“绿途”系统: 公民科学时代的植被调查制图新工具

金时超^{1,2,3}, 胡天宇^{1,2}, 苏艳军^{1,2}, 马勤^{1,2}, 关宏灿^{1,2}, 杨默含^{1,2}, 郭庆华^{1,2*}

1. 中国科学院植物研究所, 植被与环境变化国家重点实验室, 北京 100093;

2. 中国科学院大学, 北京 100049;

3. 南京农业大学作物表型组学交叉研究中心, 南京 210095

* 联系人, E-mail: qguo@ibcas.ac.cn

收稿日期: 2020-06-23; 接受日期: 2020-08-27; 网络版发表日期: 2020-12-24

中国科学院战略性先导科技专项(A类)“地球大数据科学专项”(批准号: XDA19050401)资助

摘要 植被图是了解植被空间分布状态和地理格局的地图, 是自然资源研究和管理决策的重要辅助信息源. 植被制图是以植被调查数据为基础, 辅以3S(遥感、地理信息系统、全球定位系统)技术, 实现连续区域的植被制图. 尽管3S技术得到了快速的发展, 但是获取海量的植被群系分布信息依然是大尺度植被制图至关重要的前提, 并成为当前植被制图的关键瓶颈. 随着智能终端的流行和公民科学的兴起, 以个人遥感手段为核心的众包模式为植被群系调查带来了新的发展机遇. 本研究开发了一套基于智能手机和网页端服务的植被类型众源采集与分析平台——“绿途”. “绿途”系统的手机端支持安卓和iOS系统, 主要功能包括数据收集、图层管理、兴趣点管理、轨迹管理、用户信息、数据展示以及其他辅助功能, 能高效地在野外采集带有地理位置信息的照片和点位记录. “绿途”系统的网页端具备与手机端相应的管理功能, 还包括对数据的分级审核功能. 2018年7月至今, “绿途”系统已累计采集数据记录108083条, 分布于全国的34个省市自治区, 覆盖7种植被型组和582个植被群系类型. 该系统单日上传的最大数据量为5358条、单人单日上传最大数据总量为2543条. “绿途”系统采集的数据作为地面调查样本被成功应用于1:100万中国植被图的更新, 并正在服务于1:50万中国植被图的绘制, 有望促进更多植被生态领域的科学研究和管理政策制定.

关键词 绿途, 植被图, 众包, 公民科学, APP, 网页端

植被是陆地生态系统的重要组成部分, 具有涵养水源、防风固沙和调节气候等功能. 植被图是反映特定(自然或行政)地理区域内, 不同分类等级的所有植被类型及其空间范围和分布格局的地图, 对于生态学研究、生物多样性保护和自然资源决策管理具有重要意义^[1,2]. 随着我国国民经济的快速发展, 植被的分布格局发生了巨大的变化. 目前, 我国使用的植被图本底

数据《中华人民共和国植被图(1:1000000)》是20世纪80年代起由逾250位生态学家耗时20余年制作完成的^[3]. 由于传统的制图方法和可用的遥感数据源有限, 加上过去几十年植被类型本身随着自然和人为活动的影响发生了明显的变化, 上一代植被图包含的植被群系类别和边界都与现实明显不符^[4-9](这里的群系指的是植被型或植被亚型分类级别以下的中级分类单位,

引用格式: 金时超, 胡天宇, 苏艳军, 等. “绿途”系统: 公民科学时代的植被调查制图新工具. 中国科学: 生命科学, 2021, 51: 362–374

Jin S C, Hu T Y, Su Y J, et al. LiVegetation: an investigative tool for vegetation mapping in the era of citizen science (in Chinese). Sci Sin Vitae, 2021, 51: 362–374, doi: [10.1360/SSV-2020-0209](https://doi.org/10.1360/SSV-2020-0209)

指在植物群落优势层中具有相同建群种或共建种的植物群丛的集合^[10,11]。植被图的更新和高精度绘制成为迫切需求。

植被类型调查是植被制图的前提和难点。传统的植被制图主要通过实地调查和文献调查获得基础数据,采用专家知识与高分辨率航空影像人工解译相结合实现^[12-14]。这种方法在大尺度制图过程中的时间成本和人力成本都过高。遥感技术的出现为大尺度植被制图带来了新的方式,即融合多源遥感数据与地面调查数据实现大尺度植被制图^[15,16]。但是,由于植被制图通常是精细到群系甚至更低的级别,大尺度植被制图还存在诸多难点:(i) 遥感数据存在同物异谱和同谱异物等问题,对于群系级别的植被制图能力有限;(ii) 大尺度植被制图需要海量的现实性地面调查样本作为基础;(iii) 大尺度地面调查费时、费力、效率低,难以由少数人完成。

公民科学,也叫公众科学、众包,是指公众参与的科学研究^[17]。公民科学的出现为科学、社会和政策制定等领域带来了新的潜力^[18],开创了公民参与科学研究的新范式^[17,19],尤其是在生态环境保护等领域发挥着重要的知识获取作用^[18,20,21]。以手机等智能终端为媒介、以APP(应用程序)为记录工具的方式是目前公

民科学开展的主要方式之一^[20]。这种方式模糊了专业科学家与大众的边界,降低了数据采集的门槛,有效地提高了数据采集的效率和数据的代表性。

除了使用微博、Twitter和Facebook等社交工具记录的位置数据进行科学研究以外^[22],国内外开始出现了为科学研究定制化的数据采集系统(表1)。例如在生态科学领域,中国科学院植物研究所开发的用于植物分类调查的“标本伴侣”、中国科学院动物研究所开发的用于野生生物观测与调查的“生物记”、中国科学院昆明植物研究所标本馆开发的“Biotracks”、美国康奈尔大学开发的用于鸟类分布调查的“eBird”以及加州科学院和美国国家地理学会开发的用于生物多样性数据记录的“iNaturalist”。这些定制化的众源数据采集系统积累了大量的数据,辅助了科学研究的开展^[20]。此外,随着数据的积累和人工智能算法的进步^[23],还出现了一系列为物种识别任务开发的APP(如“花伴侣”“形色”“百度识图”和“花帮主”),它们也同时收集了大量的数据。但是,因为上述系统采集数据的目的不同,其积累的数据无法直接应用于植被制图任务。同时,上述众源APP采集的数据多为管理员所拥有和使用,对于数据的共享使用程度还有待提升。因此,开发适用于大尺度植被类型众源采集与分析的工具有利于

表 1 部分代表性的众源生态数据采集系统

Table 1 Some representative crowdsourcing systems for ecological data collection

名称	用途	开发者/作者	时间	数据量	资料来源
绿途	植被调查	中国科学院植物研究所	2018年至今	108888条	https://survey.chinavegetation.cn/survey
标本伴侣	植物分类野外调查	中国科学院植物研究所	2018年	未知	http://www.iplant.cn/app1/
生物记	野生生物观测与调查	中国科学院动物研究所	无	未知	http://nol.especies.cn/
Biotracks	生物多样性调查	中国科学院昆明植物研究所标本馆	2016年	1414444+条	http://www.biotracks.cn/
eBird	鸟类分布调查	The Cornell Lab of Ornithology	2002年	未知	https://ebird.org/home
iNaturalist	植物和动物识别	The California Academy of Sciences and the National Geographic Society	2011年	40943650条	https://www.inaturalist.org/
Pl@ntNet	植物识别	Agropolis Foundation 等	2014年	3058367条	https://plantnet.org/
Leafsnap	树叶识别	哥伦比亚大学、马里兰大学和史密森学会	无	23147室内照片和7719张室外照片	http://leafsnap.com/
花伴侣	花卉识别	中国科学院植物研究所	2016年	约5000种	http://hbl.nongbangzhu.cn/
形色	植物识别	杭州睿琪软件有限公司	2018年	约4000种	http://www.xingseapp.com/
百度识图	图片识别	百度公司	2010年	未知	http://image.baidu.com/?fr=shitu
花帮主	养花人的大本营	北京联星时代科技有限公司	2015年	未知	http://www.nongbangzhu.cn/portal/index.html

弥补现有研究的不足,对于帮助大尺度植被图的野外数据采集和提升制图效率具有重要意义。

本研究旨在开发一套适用于大尺度的众源植被调查系统(“绿途”),用于植被图的绘制。“绿途”系统包括手机端数据采集APP和网页端数据审核和管理系统。“绿途”APP的核心是服务于野外数据采集,主要通过拍照采集和点位采集两种途径实现,分别实现带有经纬度属性的植被群落照片和文本记录信息的采集(详见1.3节)。此外,设计的辅助功能包括图层管理、兴趣点管理、轨迹管理和用户管理。“绿途”网页端的核心功能是数据审核、数据管理以及野外调查的规划,同时具有手机端同样的辅助功能。

1 系统设计与核心功能

“绿途”系统架构设计主要包括数据存储设计、服务端设计、APP设计和网页端设计。

1.1 数据存储设计

“绿途”数据种类丰富,包含空间几何数据、业务表单数据、文档数据,“绿途”系统主要使用两种方式存储所收集到的数据,分别是数据库存储和文件存储。

数据库采用目前最先进的开源的对象-关系型数据库PostgreSQL^[24,25],它有丰富的编程接口,支持广泛的数据类型、SQL标准、SQL查询以及多种数据源的外部数据表等多种属性。数据库存储主要用于存储点位数据和业务逻辑等信息。为了提高数据存储效率、扩展性和高可用性等,数据库设计时综合分析了关系数据库数据强一致性的优点和非关系型数据库易扩展、数据结构灵活、易扩容等优点,最终采用了两者相结合的方式存储^[26],具体详见数据库结构设计图(网络版附图1)。

文件存储主要存储用户采集的照片数据。文件存储采用云端可扩容式存储方式,按需动态扩容。采用文件存储的照片数据有三种状态:未审核、已初审和已终审(图1)。照片数据上传时为未审核数据,经过相应的专家用户进行数据审核之后变成初审数据,初审状态的数据经过终审变成终审数据。相比之下,采用数据库存储的点位数据默认为已终审数据。

“绿途”系统的数据库中存储的关系型信息主要包括:用户与数据的关系、用户与归属地之间的关系(图

1)。这些关系同时与用户类型对应的用户权限相关。本系统对于数据的收集、处理、审核、下载等对不同用户提供了不同的权限类型,并设置不同的默认权限组与之对应。同时给出调整用户权限的接口,管理员可通过权限的自由组合分配不同权限给其他用户,这样保证了权限系统的灵活性以及鲁棒性。

“绿途”系统的用户类型包括普通用户、初审专家、终审专家和系统管理员(图1)。普通用户是用户自行注册申请的基本用户角色。普通用户的归属地是变化的,由其采集数据时的经纬度位置确定;普通用户与数据的关系是对所有数据有查看权限、只对自己采集的数据有编辑和使用权限。初审专家是普通用户经过申请实现权限提升的一种用户角色(图1)。具体而言,初审专家的归属地由其申请成为专家用户时填写,初审专家对所有数据具有查看权限、对所在省份的未审核数据具有初级审核的权限、对自己采集的数据和审核后的数据具有编辑和使用权限。终审专家是管理员创建的、对某一行政区域具有最高数据操作权限的用户角色。终审专家的归属地由管理员分配,终审专家对所有数据具有查看权限、对所在省份的数据具有初级审核和终极审核权限、对自己采集的数据和所在省份他人数据具有编辑和使用权限。在终审专家级别,系统管理员可指定一类专家具有对全国所有省份的数据进行初级审核和终极审核的权限,但是只对自己上传过的数据和审核过的数据具有下载使用权限。系统管理员是“绿途”系统中唯一的超级用户(root用户),具备上述用户的所有功能,同时可以修改用户的权限和关系类型。

“绿途”系统在数据库中存储有辅助调查作业的图层数据、兴趣点记录、轨迹记录 and 用户个人信息等。其中,“绿途”系统自行发布的图层数据包括用于1:50万植被图制图的全国矢量斑块和1:100万植被图数据。为解决数据展示浏览问题,本研究根据国家基本比例尺最新标准在1:25万比例尺的分幅标准下对数据进行分幅,将其均匀分成700多个矢量块,并使用Geo-Server分别发布图层^[27]。当客户端请求展示图层时,系统根据请求展示GPS坐标,计算1:25万比例尺分幅标准下展示窗口所涉及分幅,读取相应文件实时渲染图片并返回前端。兴趣点记录、轨迹记录 and 用户个人信息等数据为用户在实际使用过程中创建的个人数据。

上述文件存储、数据库关系(用户有数据关系、用户与归属地关系)以及数据库辅助图层和功能服务

图 1 “绿途”关系数据表及主要功能服务设计(网络版彩图)

Figure 1 Design of relationship databases and main functions of “LiVegetation” system (color online)

设计为后续APP设计和Web设计提供了应用接口。

1.2 服务端设计

本系统的服务端基于当前最流行的Linux容器解决方案Docker搭建,具有简单易用的特点^[28]。使用面向性能设计的HTTP服务器Nginx作为网络服务器,其具有并发能力强、稳定性高等特点^[29]。采用uWSGI作为代理网关,将HTTP协议转化成语言支持的网络协议。使用Python语言编写的轻量级网页端应用框架Flask实现服务编写,实现应用程序与服务器之间的信息传输^[30]。

1.3 APP设计与主要功能

“绿途”APP被设计为支持iOS和Android两种平台,适应中英文双语系统。为保证账户密码安全和数据安全,用户需要通过手机号注册账号,获取与手机号绑定的账号密码。已注册登录的账户,可绑定第三方登录软件账户(如QQ账户和微信账户)实现快捷登录。此外,为方便用户体验和离线模式使用,APP还支持游客

登录模式。目前,“绿途”APP在AppStore、小米应用商店、华为应用商店和应用宝均可下载。

APP主要功能包括数据收集、图层管理、兴趣点管理、轨迹管理、用户管理、数据展示以及其他辅助功能(图2)。数据收集是APP的核心功能模块,主要有拍照采集功能和点位采集功能。拍照采集功能指引用户上传三张照片,包括群落全景、优势种整体和优势种细节。照片可以通过直接拍照或从用户相册选择,系统能够自动获取已有照片的经纬度信息或当前位置经纬度作为本次拍照记录的水平位置属性,并获取手机自身记录的海拔信息或根据经纬度获取记录所处的海拔信息。照片记录的属性还包括全国群系名称和其他信息,这些属性支持快速模糊匹配(网络版附图2)。相比之下,点位采集不需要进行拍照操作,效率大大提升。点位采集是用户通过判断周边植被的类型,在APP界面的地图上用图钉标记所见植被的位置,图钉所在位置的经纬度和海拔作为该记录的位置信息(网络版附图3)。由于没有照片用于后期的群系类型的审核,因此点位数据在保存时必须包含与系统内置群系系统匹配的群系类型名称,对于无法匹配的特殊类型,用户可

图2 “绿途”APP主要功能设计(网络版彩图)

Figure 2 Main function of “LiVegetation” mobile application (color online)

以上传自定义群系系统,同时可以在其他信息中进行补充记录。照片数据和点位数据都支持离线情况下的本地保存。本地保存的数据可以进行编辑和删除等操作,在有网络情况下进行数据上传,这一设计主要是为了解决野外调查时,手机端容易出现无网络状态的问题。

图层管理功能主要是为了辅助用户开展野外调查,通过不同的图层方便用户查看地物类型和调查参考边界等信息。“绿途”APP图层包括2类:共有图层和自定义图层。共有图层是系统内置的,包括高德影像、谷歌影像、1:100万植被图群系类型图和1:50万植被斑块边界图;自定义图层可以联系“绿途”开发团队进行发布,也可以在“绿途”网页端上传斑块图层实现发布,前者适用于大尺度图层发布,后者适用于小尺度。自定义图层与用户绑定,实现在APP和网页端的自动同步。为了在离线模式下可以正常加载使用上述图层,该系统设计了所有图层的离线下下载功能。

兴趣点管理功能主要是为了满足用户在野外调查时或者规划调查时可以对特定区域进行标记。兴趣点包括经纬位置和用户定义的标签、图标、颜色等属性信息。兴趣点设计为本地和云端两种状态,可以相互同

步。兴趣点支持修改、导航、分享、同步和删除操作。

轨迹管理功能主要是为了满足用户在野外调查时可以记录调查路线。记录的轨迹包括轨迹的矢量数据信息和用户定义的属性信息。轨迹设计为本地和云端两种状态,可以相互同步。轨迹支持修改、导航、分享、同步和删除操作。

用户管理包括查看用户角色、修改个人信息(昵称等),查看个人贡献、轨迹和兴趣点,管理本地数据上传,查看排行榜、消息中心,提交帮助反馈,分享APP和申请专家权限。

数据展示主要集中在首页地图页面,照片数据以照片缩略图显示,点击之后可以放大查看,并显示数据群系名称和上传用户。点位数据以系统设计的文本框缩略图显示,点击之后无照片详情,显示群系名称和上传用户。此外,首页上还可以加载用户创建的兴趣点和轨迹,其显示的图标、颜色、形状等读取自用户保存时的设置。

其他辅助功能主要包括“扫一扫”登录网页端和设置中的配置功能。用户设置中可以进行第三方账号绑定、修改密码、保存图片 and 点位记录到本地、自定义群系匹配、地图默认缩放级别设置、检查更新和语言设置等。其中,保存图片 and 点位记录到本地是为了防止数据在特殊情况下丢失,这是将数据保存到手机本地文件系统的一种特殊设计。自定义群系匹配是为了满足用户新增自定义群系系统用于群系模糊匹配的需求而设计的功能。

1.4 Web设计与核心功能

Web选用的前端框架是目前广泛使用的Vue.js^[31]。Vue是一套轻量的用于构建用户界面的渐进式框架,可以自底向上逐层应用。Vue采用组件化开发,相同的组件可以复用,易于开发以及后期维护。同时,Vue提供了MVVM(Model-View-ViewModel)开发模式,支持View和ViewModel的双向数据绑定,这使得ViewModel的状态改变可以自动传递给View,从而保证视图和数据的一致性,使前端的开发更加高效便捷。

Web支持账号登录和手机扫码登录(<https://survey.chinavegetation.cn/survey>),主要功能包括数据管理、兴趣点管理、轨迹管理、斑块管理、底图管理、用户管理、数据展示以及其他辅助功能(图3)。

数据管理是Web的核心功能模块,主要包括数据

图 3 “绿途”网页端主要功能设计(网络版彩图)

Figure 3 Main function of “LiVegetation” Web system (color online)

上传、数据审核和数据导出。数据上传包括照片数据上传和表格数据上传。照片数据上传支持单条照片记录上传和多条照片记录的批量上传。多条照片记录批量上传时需要保证每条记录的照片数目相同,且照片记录是按名称顺序排列。表格数据的上传类似于点位记录批量上传,根据点位记录获取的设计,只接受与内置群系系统匹配的数据,因此在上传前需要按照上传页面提供的模板对拟上传数据进行标准化处理。数据审核主要是对未审核的照片数据进行初审和终审,这一过程有明显的权限要求,详见1.1节中数据状态和用户权限设计。数据导出可以按照时间、地点和上传人进行筛选导出。导出结果可以选择只导出数据记录表和同时导出数据记录表和对应的照片。数据记录表的信息包括经度、纬度、水平定位误差、数据时间、用户描述、植被型组、植被型、植被亚型、群系组、

群系、审核专家、审核状态、数据类型、上传用户类型、上传用户名、所在省份和图片编号。其中,图片编号用于对应表格记录数据与照片详情。

兴趣点管理主要是管理存储在云端的兴趣点。网页端支持兴趣点的创建、查看、修改、分享和删除等功能。网页端创建的兴趣点保存在云端,手机端可以同步至本地使用。

轨迹管理主要是为了管理用户在野外调查时的调查路线,与手机端轨迹管理对应。网页端支持轨迹的创建,用于规划野外调查路径(创建后的轨迹保存在云端,可以进行管理并同步至手机端的本地使用),同时也支持对手机端本地同步过来的轨迹进行管理。轨迹管理的内容包括修改、分享和删除等。

斑块管理对应于手机端自定义图层的创建。网页端支持斑块的在线创建和本地上传,创建/上传后的轨

迹保存在云端,可以进行管理和同步至手机端的本地。

底图管理、用户管理、数据展示和其他辅助功能与手机端类似。此外,地图首页还设计有快捷的交互式操作,包括数据上传、绘制兴趣点、绘制轨迹、绘制斑块、选择地图和数据审核。

2 “绿途”系统采集数据现状

自2018年7月投入使用至数据统计时间(2020年6月13日)，“绿途”系统已经累计注册1010位用户，其中管理员1人，全国专家1人，终审专家34人、初审专家50人，普通用户924人。累计采集数据记录为108083条，其中部分专家通过“绿途”网页端对数据开展了审核工作，最终“绿途”平台包含已审核数据98606条，未审核数据9477条。通过去除3条经纬度异常数据，用于分析数据空间分布和类型分布的有108083条，其中点位记录98270条，拍照记录9813条。去除部分时间属性丢失的数据，用于分析时间分布规律的数据有97593条，其中点位记录88630条，拍照记录8963条。

2.1 系统数据空间分布

从空间分布上看，目前“绿途”经过植被类型标准化处理的108083条记录已经分布到全国各省市自治区，尤其集中分布在中国的东南部(图4A)。从绝对数据量分布上看(图4B)，排在前十位的地区分别是广西壮族自治区、山东省、湖南省、四川省、山西省、甘肃省、江苏省、福建省、内蒙古自治区和陕西省。这前十个地区的数据量占全国总数据量的93.35%。同时，考虑到各省市自治区面积差异较大，本团队分析了各省单位面积的数据分布(图4C)，结果显示，排在前十位的地区分别是广西壮族自治区、山东省、湖南省、上海市、山西省、北京市、江苏省、宁夏回族自治区、福建省和四川省，这前十个地区的数据量占全国总数据量的91.65%。相比于绝对数据分布，单位面积的相对数量分布体现出上海市、北京市和宁夏回族自治区等面积相对较小的地区的数据分布密度较大，而绝对数量排名前十名中的甘肃省、内蒙古自治区和陕西省等地区的数据密度相对较低。

2.2 系统数据时间分布

从时间分布上看，经过植被类型和时间标准化处

理的97593条数据中包括点位记录88630条和拍照记录8963条。标准化后的数据统计时间是2018年7月1日至2020年6月13日。从累积数据量增长曲线(图5A)中可以看出，“绿途”数据的累积增长主要集中在2019年11月至2020年1月，这可能是由于秋季以后植被外貌差异明显便于调查，同时野外调查人员的参与程度较高。2020年1~4月，数据呈现出缓慢增长的平台期，这可能是由于冬季气候条件不适宜于开展大量野外调查。自2020年4月22日开始，“绿途”系统的数据总量呈现爆发式增长，且增长趋势保持在平均每日大于1000条的增长速度。另一方面，从所有数据、点位数据和照片数据的累积增长曲线可以发现，“绿途”系统数据量的增长主要受益于点位采集功能的开发和使用。

从单日净增长数据量曲线图(图5B)可以看出，数据的两次增长高峰分别位于2019年12月左右和2020年5月左右。总体而言，单日净增长率最大的是2020年4月24日，共有10位用户上传数据，数据量为5358，其中点位数据量为5353(8位用户)、照片数据量为5(3位用户)。进一步分析发现，截至目前，单日单人上传最大数据总量、最大点位数据量和最大照片数据量分别为2543条、2543条和95条(285张照片)。

2.3 系统数据类型分布

从植被类型分布上看，经过植被类型标准化处理的用于分析植被类型分布的数据总量为108083条，其中点位记录98270条，拍照记录9813条。这些数据覆盖582种群系、291种群系组、83种植被亚型、35种植被型和7种植被型组。

群系水平统计结果显示(图6A)，数据量排名前十的群系类型分别为马尾松林、毛竹林、华北落叶松林、芦苇草甸、杉木林、刺槐林、侧柏林、柏木林、油松林和黑松林，占据所有数据的34.27%，其中数量最大的马尾松有9598条。

群系组水平统计结果显示(图6B)，数据量排名前十的群系组类型分别为暖性松林、温性松林、侧柏林、杉木林、毛竹竹林、刺槐林、芦苇典型草甸、柏木林、落叶松林和羊蹄甲灌丛，占据所有数据的37.00%，其中数量最大的暖性松林有9607条。

植被亚型水平统计结果显示(图6C)，数据量排名前十的植被亚型分别为暖性常绿针叶林、温性常绿针叶林、暖性落叶阔叶林、暖性竹林和竹丛、温性落叶

图 4 “绿途”数据的(A)空间分布规律及(B)绝对数量和(C)单位面积相对数量百分比在全国各省市自治区的前十名排序

Figure 4 (A) The spatial distribution law of data collected by the “LiVegetation” system, (B) the absolute quantity ranking of the data, and (C) the relative quantity per unit area ranking of the data of the top ten provinces, municipalities, and autonomous regions in China

阔叶灌木、根茎草类典型草甸、温性落叶阔叶林、寒温性落叶针叶林、石灰岩山地常绿阔叶灌木和暖温性常绿阔叶林, 占据所有数据的47.05%, 其中数量最大的暖性常绿针叶林有16469条。

植被型水平统计结果显示(图6D), 数据量排名前十的植被型分别为常绿针叶林、落叶阔叶林、竹林与竹丛、落叶阔叶灌木、根茎草类草甸、常绿阔叶林、落叶针叶林、常绿与落叶阔叶混交林、常绿阔叶灌木和雨林, 占据所有数据的53.68%, 其中数量最大的常绿针叶林有30864条。

植被型组水平统计结果显示(图6E), 各植被型组数量从多到少依次为森林、栽培植物、灌木、草甸与

草丛、荒漠、草原、湿地植被和未匹配植被型组, 其中数量最大的森林类型有50657条。

3 讨论

3.1 众源数据采集的高效性、准确性和实用性

众源数据采集的方式大幅提高了数据采集的效率^[32,33]。从“绿途”系统数据的空间分布可以发现, 在过去近2年时间里, 通过“绿途”平台采集的数据已经广泛分布于全国, 这是传统植被调查时依靠单人或个别组织进行数据调查难以实现的。同时, “绿途”数据的时

图5 “绿途”总数据、点位数据和照片数据随时间的(A)累积分布和(B)单日净增长分布曲线图

Figure 5 (A) Cumulative growth curve and (B) single-day net growth curve of “LiVegetation” total data, point data, and photo data over time

间分布规律进一步表明,通过智能终端进行数据记录和上传的方式可以实现单人单日高达2543条点位数据和379条照片数据的效率。这主要得益于“绿途”APP设计有自动引导的拍照模式和快捷的点位记录方式。在拍照模式中,用户可以按照既定的拍照规则快速完成照片的采集,系统会自动记录照片的经纬度信息,用户可以快速匹配需要输入的群系信息,从而大大减少了传统植被调查时通过相机拍照、纸质记录和内业整理的时间。在点位采集模式中,对于在用户目视可见的较大范围区域内的植被类型,用户可以通过在地图上标记点的方式快速记录植被类型及其位置,这大幅度减少了调查时的行走路程,在一定程度上提高了对于部分难以到达的调查区域的可行性,增加了数据的丰富性,提高了整体的作业效率。

“绿途”众源数据的准确性主要来源于APP端和

Web端两个方面的设计。在APP端,为了保证数据的准确性,系统要求照片数据具备优势种整体、优势种细节和伴生种照片,这有利于保障专家对数据进行后期审核,同时照片数据和点位数据的群系名称都采用了标准的群系匹配模板,进一步保障了数据的规范性和准确性。在Web端,为了进一步提高照片数据的准确性,“绿途”系统设置了多层级的审核机制,即对于未审核数据,需要经过初级审核,再经过终极审核。

“绿途”众源数据的有效性主要体现在植被制图领域。作为一款为植被群系调查定制的APP,目前采集的丰富的植被类型数据(图6)已经被应用于1:100万中国植被图的更新^[34]。更新后的植被图在植被型组、植被(亚)型和(亚)群系水平上的精度较上一代1:100万植被图有了10%~15%左右的提升。此外,“绿途”采集的众源植被数据也正在被用于绘制1:50万中国植被图。

3.2 下一步研究计划

尽管“绿途”系统在数据采集上已经呈现出较好的高效性、准确性和实用性,当前研究还存在一些值得进一步研究的方向。主要包括以下3个方面。

(1) 高效性。为进一步提升“绿途”系统用户的使用体验,实现用户在全国范围的广泛分布。未来将重点从数据采集高效性和趣味性出发。群系名称的自动识别是正在开展的一项工作。目前已经完成基于花伴侣API识别植物物种的功能原型开发,用户可以在拍照阶段选择性地识别照片所对应的物种名称。在此基础上,本团队计划构建物种名称与植被调查群系系统的语义关系库,实现从自动识别的物种名称记录群系名称,从而提高数据采集记录的效率。这一改进也将有利于降低“绿途”系统的使用门槛,推动植被众源调查的流行。

(2) 准确性和代表性。众源数据的质量是这类系统设计时需要考虑的重点和难点^[35]。在本研究中,数据质量主要指数据的准确性和代表性。数据的准确性体现在位置的准确性和群系类型的准确性。位置的准确性决定于数据采集时的系统定位精度。“绿途”系统采集的数据均有万分之一以上的经纬度精度要求。由于部分机型定位误差低,外业使用时建议用户选用中高端手机品牌。为了在不同情况下都能达到要求的数据定位精度,系统在采集数据时提供了人工修改数据坐标的功能;对于存在系统性定位误差的数据,“绿途”

图 6 “绿途”所有数据在(A)群系、(B)群系组、(C)植被亚型、(D)植被型和(E)植被型组水平的分布

Figure 6 Distribution of “LiVegetation” data at the level of (A) formation, (B) formation group, (C) vegetation-subtype, (D) vegetation-type, and (E) vegetation-type group

在后台设计了统一的坐标转换功能来提高定位精度。此外,在数据使用阶段,用户可根据数据采集时记录的手机定位误差对数据质量进行筛选和控制,从而满足不同任务的需求。

影响群系类型准确性的一个重要因素是系统提供的群系匹配系统。目前,“绿途”系统在采集过程中,用于保证数据群系名称准确性的标准匹配系统存在不全面和不定期更新的问题。群系系统的动态调整和关联是一项重要的工作。一方面,部分数据在上传时,用户会自动输入自定义的描述信息导致后期数据标准化的困难。另一方面,标准群系系统的不定期更新也导致不同时期采集的数据匹配困难。目前,已经设计有自定义群系系统上传的功能,在未来使用过程中,用户可以将自定义群系系统进行合并,从而提高匹配的准确性。此

外,本团队正在设计不同时期标准群系系统之间的动态调整关系,使群系系统变化时,能自动对不同系统匹配的群系名称按照最新标准进行匹配。最后,对于人为操作带来的数据群系名称和描述错误等问题,“绿途”系统在APP端设计有3天的数据群系名称修改权限和用户描述信息永久修改权限,同时在网页端支持数据评论和交流。

数据的代表性主要是空间位置和群系类型的代表性。目前,“绿途”系统采集的数据存在空间和类型分布不均衡的问题。已有数据的空间分布主要集中在我国的东南部地区,局部地区较为集中分布在道路沿线。类型分布包括582种群系类型,与上一代1:100万植被图的800余种群系类型还有一定的差距。为提升数据的代表性,除了采集更多的数据量以外,还需要指引

用户前往数据分布少的地区进行采集。未来计划通过设置调查奖励计划、根据数据位置调整积分权重等机制进行引导和改进。这种激励措施也有望进一步提升用户兴趣,促进用户群体的增长和数据的增长。

(3) 安全性、有效性和可持续性。数据安全是众源调查系统的重中之重^[36]。除了对服务器系统进行基本的安全配置之外,目前“绿途”系统还采用系统和数据分离的开发方式提升数据存储的稳定性,同时,采用多个服务器定期备份的方式提高数据的安全性。

为了提升“绿途”系统的有效性,促进系统的可持续性发展,一方面本团队正在设计完善的数据共享机制,使得用户在采集数据的基础上,能够充分利用和发挥数据的科学价值。现阶段,各省数据的获取可通过联系该省份专家进行合作获取,全国尺度的数据可通过联系“绿途”管理团队进行协调。未来,本团队将构建自动化的数据申请、协调和下载机制。另一方面,本团队也期望在定位于服务植被制图的基础上,提高用户使用的动机^[37],拓展植被调查时的样方记录功能,服务于森林、农业、草地和城市等领域的植被和生态调查,促进生物多样性保护研究和管理政策的制定。更多需求和建议,请访问“绿途”项目空间进行反馈交流

(<https://github.com/ShichaoJin/LiVegetation-issue-reporting/issues>).

4 总结

“绿途”是基于公民科学时代下众源采集的理念进行设计和开发的一款植被调查软件系统。“绿途”系统设计的最初目的是直接服务于1:50万中国植被图绘制所需的全国尺度海量植被群系分布信息的采集。“绿途”系统包含数据采集的APP端和数据审核管理的网页端。APP端主要功能包括对照片数据和点位数据的采集,同时设计有辅助野外数据采集的图层管理、兴趣点管理、轨迹管理、用户管理、数据展示等辅助功能。网页端的核心功能是数据管理,主要包括数据上传、数据审核、数据导出,同时也包括与APP端相应的管理功能。目前,“绿途”系统已经在全国34个省级行政单元投入使用,其采集的群系信息数据也被作为地面样本应用于1:100万中国植被图的更新。未来,本团队将重点提升系统的高效性、准确性、代表性、安全性、有效性和可持续性,为植被制图、生态调查、生物多样性保护研究和管理政策制定等提供工具和数据支持。

参考文献

- 1 Lenihan J M, Drapek R, Bachelet D, et al. Climate change effects on vegetation distribution, carbon, and fire in California. *Ecol Appl*, 2003, 13: 1667–1681
- 2 Gaston K J, Blackburn T M, Goldewijk K K. Habitat conversion and global avian biodiversity loss. *Proc R Soc Lond B*, 2003, 270: 1293–1300
- 3 The Editorial Committee of Vegetation Map of China, Chinese Academy of Sciences. Vegetation Map of the People's Republic of China (1:1000000) (in Chinese). Beijing: Geological Publishing House, 2007 [中国科学院中国植被图编辑委员会. 中华人民共和国植被图 (1:1000000). 北京: 地质出版社, 2007]
- 4 Allen F, Qian J, Qian M. Law, finance, and economic growth in China. *J Financ Econ*, 2005, 77: 57–116
- 5 Chow G C. Capital formation and economic growth in China. *Quart J Econ*, 1993, 108: 809–842
- 6 Gao Q, Guo Y, Xu H, et al. Climate change and its impacts on vegetation distribution and net primary productivity of the alpine ecosystem in the Qinghai-Tibetan Plateau. *Sci Total Environ*, 2016, 554–555: 34–41
- 7 Peng S, Chen A, Xu L, et al. Recent change of vegetation growth trend in China. *Environ Res Lett*, 2011, 6: 044027
- 8 Wang J, Wang K, Zhang M, et al. Impacts of climate change and human activities on vegetation cover in hilly southern China. *Ecol Eng*, 2015, 81: 451–461
- 9 Zhang Y, Zhang C, Wang Z, et al. Vegetation dynamics and its driving forces from climate change and human activities in the Three-River Source Region, China from 1982 to 2012. *Sci Total Environ*, 2016, 563–564: 210–220
- 10 Guo K, Liu C C, Xie Z Q, et al. China Vegetation Classification: concept, approach and applications. *Phytocoenologia*, 2018, 48: 113–120
- 11 Song Y C. Vegetation Ecology (in Chinese). Shanghai: East China Normal University Press, 2001 [宋永昌. 植被生态学. 上海: 华东师范大学出版社, 2001]
- 12 Beard J S. The vegetation survey of Western Australia. *Vegetatio*, 1975, 30: 179–187

- 13 Naidoo R, Hill K. Emergence of indigenous vegetation classifications through integration of traditional ecological knowledge and remote sensing analyses. *Environ Manage*, 2006, 38: 377–387
- 14 Molnár Z, Barthá S, Seregélyes T, et al. A grid-based, satellite-image supported, multi-attributed vegetation mapping method (MÉTA). *Folia Geobot*, 2007, 42: 225–247
- 15 Lawley V, Lewis M, Clarke K, et al. Site-based and remote sensing methods for monitoring indicators of vegetation condition: An Australian review. *Ecol Indic*, 2016, 60: 1273–1283
- 16 Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: a review. *J Plant Ecol*, 2008, 1: 9–23
- 17 Bonney R, Cooper C B, Dickinson J, et al. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 2009, 59: 977–984
- 18 Turrini T, Dörler D, Richter A, et al. The threefold potential of environmental citizen science—Generating knowledge, creating learning opportunities and enabling civic participation. *Biol Conserv*, 2018, 225: 176–186
- 19 Silvertown J. A new dawn for citizen science. *Trends Ecol Evol*, 2009, 24: 467–471
- 20 Soroye P, Ahmed N, Kerr J T. Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. *Glob Change Biol*, 2018, 24: 5281–5291
- 21 Dickinson J L, Zuckerberg B, Bonter D N. Citizen science as an ecological research tool: challenges and benefits. *Annu Rev Ecol Evol Syst*, 2010, 41: 149–172
- 22 Çolak E, Sunar F. The importance of ground-truth and crowdsourcing data for the statistical and spatial analyses of the NASA FIRMS active fires in the Mediterranean Turkish forests. *Remote Sens Appl Soc Environ*, 2020, 19: 100327
- 23 Guo Q H, Jin S C, Li M, et al. Application of deep learning in ecological resource research: Theories, methods, and challenges (in Chinese). *Sci Sin Terrae*, 2020, 50: 1354–1373 [郭庆华, 金时超, 李敏, 等. 深度学习在生态资源研究领域的应用: 理论、方法和挑战. 中国科学: 地球科学, 2020, 50: 1354–1373]
- 24 Drake J D, Worsley J C. Practical PostgreSQL. Boston: O'Reilly Media Inc, 2002
- 25 Luo Z W. Using ORDBMS to store GIS data (in Chinese). *Earth Sci: J China Univ Geosci*, 2002, 27: 267–270 [罗忠文. 应用对象关系型数据库存储GIS数据. 地球科学: 中国地质大学学报, 2002, 27: 267–270]
- 26 Zhao W S. Research on application of relational and non-relational database (in Chinese). Dissertation for Master's Degree. Beijing: North China Electric Power University (Beijing), 2016 [赵文硕. 关系型与非关系型数据库的应用研究. 硕士学位论文. 北京: 华北电力大学(北京), 2016]
- 27 Youngblood B. GeoServer Beginner's Guide. Birmingham: Packt Publishing Ltd, 2013
- 28 Merkel D. Docker: lightweight linux containers for consistent development and deployment. *Linux J*, 2014, 1: 2
- 29 Nedelcu C. Nginx HTTP Server: Adopt Nginx for Your Web Applications to Make the Most of Your Infrastructure and Serve Pages Faster Than Ever. Birmingham: Packt Publishing Ltd, 2010
- 30 Aggarwal S. Flask Framework Cookbook: Over 80 Proven Recipes and Techniques for Python Web Development with Flask. Birmingham: Packt Publishing Ltd, 2019
- 31 Filipova O. Learning Vue.js 2. Birmingham: Packt Publishing Ltd, 2016
- 32 Zapponi L, Cini A, Bardiani M, et al. Citizen science data as an efficient tool for mapping protected saproxylic beetles. *Biol Conserv*, 2017, 208: 139–145
- 33 Mourafetis G, Apostolopoulos K, Potsiou C, et al. Enhancing cadastral surveys by facilitating the participation of owners. *Surv Rev*, 2015, 47: 316–324
- 34 Su Y, Guo Q, Hu T, et al. An updated Vegetation Map of China (1:1000000). *Sci Bull*, 2020, 65: 1125–1136
- 35 Alabri A, Hunter J. Enhancing the quality and trust of citizen science data. In: 2010 IEEE Sixth International Conference on e-Science. Brisbane. 2010. 1–10
- 36 Bowser A, Wiggins A, Shanley L, et al. Sharing data while protecting privacy in citizen science. *Interactions*, 2014, 21: 70–73
- 37 Tiago P, Gouveia M J, Capinha C, et al. The influence of motivational factors on the frequency of participation in citizen science activities. *Nat Conserv*, 2017, 18: 61–78

LiVegetation: an investigative tool for vegetation mapping in the era of citizen science

JIN ShiChao^{1,2,3}, HU TianYu^{1,2}, SU YanJun^{1,2}, MA Qin^{1,2}, GUAN HongCan^{1,2}, YANG MoHan^{1,2} & GUO QingHua^{1,2}

1 State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China;

2 University of Chinese Academy of Sciences, Beijing 100049, China;

3 Plant Phenomics Research Centre, Nanjing Agricultural University, Nanjing 210095, China

Vegetation maps are important for understanding the spatial distribution and geographical pattern of vegetation. Vegetation maps also provide important auxiliary information for natural resource research and management decision-making. Vegetation mapping is based on vegetation survey data. To realize vegetation mapping in continuous areas, the data are supplemented by 3S technology, i.e., remote sensing, geographic information systems, and global positioning systems. Despite the rapid development of 3S technology, obtaining massive distribution information of vegetation groups remains a crucial prerequisite and also represents a bottleneck for large-scale vegetation mapping. With the increased availability of smart terminals and personal remote sensing devices, combined with the rise of citizen science, the crowdsourcing model has created new opportunities for vegetation formation investigation. In this study, we developed a collection and analysis system for vegetation group information based on smartphones and web services, which we refer to as LiVegetation. The LiVegetation application supports Android and iOS systems, and the main functions include data collection, layer management, point of interest management, trajectory management, user information, and data display. The LiVegetation application can efficiently collect photo and text records with geographic information in the field. The LiVegetation web system provides management functions that correspond to the LiVegetation application as well as a data inspection function with hierarchical experts. Since July 2018, the LiVegetation system has collected 108,083 data records, which are widely distributed across 34 provinces, municipalities, and autonomous regions in China. These data cover seven vegetation group types and 582 vegetation formation types. The maximum number of data uploaded by the system in a single day is 5,358, and the maximum total number of data uploaded by a single person in a single day is 2,543. Data collected by the LiVegetation system have been successfully applied to update the 1:1,000,000 Vegetation Map of China and are serving the mapping of the 1:500,000 Vegetation Map of China. In future, it is expected that LiVegetation data will continue to support vegetation ecology research and policy management.

LiVegetation, vegetation map, crowdsourcing, citizen science, APP, Web

doi: [10.1360/SSV-2020-0209](https://doi.org/10.1360/SSV-2020-0209)