

•综述•

大数据时代下的生态系统观测发展趋势与挑战

戴圣骐 赵 斌*

(复旦大学长江河口湿地生态系统野外科学观测研究站, 生物多样性与生态工程教育部重点实验室, 上海 200438)

摘要: 随着观测技术的发展, 生态学研究尺度不断扩大。生态系统观测从小规模合作、短时间个人观测向大规模、长时间、跨学科、多因子联合观测转变。传感器技术的革新带来了生态观测在时空尺度的扩展与精确度上的提升, 致使生态学观测数据的容量、产生速度与数据种类飞速增长。对生态系统数据获取、存储与管理的传统方法无疑不再能满足现代生态学研究的要求。因此, 我们建议以大数据时代的数据存储、管理与处理技术为基础, 整合生态物联观测网络(Internet of Ecology)、公民科学观测网络以及基于标准化数据管理的研究者网络互联, 建立整合生态系统观测平台来应对这一困境。为生态学研究打造一站式生态观测服务, 是大数据时代下生态系统观测的大势所趋。

关键词: 传感器; 观测网络; 物联网; 公民科学; 社交网络

Trends and challenges of ecosystem observations in the age of big data

Shengqi Dai, Bin Zhao*

Coastal Ecosystems Research Station of the Yangtze River Estuary, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Fudan University, Shanghai 200438

Abstract: With the development of observation technology, the scale of ecological research is increasing. Ecological observations have turned from small-scale, short-time individual observations into broad-scale, long-term, interdisciplinary, multi-factor group observations. Innovation in sensor techniques has led to a profound evolution in time and space precision of ecological observations, while the volume, type, and generating speed of these observational data are increasing, which indicates that traditional ecological data acquisition, storage and management methods cannot afford the demands of modern ecological research. With the assistance of new big data storage, management and processing techniques, integrated with the Internet of Ecology, a citizen science observational network and standardized data management network, we can build an ecological observation system to resolve these issues. The concept to provide a one-station ecological observation service to researchers represents the general trend of the development of ecological observations in the age of big data.

Key words: sensors; observation network; internet of things; citizen science; social network

21世纪初, 人类满怀信心地跨入了信息时代, 许多科学研究的前沿领域面临新机遇和新发展, 新理念层出不穷, 研究方式迎来了天翻地覆的变化。同时, 全球范围内的重大问题, 从全球气候变化、人口增长、粮食安全、疾病传播、清洁水短缺到生物多样性以及生态系统服务(ecosystem services)功能的丧失, 为生态学家步入这个新时代带来了压力

(赵斌, 2014)。生态系统变化正跨越局域尺度扩展至更大范围, 驱使生态学者们去发展、整合并适应新的研究方法(Heffernan et al, 2014)。而宏观生态系统研究旨在从更广的时空尺度上推进对生态系统的理解, 其研究内容不仅包括传统意义上的小样地或季节变化的问题, 还包含更大时空跨度(数年、数十年或数世纪; 数十至数百平方公里甚至全球)

收稿日期: 2015-06-03; 接受日期: 2015-09-05

基金项目: 国家自然科学基金(31170450)

* 通讯作者 Author for correspondence. E-mail: zhaobin@fudan.edu.cn

上的重大问题(Heffernan et al, 2014)。

现代生态学之所以发展成一门综合性大学科,首先源于生态系统的内在复杂性,其次得益于地球、生命和社会科学等诸多领域的共同发展诉求。现代生态学不仅与生命科学相关领域(如地理学、海洋学、气候学、基因与进化生物学等)产生了千丝万缕的联系,还与数学、经济学、信息科学等学科进行频繁交流与跨领域合作。这些都促使生态学家们在更大时空尺度、更多领域收集数据,进而完成更为复杂的分析(Goring et al, 2014)。

1 生态系统观测发展历程

生态系统观测,实质上是将自然界的现象及其变化以显式数据(所有可转化为数字表达的变量)或隐式数据(例如文本或图像信息)的方式记录下来的过程。该过程可以将生态系统中生物体与环境的相互作用以数据形式映射到数据界(data kingdom),进而挖掘数据中显示出的生物与环境相互关系(Watt, 2013)。

生态系统观测方式大体分为两种:人工实地观测与传感器自动记录。人类对生态系统的直接观察是经典研究方法,灵活性较高,但受限于人类的身体条件。面对大规模、长时间、高精度的观测需求,以及恶劣的野外条件,人工实地观测往往无法达到现代生态系统观测的需求(Wang et al, 2007),传感器数据收集逐渐变成主要途径。近年来,针对生态系统观测而开发的传感器不断涌现,由它们组成的传感器网络能更有效、更全面地获取观测数据(Conant, 2006; Hart & Martinez, 2006)。自此,生态系统观测跨入快速发展的黄金期(Shamoun-Baranes et al, 2014)。

一个新的现象是,在过去数十年里,生态学家与环境学家们不断诟病相关传感器数量不足与性能不佳等问题(Baraniuk, 2011)。诚然,由于传感器技术不成熟与成本过高等原因(Collins et al, 2006),曾经的传感器覆盖的观测范围小、观测精度差,所收集的数据量十分有限(Porter et al, 2005)。但是,我们应该看到,新型传感器不仅很好地控制了成本,而且其种类也呈爆炸式的增长态势(Rundel et al, 2009),这种革新得益于传感器开发与行业应用的相互促进。例如,数码相机中的图像传感器(charge-coupled device, CCD)原本是从航天遥感器

的技术转化而来,由于商业化普及,其技术革新反过来又促使新型廉价微型传感器的产生,最终促成了低成本卫星群的出现(Butler, 2014; Jones, 2014)。搭载廉价的小型无线传感器与高带宽无线数据传输设施(Rundel et al, 2009)的无线传感器网络,将成为现代生态系统观测中提高时空尺度与精度的优秀解决方案(Porter et al, 2005; Hart & Martinez, 2006)。总地来看,现代生态系统观测的发展方向主要有两个:一是观测尺度扩大,二是观测精度提高。

1.1 观测尺度扩大

现代生态系统观测所跨越的空间尺度越来越大,不仅可以分析小区域尺度动物种群或个体的动态(Turner, 2014),还能在整个地球生物圈尺度考察气候变化(Michener & Jones, 2012)。许多生态学研究依赖于观测数据的时空连续性,所以选择高效且系统化的观测手段显得尤其重要(Turner et al, 2015)。目前,系统化的生态系统观测方法主要有两种:对地遥感观测与网络化陆地观测。

遥感观测是无值守、无干扰、无偏差的数据获取方法(宫鹏, 2010; Willis, 2015)。无论是计算区域尺度上的土地覆盖利用变化(Horning et al, 2010),还是调查全球陆地水储存分布(Velicogna & Wahr, 2006),甚至是评估全球生物多样性(Gillespie et al, 2008),都离不开遥感观测所获取的大量连续数据(Turner et al, 2015)。覆盖度高的多雷达观测数据还可用于在大陆尺度上长期监控飞行生物的迁徙活动(Shamoun-Baranes et al, 2014)。

陆地观测经历了从单站点到大规模观测网络的转变(Conant, 2006; Rundel et al, 2009)。创建于1997年的全球碳通量网(FLUXNET, <http://fluxnet.ornl.gov/>)是生态领域网络化观测的先驱者,旨在将微气象学区域站点整合为全球化观测网络并免费发布(Baldocchi et al, 2001)。如Jung等(2009)曾利用该网络数据进行分析,绘制了生物-大气圈物质通量分布图。

生态系统观测中遥感观测与陆地观测相辅相成,在这方面最成功的案例就是美国国家生态观测网络(National Ecological Observatory Network (NEON), 2011)。该网络使用各类传感器约15,000个,其中陆地观测指标约有500个大类,包含气象、土壤、植被、大气化学和水体等(Kampe et al, 2010)。

遥感观测覆盖全球, 不仅包括传统的多光谱数据, 还包括新型的高分辨率和高光谱数据, 甚至扩展至荧光数据与重力场数据。借助这些数据, 科研人员从大尺度着手进行了诸如综合考察生物入侵模式(Melbourne et al, 2007)、预测物种丰富度(Elmendorf & Moore, 2008)、统计森林覆盖变化(Huang et al, 2009)与记录各生态区碳动态(Zulueta et al, 2011)等研究。

1.2 观测精度提高

除了观测尺度的扩大, 观测理论的发展(Baldocchi et al, 2001; Field et al, 2006)与传感器技术的进步(Collins et al, 2006; Hart & Martinez, 2006)还带来了观测精度的提升。

首先, 实时高频传感器提高了观测的时间精度。例如, 在碳通量测量中, 传统气室法或闭路碳通量测量方法每几天或十几分钟抽样1次。这种低频的测量不仅容易丢失通量变化热点数据, 还存在对生态系统干扰大、观测尺度小等局限性(Baldocchi, 2014)。新兴的涡度通量测量装置, 其采样频率高达10 Hz, 具有高精度、快速响应、连续原位观测、对生态系统干扰低等优点(Anthoni et al, 2002; Baldocchi, 2014), 该方法的出现无疑为通量研究带来了革新。

其次, 得益于遥感器观测精度的提升, 具备超高空间分辨率的大尺度观测逐渐成为主流(Hyyppä et al, 2000; Horning et al, 2010)。过去, 遥感观测的空间分辨率低于1 km, 在生态学研究只能用于粗略判断。而现代超高分辨率卫星产品的空间分辨率达到米级甚至亚米级, 如IKONOS或QuickBird产品的空间分辨率分别达到1 m或0.61 m。生态学家们可将这些高精度遥感数据用于模型训练与验证。例如, Mairota等(2015)使用超高分辨观测(Very High Resolution Earth Observation, VHREO)影像在大空间尺度上生成了精细的动物栖息地斑块分布图; 而Shamoun-Baranes等(2014)则借助于新型光学雷达监测飞行生物的飞行高度、速度、空间分布与飞行方向等信息, 大幅提升了鸟类迁徙监测的精确度。

总之, 生态学已从一门依靠小规模合作与短时间个人观测的实验学科演化为一门实施长时间大规模观测、跨学科合作的大数据学科(Schatz, 2014)。

2 生态系统观测现状与困境

伴随生态系统观测尺度与精度的不断上升, 以及数据科学思维在生态学领域的不断深化, 大量曾经隐藏的信息被挖掘出来(Hampton et al, 2013), 生态系统观测中数据总量及其产生速度与种类呈爆炸式增长(Jones et al, 2006; Reichman et al, 2011)。生态系统观测将面临数据传输、数据存储、数据整合和数据整合等4个方面的挑战。

2.1 高通量数据传输

现代生态系统观测中, 高频传感器与富媒体传感器使用比例上升, 导致数据量剧增。以我们搭建于上海崇明东滩的碳通量观测站为例, 每天产生的观测数据量就高达150 MB(表1)。面对如此高的数据流量, 现代观测数据传输却主要依靠存储卡存储与网络传输(图1)。因为大多数生态观测站点都位于偏远地区, 难以架设稳定运行的有线传输网络, 往往只能通过频繁更换存储卡来维持数据交换(Porter et al, 2012)。作为另一种解决方案的无线网络传输, 虽然能将实时观测数据传输至远程数据库, 但成本高且易受到通讯问题的影响, 数据传输稳定性大打折扣(宫鹏, 2010)。因此, 大部分野外观测站实际上采用了存储卡存储(异步传输)与网络传输(同步传输)相结合的方案来完成数据传输。可见, 若要从本质上解决高通量数据的传输问题, 除了使用更高效的传输技术外, 还应从生态学研究方式出发改进信号处理和数据压缩技术, 例如引入传感器端数据预处理技术, 充分利用观测平台中闲置的计算资源生成数据产品, 成倍减少待传输数据量; 还可以预整合并筛选原始数据, 减少无效数据和冗余数据的传输(Baraniuk, 2011)。

2.2 生态大数据存储

国际数据公司(International Data Cooperation, IDC)报告显示, 截至2013年, 世界范围内传感器获取的数据总量以每年58%的速度激增。而根据摩尔定律, 数据的存储能力(指硬盘、记忆芯片以及其他存储介质)每年仅增加40%, 数据介质成本每年只下降30% (图2)。而NEON、LTER等大型观测网络的建成又将生态学数据容量提高到了新的层次(Michener & Jones, 2012), 这意味着传感器获取的数据总量与现有数据存储能力之间的差距正在逐渐拉大, 导致部分观测数据将无法得到合理存储

表1 典型的涡度相关碳通量观测站仪器搭载, 测量参数以及数据流量概览
Table 1 Introduction to typical instruments carried by an eddy carbon flux observation tower, their measure parameters and data flux

仪器名称 Instrument name	测量参数 Measure parameters	工作频率 Operating frequency	数据流量 Data flow
二氧化碳与水汽浓度测量仪 Carbon dioxide and vapor analyzer (LI-7500A)	CO ₂ , 水汽浓度与气温等 Carbon dioxide, vapor solution, temperature, etc.	每秒采集20次 Measure frequency at 20 Hz	每秒15 KB 15 KB/s
甲烷浓度测量仪 Methane analyzer (LI-7700)	CH ₄ 浓度 Methane solution in the air	每秒采集50次 Measure frequency at 50 Hz	每秒10 KB 10 KB/s
Gill 风速测量仪 Gill windmaster Pro	三维风速 3-dimensional wind speed	每秒采集10–20次 Measure frequency at 10–20 Hz	每秒5 KB 5 KB/s
CMP3与PQS1辐射测量探头 CMP3 & PQS1 radiation sensor	太阳辐射与光合有效辐射 Solar radiation & photosynthetic active radiation	30分钟采集1次 Measure once per 30 minutes	每天2 KB 2 KB/day
109号土温测量探头 No. 109 soil temperature sensor	多层土壤温度 Multi-layer soil temperature	30分钟采集1次 Measure once per 30 minutes	每天20 KB 20 KB/day
物候观测摄像头 Phenological observation cameras	站点周围物候变化 Phenological change around the station	每天拍摄2次 Two shoots per day	每天6 MB 6 MB/day
复合数据采集器 Integrated data logger (CR5000, Li7550)	收集所有仪器数据 Collection of all observation data	每秒汇总10次 Collection frequency at 10 Hz	每个月4 GB左右 Around 4 GB/month

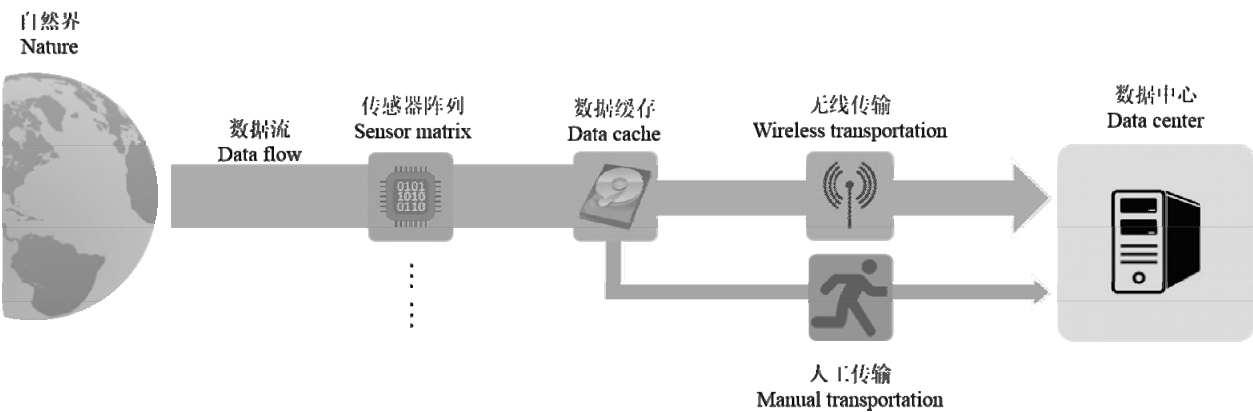


图1 传感器阵列从自然界获取数据示意图。多功能传感器组成的传感阵列拥有极强的数据抓取能力, 使单位时间内获取的数据量剧增, 进而为下游数据操作带来挑战。
Fig. 1 Sketch map for the sensor matrix that acquires data from nature. The sensor matrix contains multifunctional sensors which can efficiently fetch environmental data. For its high data collection speed, this data collection system has brought challenges to the downward data processing works.

(Gantz & Reinsel, 2011, 2012)。

此外, 生态学家们目前还面临一个现实问题: 总体上, 生态学研究产生海量数据, 但尚未形成进行透明数据交换和聚合的文化。集中体现在如下方面: (1)大多数生态学研究由个人调查完成, 绝大多数数据分散在众多的科研工作者手中; (2)生态学有广阔的研究领域, 但不同研究者遵循不同实验规范, 造成生态学数据的异质性; (3)许多生态学研究是在有限的时空尺度上进行的, 离开了其原始的调查环境, 其数据难以理解。除此之外, 生态学数据共享所面临的社会与文化壁垒更是困难重重, 生态学家鲜有分享研究数据的动机。在缺乏合作数据共享激励机制、数据政策和共享文化的背景下, 这些

问题变得更为复杂(赵斌, 2014)。因此, 缺少专用生态系统观测数据库与数据管理制度不完善已经成为该领域数据开放获取的最大障碍(Wallis et al, 2013)。合理规范地存储生态大数据成为当务之急。

2.3 流程化数据处理

除生态观测数据存储落后外, 数据分析技术也未能跟上生态学大数据的脚步。众多新技术被用于生态系统观测数据处理中, 但研究者们却不清楚这些方法的优势与限制所在。虽然观测技术的进步带来了大量宝贵数据, 但生态学家们却仍深陷于信息挖掘困扰中(Barseghian et al, 2010)。因此, 已有的观测数据只有得到妥善处理才能找出数据之间的相关关系或不合常理之处, 最终推动学科发展。该数

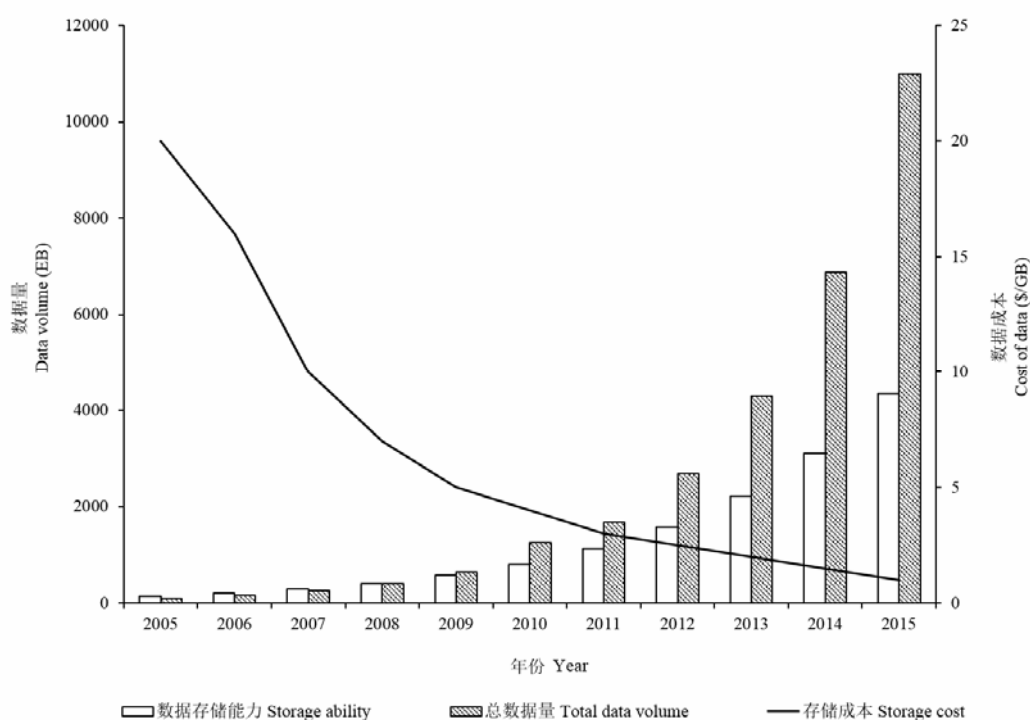


图2 全球数据总量与数据存储能力对比。早在2008年传感器获取的数据就已经大到不能被完全存储(Baraniuk, 2011), 预计2015年总数据量将超越存储能力两倍甚至更多。

Fig. 2 Global total data volume in contrast with the ability of data storage. Early in 2008, the data collected by sensors could not be completely stored (Baraniuk, 2011). It is estimated that the total data volume will be twice bigger than the storage ability in 2015.

据处理过程需要标准化与自动化系统, 例如 Michener 开发的基于 Kepler 工作流的 SEEK 方案, 它可能成为生态学中 workflow 管理与数据整合的标准 (Jones, 2003; Michener et al, 2005)。也有研究者将试验工作室系统 (Experimental Laboratory Systems, ELS) 与自动图像处理系统结合, 使实验室的工作流程标准化 (Pennekamp & Schtickzelle, 2013)。同样, Palazzo 等 (2014) 也开发了针对大量生态学图片、视频数据处理的标准工作流。可见, 在今后大规模生态系统研究中, 形成完善的数据生命循环并建立统一的信息管理制度已迫在眉睫 (Rüegg et al, 2014)。

2.4 多源数据整合分析

从基因尺度跨越到生物圈尺度的复杂生态学问题逐渐凸显, 使得多源数据的整合与分析成为生态系统研究中的重要议题 (Jones et al, 2006)。虽然已有项目着手整合多源数据, 但其主要任务仍是收集并提供异质性数据存储场所 (表2), 暂不具备深层数据挖掘能力。

这些多源数据集需首先经过数据杂质过滤并

精炼为数据产品, 再重编码为可统一分析的语义形式才可使用 (Michener & Jones, 2012)。首先, 进行这项工作需要一个拥有丰富的跨领域知识、并能理解不同类型数据的团队 (Chattopadhyay & Lipson, 2015); 其次, 该工作中处理数据量大, 计算成本高。生态学家们需要与其他领域的研究者联合才能攻克多源数据难关。我们欣喜地看到, 正因为多源数据整合分析以及跨领域合作逐渐成为主流, 共享合作的思想才在生态观测领域不断深化 (Goldman, 2014)。例如, NEON、LTER 等大型观测网络的数据团队正在开发数据预处理、实时分析与可视化综合平台。他们不仅制定出统一的数据格式和数据共享制度, 还着手对异质性数据进行标准化和整合, 最终将多源数据归档至公共数据仓库 (Rüegg et al, 2014)。

3 生态系统观测新方案

解决问题自然离不开新技术, 而新兴信息处理技术正是各行业的关注焦点, 物联网 (Internet of Things, IoT)、无线传感器网络 (Wireless Sensor Networks,

表2 实施多源数据整合的代表性项目
Table 2 Outstanding projects that intended to integrate multiple-source data

项目名称 Project name	数据库状态 Database status			
	数据接口 Data portal	开放获取 Open access	引用规则 Citing rules	整合数据集的关键元数据标签 The key metadata tag of data integration
长期生态学研究网络 The Long Term Ecological Research Network (LTER) https://www.lternet.edu/	单个接口 Single data portal	完全开放 Completely open	引用数据集DOI Cite the DOI of dataset	站点名称, 数据包编号, 地理位置, 发布单位等 Site name, package identifier, spatial location, publisher name, etc.
国家生态学观测网络 The National Ecological Observatory Network (NEON) https://www.neoninc.org/	单个接口 Single data portal	完全开放 Completely open	引用NEON名称 Cite the name of NEON	日期, 站点名称, 行政州名, NEON地域划分, 数据集主题 Date, site name, state name, NEON domain, dataset subject
全球生物多样性信息中心 Global Biodiversity Information Facility (GBIF) https://www.gbif.org/	单个接口 Single data portal	完全开放 Completely open	引用数据集DOI Cite the DOI of dataset	数据集名称, 关键词, 发布单位, 国家等 Dataset name, key words, publisher, country, etc.
全球观测系统信息中心 Global Observing System Information System (GOSIC) https://www.ncdc.noaa.gov/gosic	多个接口 Multiple data portals	部分开放 Partly open	联系数据集发布者 Contact the dataset publisher	子项目名称, 数据集名称 Name of the child project, name of dataset

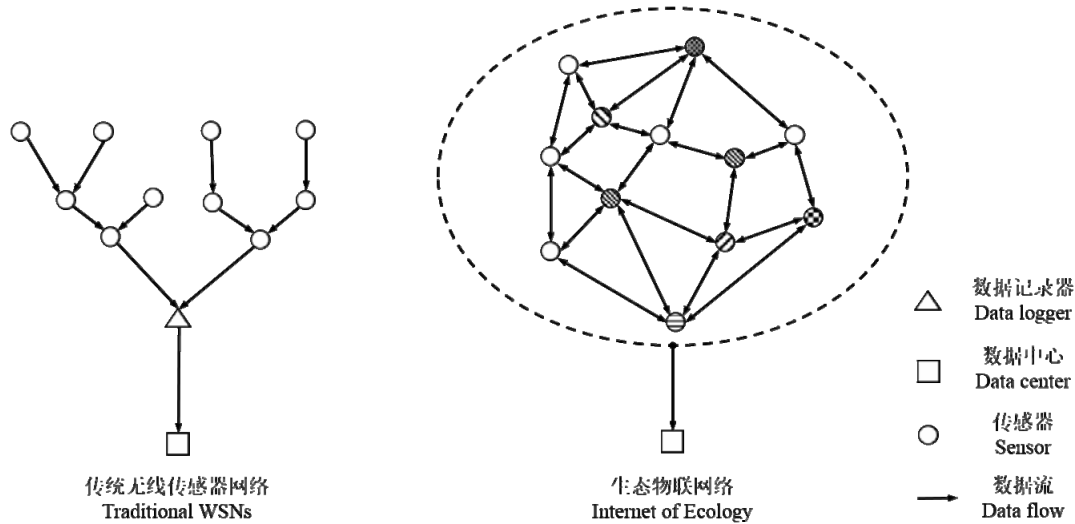


图3 传统无线传感器网络与生态物联网对比, 图中不同类型圆圈代表不同传感器。传统无线传感器网络数据流呈树形分层, 其数据单向流动且传感器功能单一。生态物联网支持传感器间的双向数据交流, 其中物联网传感器具备优秀信息预处理能力与网络信息共享能力, 能够自动调节测量参数并预处理数据。
Fig. 3 Traditional Wireless Sensor Networks (WSNs) and the Internet of Ecology (IoE). In this figure, different types of circles stand for different types of sensors. Traditional WSNs have its data flow in hierarchical tree shape, with single data direction and function sensors inside. The IoE supports mutual communication between the sensors, these internet based sensors are equipped with comprehensive data processing functions, which can share information throughout the network, feedback regulate the parameters in measurement and pre-process the data.

WSNs)、社交网络(Social Networks)等技术已渗透到各行各业并形成了完整的解决方案(Stankovic, 2014)。例如, 美国通用电气公司利用物联网技术助其在能源、制造业、卫生保健、飞机制造等行业保持快速增长, 而IBM与西门子公司则试图将他们的商业分析能力运用在机器与传感器产生的大数据上。下一次工业革命的“燃料”将由数据组成, 数据

驱动理念将永久地改变人类的生产方式。借助这些新理念与新技术, 我们提出了3个生态系统观测发展的方案来应对大数据时代下生态系统观测面临的挑战。

3.1 生态观测物联网

野外观测系统需要应对复杂多变的数据请求, 但传统观测网络在数据传输与标准化等方面都已

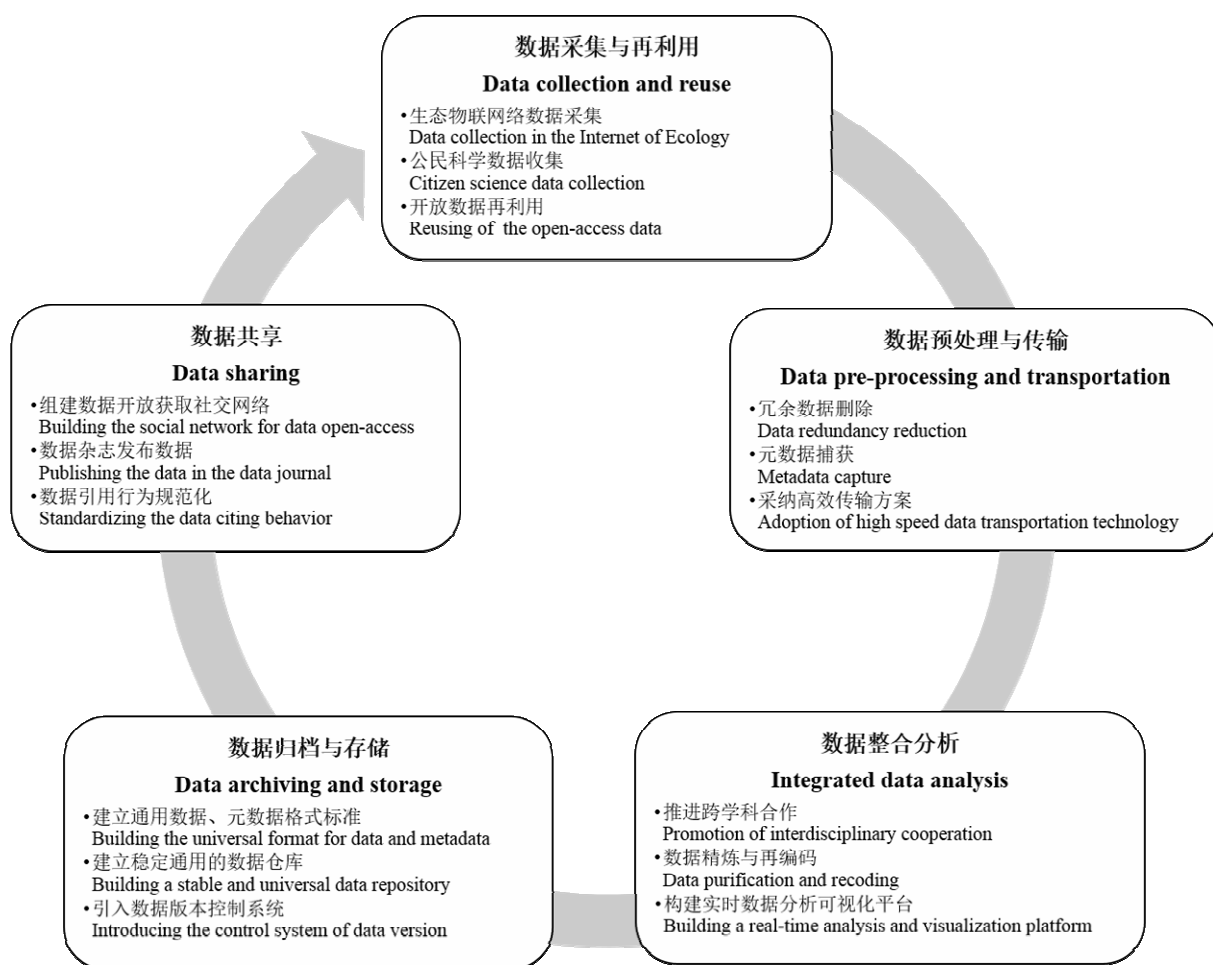


图4 生态系统观测数据生命周期示意图。生态物联网、公民科学、通用数据格式、数据开放获取社交网络与数据版本控制系统等设想的实施可以改变生态观测数据的生命周期, 这将形成有利于生态学发展的良性循环。

Fig. 4 The life cycle of ecosystem observation data. The adoption of the IoE, citizen science, universal data format, social network for data open-access and the control system of data version can improve the life cycle of ecosystem observation data, which also builds a virtuous circle for the development of ecology.

捉襟见肘。因此, 我们提出生态物联网(the Internet of Ecology, IoE)的理念来解决这一问题。该网络借鉴了生物社会网络构造, 即许多生物群形成社会网络进行信息交流, 这种网络关系将众多个体特性赋予整体, 进而提高种群的环境适应性。类似地, 生态物联网中每个传感器都是独立个体, 但其数据交流不是单向层级式、指令式交流, 而是网络化的双向数据交流。每个传感器执行的决策都是整个物联网权衡后的最优决策, 因此该网络是具备集体智能的自适应观测网络(图3)。

在该设计中, 生态物联网解决了数据标准化与数据传输问题。首先, 异质性生态学数据的标准化是生态学观测中的研究热点。代表成果有基于

XML架构的生态学元数据语言(ecological metadata language, EML), 它协助整合并标准化处理生态学元数据, 以及用于标准化生物多样性数据的达尔文核心(Darwin Core)数据标准等。因此, 为了更有效地进行内外部数据交流, 生态物联网中不同传感器间数据交流也利用标准化元数据语言实现统一。其次, 得益于数据交流双向化与灵活的传感器配置, 该生态物联网中摒弃了原本用于数据收集与传输的中转节点, 网络中每个传感器都能与数据中心进行直接数据交流。如此, 该网络将数据传输资源进行动态分配, 避免了集中传输带来的瓶颈。

3.2 公民科学

在非科研热点区域中存在大量有价值数据, 可

惜这些区域往往没有被自动化传感器网络覆盖。因此,选择借助公众力量来收集这些难以获取的数据(Tulloch et al, 2013),进而整合这些重要信息来改进已有研究结果(Turner et al, 2015)。近年来有些新研究项目开始将注意力转向大范围的潜在科研参与者,虽然他们并不是训练有素的科学家,但依旧可以借助他们的力量来收集、分类、转化或分析数据,此类项目被称为公民科学项目(citizen science)(Bonney et al, 2014)。

利用公民科学进行生态系统观测是一种革新:首先,公民科学是传感器网络的扩展。例如iSPEX项目将专业传感器改造为便携的手机附属装置,用于测定和收集大气中的气溶胶数据(de Boer, 2012),而美国国家极端风暴实验室(NSSL)团队开发的mPING手机应用则通过手机内置传感器获取信息,并将数据分析结果以天气预报的形式反馈给用户(NSSL, 2013)。其次,公民科学又是大数据时代科学观测的一种新方式,可用于收集种群分布、种群大小、物种存在性与物种丰富度等数据(Barlow et al, 2015)。例如,创始于2003年的eBird项目最早致力于统计鸟类分布与丰富度,现已从公民科学项目发展为成熟的数据产业平台(Sullivan et al, 2014),不仅加深了生态学家对鸟类种群分布动态的理解,同时还提高了民众对鸟类及其栖息地的保护意识。类似的项目还有哥伦比亚大学与马里兰大学联合开发的Leafsnap,该软件利用机器识别并辅以专家人工鉴定,为用户提供近乎实时的植物识别、分类和高分辨率图像下载服务并实现了大空间、长时间尺度的植物多样性数据收集(Belhumeur et al, 2011)。

3.3 通用数据平台

生态系统观测面临多方面挑战。客观上看,生态观测数据本身具有高异质性与低通用性;主观上讲,生态学家们仍然对数据结果的第一发布权感到担忧,同时由于缺乏数据引用规范,生态系统观测的发展面临巨大障碍(Wallis et al, 2013)。因此,为了构建良性的生态系统观测数据生命循环,应充分利用现代数据传输与处理中已发展出的高带宽、低延迟等优势,设计更为通用的观测数据平台,使之成为集科研社交网络、科研文献网络、数据仓库与数据版本控制平台(如Github)为一体的有机整体(图4)。该平台首先基于社交网络,为用户提供双向且自由的信息交流服务,着重加强观测同行间的经验

交流,建立共享互信的良好领域氛围。其次,该平台采用版本控制方式来管理众多异质性的观测数据集。该系统不仅可督促生态学家们对自己的数据集进行规范化存储与管理,还可确保在数据生产者拥有知识产权的前提下便于规范化的数据下载与引用,这样可形成双赢局面来构建良性数据生命循环。最后,该平台还应提供科研信息提取与处理服务,通过富媒体结构化、复杂数据检索与云计算等技术挖掘新的生态学数据,进而用于整合分析或生态计量学研究等。

综上所述,我们认为“为生态学研究提供一站式的生态系统观测服务”这一理念才是大数据时代生态系统观测发展的大势所趋。

参考文献

- Anthoni PM, Unsworth MH, Law BE, Irvine J, Baldocchi DD, van Tuyl S, Moore D (2002) Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems. *Agricultural and Forest Meteorology*, 111, 203–222.
- Baldocchi DD, Falge E, Gu L, Olson R, Hollinger D, Running S, Anthoni P, Bernhofer C, Davis K, Evans R, Fuentes J, Goldstein A, Katul G, Law B, Lee X, Malhi Y, Meyers T, Munger W, Oechel W, Paw KT, Pilegaard K, Schmid HP, Valentini R, Verma S, Vesala T, Wilson K, Wofsy S (2001) FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82, 2415–2434.
- Baldocchi DD (2014) Measuring fluxes of trace gases and energy between ecosystems and the atmosphere—the state and future of the eddy covariance method. *Global Change Biology*, 20, 3600–3609.
- Baraniuk RG (2011) More is less: signal processing and the data deluge. *Science*, 331, 717–719.
- Barlow KE, Briggs PA, Haysom KA, Hutson AM, Lechiara NL, Racey PA, Walsh AL, Langton SD (2015) Citizen science reveals trends in bat populations: the National Bat Monitoring Programme in Great Britain. *Biological Conservation*, 182, 14–26.
- Barseghian D, Altintas I, Jones MB, Crawl D, Potter N, Gallagher J, Cornillon P, Schildhauer M, Borer ET, Seabloom EW (2010) Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics*, 5, 42–50.
- Belhumeur P, Jacobs D, Kress J (2011) Introduction to Leafsnap. <http://leafsnap.com/> (accessed 2015-6-15)
- Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Next steps for citizen science. *Science*, 343, 1436–1437.
- Butler D (2014) Earth observation enters next phase. *Nature*,

- 508, 160–161.
- Chattopadhyay I, Lipson H (2015) Data smashing: uncovering lurking order in data. *Journal of the Royal Society*, 11, 1–11.
- Collins SL, Bettencourt LMA, Hagberg A, Brown RF, Moore DI, Bonito G, Delin KA, Jackson SP, Johnson DW, Burleigh SC, Woodrow RR, McAuley JM (2006) New opportunities in ecological sensing using wireless sensor networks. *Frontiers in Ecology and the Environment*, 4, 402–407.
- Conant R (2006) Wireless sensor networks: driving the new industrial revolution. *Industrial Embedded Systems*, 1, 8–11.
- de Boer J (2012) Introduction to iSPEX. <http://ispek.nl/en/> (accessed 2015-6-15)
- Elmendorf SC, Moore KA (2008) Use of community composition data to predict the fecundity and abundance of species. *Conservation Biology*, 22, 1523–1532.
- Field C, DeFries R, Foster D, Grove M, Jackson R, Law B, Lodge D, Peters D, Schimel D (2006) Integrated science and education plan for the National Ecological Observatory Network. <http://www.neoninc.org/> (accessed 2015-6-15)
- Gantz J, Reinsel D (2011) Extracting value from chaos. *International Data Corporation Interview*, 11, 9–10.
- Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. *International Data Corporation iView: IDC Analyze the Future*, 20, 1–16.
- Gillespie TW, Foody GM, Rocchini D, Giorgi AP, Saatchi S (2008) Measuring and modelling biodiversity from space. *Progress in Physical Geography*, 32, 203–221.
- Goldman SL (2014) Reinventing discovery: the new era of networked science. *The European Legacy*, 19, 392–393.
- Gong P (2010) Progress in recent environmental applications of wireless sensor networks. *Journal of Remote Sensing*, 14, 387–395. (in Chinese with English abstract) [宫鹏 (2010) 无线传感器网络技术环境应用进展. *遥感学报*, 14, 387–395.]
- Goring SJ, Utz RM, Weathers KC, Dodds WK, Soranno PA, Sweet LCC, Kominoski JS, Rüegg J, Thorn AM (2014) Improving the culture of interdisciplinary collaboration in ecology by expanding measures of success. *Frontiers in Ecology and the Environment*, 12, 39–47.
- 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.
- Hart JK, Martinez K (2006) Environmental sensor networks: a revolution in the earth system science? *Earth-Science Reviews*, 78, 177–191.
- Heffernan JB, Xiao J, Harms TK, Goring SJ, Koenig LE, McDowell WHP, Richardson AD, Stow CA, Vargas R, Weathers KC, Soranno PA, Angilletta MJ, Buckley LB, Gruner DS, Keitt TH, Kellner JR, Kominoski JS, Rocha AV (2014) Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, 12, 5–14.
- Horning N, Robinson J, Sterling E, Turner W, Spector S (2010) *Remote Sensing for Ecology and Conservation*. Oxford University Press, London.
- Huang C, Kim S, Song K, Townshend JR, Davis P, Altstatt A, Rodas O, Yanosky A, Clay R, Tucker CJ (2009) Assessment of Paraguay's forest cover change using Landsat observations. *Global and Planetary Change*, 67, 1–12.
- Hyypä J, Hyypä H, Inkinen M, Engdahl M, Linko S, Zhu Y (2000) Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, 128, 109–120.
- 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.
- Jones MB (2003) SEEK EcoGrid: Integrating Data and Computational Resources for Ecology. <http://datatabs.lternet.edu/spring-2003/seek-ecogrid-integrating-data-and-computational-resources-ecology> (accessed 2015-6-15)
- Jones N (2014) Mini satellites prove their scientific power. *Nature*, 508, 300–301.
- Jung M, Reichstein M, Bondeau A (2009) Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, 6, 2001–2013.
- Kampe TU, Johnson BR, Kuester M, Keller M (2010) NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing*, 4, 510–524.
- Mairota P, Cafarelli B, Labadessa R, Lovergine FP, Tarantino C, Nagendra H, Didham RK (2015) Very High Resolution Earth Observation features for testing the direct and indirect effects of landscape structure on local habitat quality. *International Journal of Applied Earth Observation and Geoinformation*, 34, 96–102.
- Melbourne BA, Cornell HV, Davies KF, Dugaw CJ, Elmendorf S, Freestone AL, Hall RJ, Harrison S, Hastings A, Holland M (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? *Ecology Letters*, 10, 77–94.
- Michener WK, Beach J, Bowers S, Downey L, Jones M, Ludäscher B, Pennington D, Rajasekar A, Romanello S, Schildhauer M (2005) Data integration and workflow solutions for ecology. In: *Data Integration in the Life Sciences* (ed. Bodenreider O), pp. 321–324. Springer, Berlin.
- Michener WK, Jones MB (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, 27, 85–93.
- National Ecological Observatory Network (2011) History of National Ecological Observatory Network. <http://www.neoninc.org/about/history> (accessed 2015-6-15)
- NSSL (2013) Introduction to mPING. <http://mping.nssl.noaa.gov/> (accessed 2015-6-15)
- Palazzo S, Spampinato C, Giordano D (2014) Large scale data processing in ecology: a case study on long-term underwater video monitoring. In: *Parallel, Distributed and Network-Based Processing (PDP)*, 22nd Euromicro International Conference, pp. 312–316. IEEE Computer Society,

- Washington.
- Pennekamp F, Schtickzelle N (2013) Implementing image analysis in laboratory-based experimental systems for ecology and evolution: a hands-on guide. *Methods in Ecology and Evolution*, 4, 483–492.
- Porter JH, Arzberger P, Braun H, Bryant P, Gage S, Hansen T, Hanson P, Lin C, Lin F, Kratz T, Michener W, Shapiro S, Williams T (2005) Wireless sensor networks for ecology. *BioScience*, 55, 561–572.
- Porter JH, Hanson PC, Lin C (2012) Staying afloat in the sensor data deluge. *Trends in Ecology & Evolution*, 27, 121–129.
- Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and opportunities of open data in Ecology. *Science*, 331, 703–705.
- Rüegg J, Gries C, Bond-Lamberty B, Bowen GJ, Felzer BS, McIntyre NE, Soranno PA, Vanderbilt KL, Weathers KC (2014) Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, 12, 24–30.
- Rundel PW, Graham EA, Allen MF, Fisher JC, Harmon TC (2009) Environmental sensor networks in ecological research. *New Phytologist*, 182, 589–607.
- Schatz G (2014) The faces of Big Science. *Nature Reviews Molecular Cell Biology*, 15, 423–426.
- Shamoun-Baranes J, Chapman JW, Alves JA, Bauer S, Dokter AM, Hüppop OKJ, Jarmo K, Leijnse H, Liechti F, van Gasteren H (2014) Continental-scale radar monitoring of the aerial movements of animals. *Movement Ecology*, 2, 9.
- Stankovic JA (2014) Research directions for the internet of things. *IEEE Internet of Things Journal*, 1, 3–9.
- Sullivan BL, Aycrigg JL, Barry JH, Bonney RE, Bruns N, Cooper CB, Damoulas T, Dhondt AA, Dietterich T, Farnsworth A, Fink D, Fitzpatrick JW, Fredericks T, Gerbracht J, Gomes C, Hochachka WM, Iliff MJ, Lagoze C, La Sorte FA, Merrifield M, Morris W, Phillips TB, Reynolds M, Rodewald AD, Rosenberg KV, Trautmann NM, Wiggins A, Winkler DW, Wong W, Wood CL, Yu J, Kelling S (2014) The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31–40.
- Tulloch AIT, Possingham HP, Joseph LN, Szabo J, Martin TG (2013) Realising the full potential of citizen science monitoring programs. *Biological Conservation*, 165, 128–138.
- Turner W (2014) Sensing biodiversity. *Science*, 346, 301–302.
- Turner W, Rondinini C, Pettorelli N, Mora B, Leidner AK, Szantoi Z, Buchanan G, Dech S, Dwyer J, Herold M, Koh LP, Leimgruber P, Taubenboeck H, Wegmann M, Wikelski M, Woodcock C (2015) Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation*, 182, 173–176.
- Velicogna I, Wahr J (2006) Measurements of time-variable gravity show mass loss in Antarctica. *Science*, 311, 1754–1756.
- Wallis JC, Rolando E, Borgman CL (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8, e67332.
- Wang Y, Liu XN, Ju XH (2007) The difference and relativity between rainfall by automatic recording and manual observation. *Journal of Applied Meteorological Science*, 18, 412–413. (in Chinese with English abstract) [王颖, 刘小宁, 鞠晓慧 (2007) 自动观测与人工观测差异的初步分析, *应用气象学报*, 18, 412–413.]
- Watt KE (2013) *Systems Analysis in Ecology*. Elsevier, Amsterdam.
- Willis KS (2015) Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, 182, 233–242.
- Zhao B (2014) Ecology should evolve to facing global problems. *Science and Technology Review*, 32(12), 12. (in Chinese) [赵斌 (2014) 生态学必须进化以应对全球性重大问题. *科技导报*, 32(12), 12.]
- Zulueta RC, Oechel WC, Loescher HW, Lawrence WT (2011) Aircraft-derived regional scale CO₂ fluxes from vegetated drained thaw-lake basins and interstitial tundra on the Arctic coastal plain of Alaska. *Global Change Biology*, 17, 2781–2802.

(责任编辑: 纪力强 责任编辑: 闫文杰)