,100多条调查样线, 记录了420多种鸟类。通过与其他数据联合, 该项目已发表450余篇相关的研究论文(<https://www.pwrc.usgs.gov/bbs/about/>)。例如, Keitt和Stanley (1998)发表在*Nature*上的文章中使用了3,000多条样线连续31年的调查记录来分析鸟类种群的年际动态。LaDeau等(2007)发表在*Nature*上的文章中分析了西尼罗河病毒(West Nile virus)爆发与鸟类种群下降之间的关系。Dobson等(2015)分析了鸟类多样性与多度和植物生产力之间的关系, 发现了非常弱的相关性。这

些工作为评价气候变化和土地利用等对鸟类多样性的长期影响提供了重要的证据。

(3)加拿大阿尔伯塔生物多样性监测。以上两个例子是基于单个类群(taxon)的研究, 并没有考虑各类群之间如鸟类与植物、兽类与植物的相互作用。不同类群之间如何相互协调来共同维持生物多样性是一个重要的研究议题(Terborgh, 2015)。加拿大阿尔伯塔省的生物多样性监测网络(Alberta Biodiversity Monitoring Institute, ABMI, <http://abmi.ca>)为回答这一问题提供了数据支持。ABMI项目始于2003年, 该研究将整个阿尔伯塔省分成1,656个20 km × 20 km的网格, 在每个网格内分别设置1个陆地固定样点和1个湿地固定样点, 调查维管束植物、苔藓、地衣、鸟类、哺乳动物、昆虫等类群及其生境条件, 并计划每5年复查一次。同时, ABMI通过整合数十个不同的数据源来绘制全省范围的高分辨率的土地利用覆盖图(ABMI, 2016), 并每2年更新一次, 目前已经有4次数据, 单次产生的数据量在1–2 Gb。ABMI通过整合这些详细的多类群的地面调查数据和海量的高分辨率遥感数据来评价生态系统的完整性、分析不同干扰状态下的生物多样性维持机制(Mayor et al, 2012; Zhang et al, 2013, 2014b)、分析气候变化对物种分布的影响(Stralberg et al, 2015; Zhang et al, 2015)、估计区域生物量分布(Zhang et al, 2014a)等。

(4)近地面遥感与功能多样性。功能性状及其多样性是连接生物多样性与生态系统功能的关键性因素, 是近年来生态学的研究热点之一(Violle et al, 2014; 刘晓娟和马克平, 2015)。然而, 由于野外测量功能性状工作量很大, 目前大多数关于功能性状数据的采集仅限制在局域尺度上。最近, 由美国科学院院士Gregory Asner领导的研究团队在*Science*上报道了他们基于搭载激光雷达(LiDAR)扫描仪和光谱成像仪的飞机对7,600万ha秘鲁热带雨林进行的航空调查, 从空中测量了7个功能性状(比叶重、叶氮含量、叶磷含量、叶钙含量等), 获得了整个区域的功能性状变异图以及功能多样性指数(Asner et al, 2017)。在此次调查中, 光谱和LiDAR采样点的总数分别为50.1亿和204.5亿, 最终数据的空间分辨率为2 m。他们用这些数据集将研究区域分为36个功能等级和6个功能组, 以此来确定生物多样性保护的优先区域。

3 总结与展望

基于以上全球、大陆和区域尺度上的研究案例可以发现,科学大数据在生物多样性科学和宏生态学研究已经崭露头角,并有极大潜力去弥补当前生物多样性研究存在的7个知识空缺(表1)。需要指出的是,我们选取的几个研究案例并不能覆盖生物多样性大数据应用的各个方面。除此之外,还有很多重要的研究工作和研究平台。例如,全球生物多样性信息网络(GBIF)、Map of Life (MOL, <https://www.mol.org/>)、亚洲植物多样性数字化计划(马克平, 2017)、植物学信息与生态网络(Botanical Information and Ecology Network, BIEN; Enquist et al, 2017)等全球或区域物种分布信息平台;美国国家生态观测网络(National Ecological Observatory Network, NEON, <https://www.neoninc.org/>)、长期生态学研究网络(The Long Term Ecological Research Network, LTER, <https://www.lternet.edu/>)等观测平台(戴圣骐和赵斌, 2016);全球物种系统发育树的构建与研究平台,如:Jetz等(2012)构建的全球9,993种鸟类的进化树、Zanne等(2014)构建的全球31,749种种子植物的进化树,以及整合了230多万个末端分支(tips)的Open Tree of Life 项目(Hinchliff et al, 2015, <http://opentreeoflife.org/>)等。

通过这些大数据的案例可以看出,生物多样性的相关数据不仅规模庞大,分布在不同的组织机构,而且异构性和动态性高,数据不完整性和不确定性强。这些特征对数据采集、处理、保存、共享等提出了巨大的挑战(Hampton et al, 2013)。

首先,现有的生物多样性数据存在大量的冗余和噪音,包括采样地理偏差、采样类群或科属偏差、物种识别的错误、数据记录不准确或不完整等(Meyer et al, 2016; 孙航等, 2017)。这些原始数据的偏差如果考虑不周或处理不当,很可能导致错误的结论。Meyer等(2015)分析了全球21,170种陆地脊椎动物的1.57亿个物种的出现记录,发现这些数据存在着极大的区域偏差,尤其是在很多非西方国家,如中国、印度、俄罗斯、印度尼西亚等。基于我国国家标本资源共享平台(NSII, <http://www.nsii.org.cn/>)的29,012个植物物种的423万个标本和文献记录, Yang等(2013)发现91%的中国县域存在采样不足,

并表现出明显的地理差异。如何系统地评估生物多样性大数据的完整性以及如何有效地降低冗余和噪音对研究结果的影响是目前亟待解决的问题。

其次,生物多样性数据的高异构性和动态性给数据挖掘和分析带来了诸多困难。生物多样性研究涉及从基因到生物圈所有的生物组织层次,同时需要整合气候、土壤、人类干扰、社会经济等环境要素。如何快速有效地将这些多源数据转换为有用的知识是一大挑战(Jones et al, 2006)。为应对这一挑战,一方面,需要对各个层级的数据有足够的了解,围绕研究问题对多源数据进行降维;另一方面,对数据分析方法(如深度学习)和数据处理系统(如高性能计算、并行处理技术)提出了较高的要求。需要通过计算机、统计学、地理学、遥感科学等领域的专家合作来推动生物多样性这一交叉学科的发展。

第三,从以上生物多样性大数据的研究案例可以看出,目前大数据在该领域的应用还多局限于分布格局的描述、已有理论的校验印证,尚未取得理论上的突破。生态学家长期以来一直在争论群落生态学是否存在普适性的理论(Lawton, 1999; Roughgarden, 2009),但至今仍未达成共识。其中的一个原因是我们对自然了解得太少,缺乏强有力的数据支持(Houlahan et al, 2017)。大数据时代所带来的科研范式的转变可能为生物多样性领域新思想和新理论等的形成提供了温床,今后需要重视大数据研究中思想和理论的提炼总结。

综上所述,大数据时代的到来给生物多样性科学与宏生态学研究带来了诸多机会和挑战。大数据所带来的不仅仅是研究方法的革命,更重要的是思维方式的革命(Mayer-Schönberger & Cukier 2013)。面对不断加剧的人类活动和日益严峻的生物多样性保护形势,我们需要在适应大数据思维的条件下重新审视生物多样性研究的理论基础与应用问题。正如Soranno和Schimel (2014)所说:“理解和解决当前面临的大多数问题,生态学家需要大数据和大生态学”。

致谢: 非常感谢马克平主编的邀请与诸多鼓励,感谢两位审稿人提出了很有参考价值的修改意见。

参考文献

- ABMI (Alberta Biodiversity Monitoring Institute) (2016) Alberta Wall-to-Wall Vegetation Layer Including "Backfilled" Vegetation in Human Footprints (Version 6). <http://abmi.ca/>. (accessed on 2017-03-22)
- Asner GP, Martin RE, Knapp DE, Tupayachi R, Anderson CB, Sinca F, Vaughn NR, Llacayo W (2017) Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science*, 355, 385–389.
- Bechtold WA, Patterson PL (2005) The Enhanced Forest Inventory and Analysis Program: National Sampling Design and Estimation Procedures. General Technical Report, SRS-80. USDA Forest Service, Southern Research Station, Asheville, NC.
- Brown JH (1995) *Macroecology*. Chicago University Press, Chicago.
- Brown JH, Lomolino MV (1998) *Biogeography*. Sinauer Press, Sunderland, Massachusetts.
- Brito D (2010) Overcoming the Linnean shortfall: data deficiency and biological survey priorities. *Basic and Applied Ecology*, 11, 709–713.
- Cardoso P, Erwin TL, Borges PA, New TR (2011) The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, 144, 2647–2655.
- Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM (2013) Darwinian shortfalls in biodiversity conservation. *Trends in Ecology and Evolution*, 28, 689–695.
- Dai SQ, Zhao B (2016) Trends and challenges of ecosystem observations in the age of big data. *Biodiversity Science*, 24, 85–94. (in Chinese with English abstract) [戴圣骐, 赵斌 (2016) 大数据时代下的生态系统观测发展趋势与挑战. *生物多样性*, 24, 85–94.]
- Dobson LL, La Sorte FA, Manne LL, Hawkins BA (2015) The diversity and abundance of North American bird assemblages fail to track changing productivity. *Ecology*, 96, 1105–1114.
- Enquist BJ, Condit R, Peet RK, Schildhauer M, Thiers BM (2017) Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ*, 4, e2615v2.
- Falchi F, Cinzano P, Duriscoe D, Kyba CCM, Elvidge CD, Baugh K, Portnov BA, Rybnikova NA, Furgoni R (2016) The new world atlas of artificial night sky brightness. *Science Advances*, 2, e1600377.
- Fei S, Guo Q, Potter K (2016) Macrosystems ecology: novel methods and new understanding of multi-scale patterns and processes. *Landscape Ecology*, 31, 1–6.
- Gantz J, Reinsel D (2012) The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. IDC (International Data Corporation), Framingham.
- Gaston KJ, Visser ME, Hölker F (2015) The biological impacts of artificial light at night: the research challenge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140133.
- Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11, 156–162.
- Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JRG (2013) High-resolution global maps of 21st-century forest cover change. *Science*, 342, 850–853.
- Heffernan JB, Soranno PA, Angilletta MJ, Buckley LB, Gruner DS, Keitt TH, Kellner JR, Kominoski JS, Rocha AV, Xiao J, Harms TK, Goring SJ, Koenig LE, McDowell WH, Powell H, Richardson AD, Stow CA, Vargas R, Weathers KC (2014) Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, 12, 5–14.
- Hey T, Tansley S, Tolle K (translated by Pan XF, Zhang XL) (2012) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Science Press, Beijing. (in Chinese) [潘教峰, 张晓林等(译) (2012) 第四范式: 数据密集型科学发现. 科学出版社, 北京.]
- Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, Crandall KA, Deng J, Drew BT, Gazis R, Gude K, Hibbett DS, Katz LA, Laughinghouse HD, McTavish EJ, Midford PE, Owen CL, Ree RH, Rees JA, Soltis DE, Williams T, Cranston KA (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences, USA*, 112, 12764–12769.
- Hortal J, Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549.
- Houlahan JE, McKinney ST, Anderson TM, McGill BJ (2017) The priority of prediction in ecological understanding. *Oikos*, 26, 1–7.
- Hu HJ, Jiang ZG, Wang ZW (2003) Macroecology: concept and progresses. *Acta Ecologica Sinica*, 23, 1192–1199. (in Chinese with English abstract) [胡慧建, 蒋志刚, 王祖望 (2003) 宏生态学(Macroecology)及其研究. *生态学报*, 23, 1192–1199.]
- Ibisch PL, Hoffmann MT, Kreft S, Pe'er G, Kati V, Biber-Freudenberger L, DellaSala DA, Vale MM, Hobson PR, Selva N (2016) A global map of roadless areas and their conservation status. *Science*, 354, 1423–1427.
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature*, 491, 444–448.
- Jones MB, Schildhauer MP, Reichman OJ, Bowers S (2006) The new bioinformatics: integrating ecological data from the

- gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, 37, 519–544.
- Keitt TH, Stanley HE (1998) Dynamics of North American breeding bird populations. *Nature*, 393, 257–260.
- LaDeau SL, Kilpatrick AM, Marra PP (2007) West Nile virus emergence and large-scale declines of North American bird populations. *Nature*, 447, 710–713.
- Lawton JH (1999) Are there general laws in ecology? *Oikos*, 84, 177–192.
- Liang JJ, Crowther TW, Picard N, Wiser S, Zhou M, Alberti G, Schulze ED, McGuire AD, Bozzato F, Pretzsch H, de-Miguel S, Paquette A, Hérault B, Scherer-Lorenzen M, Barrett CB, Glick HB, Hengeveld GM, Nabuurs GJ, Pfautsch S, Viana H, Vibrans AC, Ammer C, Schall P, Verbyla D, Tchebakova N, Fischer M, Watson JV, Chen HYH, Lei X, Schelhaas MJ, Lu H, Gianelle D, Parfenova EI, Salas C, Lee E, Lee B, Kim HS, Bruelheide H, Coomes DA, Piotta D, Sunderland T, Schmid B, Gourlet-Fleury S, Sonké B, Tavani R, Zhu J, Brandl S, Vayreda J, Kitahara F, Searle EB, Neldner VJ, Ngugi MR, Baraloto C, Frizzera L, Bałazy R, Oleksyn J, Zawila-Niedzwiecki T, Bouriaud O, Bussotti F, Finér L, Jaroszewicz B, Jucker T, Valladares F, Jagodzinski AM, Peri PL, Gonmadje C, Marthy W, O'Brien T, Martin EH, Marshall AR, Rovero F, Bitariho R, Niklaus PA, Alvarez-Loayza P, Chamuya N, Valencia R, Mortier F, Wortel V, Engone-Obiang NL, Ferreira LV, Odeke DE, Vasquez RM, Lewis SL, Reich PB (2016) Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354, aaf8957.
- Liu XJ, Ma KP (2015) Plant functional traits—concepts, applications and future directions. *Scientia Sinica Vitae*, 45, 325–339. (in Chinese with English abstract) [刘晓娟, 马克平 (2015) 植物功能性状研究进展. *中国科学: 生命科学*, 45, 325–339.]
- Lomolino MV (2004) Conservation biogeography. In: *Frontiers of Biogeography: New Directions in the Geography of Nature* (eds Lomolino MV, Heaney LR), pp. 293–296. Sinauer Press, Sunderland, Massachusetts.
- Ma KP (2014) Rapid development of biodiversity informatics in China. *Biodiversity Science*, 22, 251–252. (in Chinese) [马克平 (2014) 生物多样性信息学在中国快速发展. *生物多样性*, 22, 251–252.]
- Ma KP (2016) Hot topics for biodiversity science. *Biodiversity Science*, 24, 1–2. (in Chinese) [马克平 (2016) 生物多样性科学的热点问题. *生物多样性*, 24, 1–2.]
- Ma KP (2017) Mapping Asia Plants: a cyberinfrastructure for plant diversity in Asia. *Biodiversity Science*, 25, 1–2. (in Chinese) [马克平 (2017) 亚洲生物多样性数字化计划. *生物多样性*, 25, 1–2.]
- Mayer-Schönberger V, Cukier K (2013) *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, Boston.
- Mayor S, Cahill J, He F, Solyomos P, Boutin S (2012) Regional boreal biodiversity peaks at intermediate human disturbance. *Nature Communications*, 3, 1142.
- Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, 6, 8221.
- Meyer C, Weigelt P, Kreft H (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006.
- Pennisi E (2005) What determines species diversity? *Science*, 309, 90.
- Roughgarden J (2009) Is there a general theory of community ecology? *Biology & Philosophy*, 24, 521–529.
- Soberon J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 689–698.
- Soranno PA, Schimel DS (2014) Macrosystems ecology: big data, big ecology. *Frontiers in Ecology and the Environment*, 12, 3.
- Stralberg D, Matsuoka SM, Hamann A, Bayne EM, Solyomos P, Schmiegelow F, Wang X, Cumming SG, Song SJ (2015) Projecting boreal bird responses to climate change: the signal exceeds the noise. *Ecological Applications*, 25, 52–69.
- Sun H, Deng T, Chen YS, Zhou Z (2017) Current research and development trends in floristic geography. *Biodiversity Science*, 25, 111–122. (in Chinese with English abstract) [孙航, 邓涛, 陈永生, 周卓 (2017) 植物区系地理研究现状及发展趋势. *生物多样性*, 25, 111–122.]
- Sutherland WJ, Adams WM, Aronson RB, Aveling R, Blackburn TM, Broad S, Ceballos G, Côté IM, Cowling RM, Dafonseca GAB, Dinerstein E, Ferraro PJ, Fleishman E, Gascon C, Hunter Jr M, Hutton J, Kareiva P, Kuria A, MacDonald DW, Mackinnon K, Madgwick FJ, Mascia MB, Mcneely J, Milner-Gulland EJ, Moon S, Morley CG, Nelson S, Osborn D, Pai M, Parsons ECM, Peck LS, Possingham H, Prior SV, Pullin AS, Rands MRW, Ranganathan J, Redford KH, Rodriguez JP, Seymour F, Sobel J, Sodhi NS, Stott A, Vance-Borland K, Watkinson AR (2009) One hundred questions of importance to the conservation of global biological diversity. *Conservation Biology*, 23, 557–567.
- Terborgh JW (2015) Toward a trophic theory of species diversity. *Proceedings of the National Academy of Sciences, USA*, 112, 11415–11422.
- Violle C, Reich PB, Pacala SW, Enquist BJ, Kattge J (2014) The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences, USA*, 111, 13690–13696.
- Wang LS, Chen B, Ji LQ, Ma KP (2010) Progress in biodiversity informatics. *Biodiversity Science*, 18, 429–443. (in Chinese with English abstract) [王利松, 陈彬, 纪力强, 马克平 (2010) 生物多样性信息学研究进展. *生物多样性*, 18, 429–443.]
- Wu JG, Shen WJ (2002) The sciences of complexity and eco-

- logical applications. In: *Lectures in Modern Ecology (II): From Basic Ecology to Environmental Issues* (eds Wu JG, Han XG, Huang JH), pp. 6–15. China Science and Technology Press, Beijing. (in Chinese) [郭建国, 申卫军 (2002) 复杂性科学及其生态学应用. 见: 现代生态学讲座 (二): 基础研究与环境问题(郭建国, 韩兴国, 黄建辉主编), 6–15页. 中国科学技术出版社, 北京.]
- Yang WJ, Ma KP, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography*, 40, 1415–1426.
- Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlenn DJ, O'Meara BC, Moles AT, Reich PB, Royer DL, Soltis DE, Stevens PF, Westoby M, Wright IJ, Aarssen L, Bertin RI, Calaminus A, Govaerts R, Hemmings F, Leishman MR, Oleksyn J, Soltis PS, Swenson NG, Warman L, Beaulieu JM (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature*, 506, 89–92.
- Zhang J, Chen SB, Chen B, Du YJ, Huang XL, Pan XB, Zhang Q (2013) Citizen science: integrating scientific research, ecological conservation and public participation. *Biodiversity Science*, 21, 738–749. (in Chinese with English abstract) [张健, 陈圣宾, 陈彬, 杜彦君, 黄晓磊, 潘绪斌, 张强 (2013) 公众科学: 整合科学研究, 生态保护和公众参与. 生物多样性, 21, 738–749.]
- Zhang J, Huang S, Hogg E, Lieffers V, Qin Y, He F (2014a) Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences*, 11, 2793–2808.
- Zhang J, Kissling WD, He F (2013) Local forest structure, climate and human disturbance determine regional distribution of boreal bird species richness in Alberta, Canada. *Journal of Biogeography*, 40, 1131–1142.
- Zhang J, Mayor SJ, He F (2014b) Does disturbance regime change community assembly of angiosperm plant communities in the boreal forest? *Journal of Plant Ecology*, 7, 188–201.
- Zhang J, Nielsen SE, Chen Y, Georges D, Qin Y, Wang SS, Svenning JC, Thuiller W (2017) Extinction risk of North American seed plants elevated by climate and land-use change. *Journal of Applied Ecology*, 54, 303–312.
- Zhang J, Nielsen SE, Stolar J, Chen Y, Thuiller W (2015) Gains and losses of plant species and phylogenetic diversity for a northern high-latitude region. *Diversity and Distributions*, 21, 1441–1454.

(责任编辑: 米湘成 责任编辑: 黄祥忠)