基于公民科学数据测算物种保护优先性的方法优化研究——应用机器学习与智能优化算法

Optimizing the Framework for Species Conservation Priorities Calculation Based on Citizen Science Data: Application of Machine Learning and Intelligent Optimization Algorithms

侯姝彧
尚轩仪
刘　彦
李　晖*
梁健超

HOU Shuyu
SHANG Xuanyi
LIU Yan
LI Hui
LIANG Jianchao

摘　　要：公民科学观测记录是生物多样性保护相关规划研究和实践中常用的数据来源，但存在记录点代表性有限和热点聚集等问题，充分了解其局限性及可能存在的偏差对于有效的保护规划至关重要。选用广州市100条鸟类系统调查样线数据作为基准，分析同时期公民科学数据的偏差情况及不同数据筛选方法和优先性测算方法的改善效果。采用3种数据稀疏方式减小公民科学数据热点聚集产生的影响，对鸟类记录点及其所处环境进行机器学习并构建275种鸟类的分布模型，基于此测算保护优先性。广州市公民科学观测记录数据热点聚集明显，对其进行稀疏有助于减小物种分布模拟偏差，但与基于系统调查数据得出的结果相比仍具有较大差距。对比传统丰富度方法与智能优化算法的保护优先性测算结果显示，智能优化算法可以更有效地识别丰富度不高但对特定物种更重要的区域，且对公民科学数据采样不均问题带来的保护优先性测算结果偏差具有良好的改善效果。因此，在利用公民科学观测记录进行规划研究和实践时，宜采用多种方式进行数据筛选、物种分布模拟及保护优先性测算，以取得更加可靠的结果。

关 键 词：风景园林；风景园林规划；生物多样性保护；物种分布模型；人工智能；公民科学

文章编号：1000-6664(2024)09-0029-07
DOI：10.19775/j.cla.2024.09.0029
中图分类号：TU 986
文献标志码：A
收稿日期：2024-05-04
修回日期：2024-06-26
基金项目：国家自然科学基金面上项目(52078222)；广东省自然科学基金面上项目(2024A1515010783)；广东省基础与应用基础研究基金项目(2021A1515110744)；2023年度广州市水务科技项目(GZSWKJ2022-008)

Abstract: Citizen science records are commonly used as data sources in planning research and practice. However, there are issues including limited representativeness of observation points and clustering of hotspots, leading to biases in the analysis based on them. It is important to fully understand their limitations and possible deviations for effective conservation planning. We used bird survey data from 100 transects in Guangzhou as a benchmark to assess how various data filtering and prioritization methods improves the results from concurrent citizen science data. Applied machine learning with a max entropy model to simulate 275 bird species' occurrence probabilities according to their records and environments, which then informed conservation priority calculations. The results revealed clustering in Guangzhou's citizen science data. Sparsity can alleviate deviation, yet it still lags significantly in representativeness compared to systematic survey-based results. Compared with the conservation priority estimation results of the traditional richness method and the intelligent optimization algorithm, it is found that the latter can identify habitats with low richness but more important for specific species effectively, and has an obvious alleviation on the bias of conservation priority estimation results caused by the uneven sampling of citizen science data. The results show that it is advisable to use a variety of methods to screen data, simulate species distribution, and estimate conservation priorities in order to obtain relatively reliable results when using citizen science records for planning research and practice.

Keywords: landscape architecture; landscape planning; biodiversity conservation; species distribution model; artificial intelligence; citizen science

1 研究背景

随着中国生态文明建设工作不断深入，生物多样性价值评估及潜在物种调查分析已成为风景园林和城乡规划实践中的重要前置环节。例如，在进行自然保护地选址划界等规划建设前，在区域尺度上分析物种信息并评估保护优先性是必要的基础内容；在生物多样性保护空间网络布局和城乡蓝绿空间规划中，对生物多样性保护空缺的识别需要基于物种分布信息及保护优先性分析结果；列入《中国生物多样性保护战略与行动计划(2023—2030年)》优先项目的“探索社区、景区、学校等生物多样性友好城市单元”及“城市绿地和公园等生物多样性体验地建设”等实践工作，其选址和规划设计都离不开潜在物种分析和所在区域生物多样性保护优先性测算。

在以上过程中，科学的物种调查或分布模拟及基于此的生物多样性保护优先性分析十分必要，其有效性高度依赖物种分布信息。然而在实际工作中，针对物种的系统性实地调查数据非常有限，通常难以支撑科学可靠的前期研究。近年来兴起的以观鸟记录为代表的公民科学数据为

* 通信作者(Author for correspondence) E-mail: ydlihui@scau.edu.cn

风景园林研究中的生物多样性要素提供了广泛的数据来源[1-2]。然而，由于公民科学记录通常在没有实验设计的情况下自发进行，缺少对调查地点的系统布局和调查方法的规范统一，增加了数据集的偏差和噪声，影响建模预测的准确性[3]。具体而言，依据物种记录数据在区域尺度上测算生物多样性保护优先性通常包含2个核心步骤：根据物种记录构建物种分布模型(species distribution model，SDM)，以及根据物种分布分析各区域对生物多样性保护的重要性。SDM通过将已知物种出现点与其所在环境联系起来，以描述和预测物种在各个位置的出现概率[4]，因此，用于构建SDM的输入点位对物种实际分布的代表性水平对模型的准确性具有重要影响。

相关研究显示，可靠的SDM往往来自结构化项目[5]，如通过系统设计的样线、样点调查等获取的数据，但这类数据较难获得，目前应用较广泛的公民科学平台如ebird、iNatrualist、中国观鸟记录中心等的数据仍以非结构化为主，少有半结构化数据，因此通常被认为质量较低[6]。目前，公民科学数据公认的局限性主要有：对于物种记录较多地区[7]、便于到达的区域[4]，以及体型较大的物种[8]的记录量显著偏高，反之记录不足。虽然在少量研究中公民科学记录能够取得与系统调查数据相近的结果[9]，但研究普遍显示基于公民科学数据的SDM的准确性与物种[10]、调查时长、调查者水平[11]和数据调查的结构性程度[8]等有较大关联，并强调了高质量的公民科学数据对物种预测的重要性[12]。因此，利用公民科学数据进行研究，需要首先使用一定方法进行数据清洗，常用的方法包括密集区域稀疏化[2, 13]、选用符合特定记录时长或记录者条件的调查记录[3, 11]，以及改变采样方式[4]等。此外，将物种“出现-未出现”数据建模与仅使用“出现”数据建模相结合[14]，以及多物种联合SDM[4, 15]等建模手段，也被用于改善公民科学记录数据的建模结果。但对于各种处理方法的效果，不同研究的结论尚存在较大差异。

基于物种分布模拟结果进行保护优先性分析时，最常用的方法是通过物种潜在丰富度识别热点区域[16-17]。但这种基于α多样性的方法在有效识别生物多样性丰富区域的同时，可能会忽略物种不是最多但对于特定物种具有重要价值的区域[18]，而由于公民科学数据常具有热点过度聚集的缺陷，这一问题将尤为显著。结合基于β多样性的优先性识别方法能够综合考虑对物种丰富度不高区域的特定物种的重要性[19-20]，可以改善由于记录点位偏差引起的热点区域过分集聚的情况。如智能优化算法[18]和深度学习框架[21]等人工智能算法，可以通过不断迭代模拟评估不同保护方案下的物种灭绝风险等信息，测算各区域的保护优先性，以减小公民科学记录数据带来的偏差。

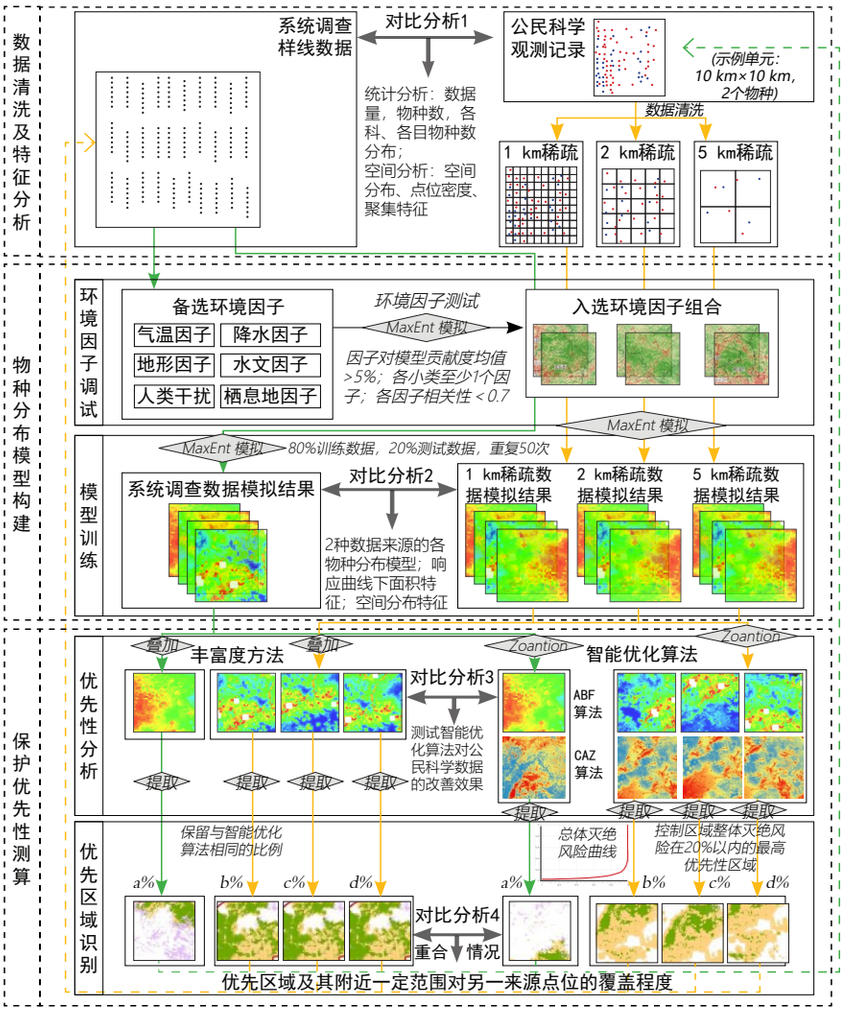

图1 研究框架及技术路线

大多数情况下，公民科学观测记录是规划设计师在实践中能够获取的最优物种记录数据，因此，充分了解其局限性对于进行符合生物多样性保护需求的规划设计至关重要。选取广州市为研究范围，鸟类为保护对象，以结构化的系统调查数据作为比较基准，分析基于非结构化的公民科学记录测算保护优先性时可能存在的局限性，重点关注2个问题：1)比较分析公民科学记录数据与系统调查数据的物种记录情况及其分布特征，定量理解基于公民科学记录构建SDM可能存在的偏差及不确定性来源；2)借助机器学习方法构建并调试物种分布模拟模型，获取适宜的数据清洗方式和环境变量组合，并基于模拟退火算法测度研究区域的保护优先性，验证这一算法能否减小非结构化的公民科学记录带来的不确定性。

2 数据和方法

研究框架及技术路线如图1所示。

2.1 物种分布模拟及保护优先性测算

2.1.1 物种记录数据特征分析

系统样线调查数据为广东省科学院动物研究所在2022年全年对100条结构调查样线的调查记录，共22 040个点位，包含275种鸟类。公民科学物种记录数据从中国观鸟记录中心(http://www.birdreport.cn)搜索同时期广州市范围内的公开鸟类调查记录报告，整理其中观鸟记录点位及相应物种信息，共获取149 882个鸟类记录点位，包含578种鸟类。选取二者重合的275种鸟

类进行分析和对比，为基于公民科学数据测算保护优先性的方法优化提供依据。

对2种来源数据进行统计分析和空间分析。统计记录数据的数量和鸟种，比较不同目和科中鸟种记录情况，分析公民科学数据存在的偏差及其可能的原因。此外，通过空间制图显示2种来源数据的空间分布情况，分析二者的核密度及其相关性，了解公民科学数据的潜在偏差情况。

2.1.2 模型选择与数据清洗

选取最大熵(MaxEnt)模型方法进行模拟，这是一种常用的机器学习方法，通过非监督学习方式解析物种记录点位与环境因子的关系，预测物种在研究区域各处的出现概率。MaxEnt方法对样本量敏感性较低，在小样本量(n<30)的研究中，显著优于广义增强回归模型、广义加性模型和广义线性模型等其他模型，而且它在大样本量计算中的表现同样优秀[22]。此外，目前获取的数据属于仅有出现记录的数据，而基于回归的模型在处理此类数据时存在“污染控制”的问题[23-24]。考虑到目前数据集中尚有多种鸟类被观测到的点位数量较少，因此选择MaxEnt方法，为尽量多的鸟类模拟出相对准确的SDM。

采用相关研究常用的稀疏化方法处理公民科学记录点位数据[4]。考虑研究区域面积和物种记录点位情况，设定1、2、5 km 3种网格，分别对应近似研究面积常用网格[2]、整体与样线数据密度近似、城市中心区密度与样线数据密度近似3种状态(以下简称“1 km稀疏”“2 km稀疏”“5 km稀疏”)。每个物种在每个网格内分别随机保留一个点位，获取清洗后的3组鸟类记录点位数据，与样线调查数据进行对比分析。

2.1.3 环境因子筛选及物种分布模型构建

参考相关研究，结合研究区环境特色及鸟类分布特点，选取气候、生境和人类干扰三大类环境因素，初选6个小类26个备选环境因子[1-2, 13](附表1①)，将其统一为100 m×100 m的空间分辨率，并基于样线数据，用MaxEnt对数据进行测试(用每组变量对每个物种进行10次模拟)，筛选环境因子。采用刀切法计算各因子作为唯一变量时的结果，以及仅将该因子去掉后的结果，反映各因子对各物种的预测结果的影响程度。去除预模拟中贡献度(percentage contribution)小于5%的因子，从每小类中选取对各物种总体影响较大的因子(permutation importance高于均值)，并在保证每小类至少包含1个因子的前提下，在每对皮尔逊相关性系数＞0.7的相似因子中至多保留一个贡献度较大的因子。筛选后共获得14个因子进行正式模拟。

将至少包含10个记录点位的物种点位信息输入MaxEnt[25]，随机选取80%的点用于模型训练，构建鸟类出现与环境变量的关系模型，其余20%的点位数据用于对预测结果进行验证。设置最大迭代次数为1 000次，每个物种进行50次模拟，将计算结果取均值，得出该物种在广州市出现概率的预测结果。同时，获得预测结果的ROC曲线(receiver operating characteristic curve)，曲线下面积(area under curve，AUC)用于衡量这一模拟结果的性能优劣，筛选测试AUC＞0.7的模型作为有效结果。

为保证分析结果的可靠性，将14个环境因子数据分别处理为200和500 m的空间精度，重复以上操作用于敏感性分析，验证在不同空间分辨率下的研究结果是否具有一致性。

2.1.4 保护优先性模拟分析

采用基于模拟退火算法的Zonation决策模型(https://www2.helsinki.fi /en/researchgroups/digital-geography-lab/software-developed-in-cbig)进行灭绝风险模拟和保护优先性分析。该方法通过每次移除不同单元格并模拟总体灭绝风险，测算每个单元格的保护优先性，越晚被移除的单元格具有越高的保护优先性。将基于样线数据和1、2、5 km稀疏的公民科学调查数据得出的3对物种SDM组合作为输入图层，用其内置的加性收益函数(ABF)和核心区域(CAZ)算法，分别优先考虑物种受威胁程度和丰富度模拟整体保护优先性，并分别生成连续栅格结果。其中，ABF算法优先考虑各单元的总体权重值；CAZ算法优先考虑各单元的最高权重值。参考相关研究[13]，将国家一级、二级保护物种的权重分别设为8、4，IUCN物种红色名录CR、EN、VU、NT级别的物种权重分别设为8、6、4、2，其余权重设为1，不同分类中权重取高值。将翘曲值设置为1，即每次移除1个单元，为最高精度的计算方式；选择“不优先从边缘移除”，使研究区边缘与中心具有同等优先级。为验证智能优化算法的效果，采用相同的权重对同样3组鸟类的SDM进行加和，获取基于潜在丰富度的保护优先性结果。

计算保护优先性后，根据模拟结果中的灭绝风险曲线，结合保护目标从连续谱中选取合适的阈值，在各方法的预测结果中，控制模拟灭绝风险在20%以内时所对应的区域作为相应鸟种在广州市的保护优先区域。基于潜在丰富度的保护优先性结果选取同样的比例作为保护优先区域。对200和500 m敏感性测试数据重复以上操作，获得相应结果。

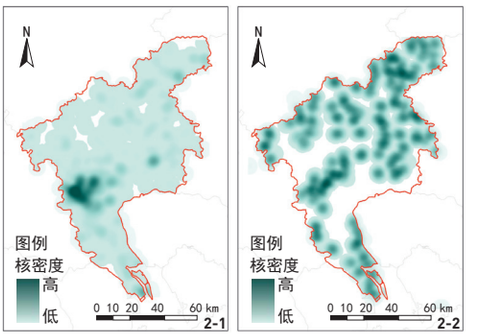

图2 2种数据来源的采样点核密度(2-1 公民科学记录点位密度；2-2 系统调查记录点位密度)

2.2 模拟结果分析及优化效果检验

对基于2种来源数据分析获取的物种分布模型与保护优先性分析结果进行3类对比分析，以判断基于公民科学数据构建SDM及计算保护优先性的结果偏差情况及偏差来源。

1)物种分布模型对比。统计分析系统调查数据与公民科学数据模拟结果的AUC值，并比较各物种在2种数据来源下的模拟结果的空间分布特征及差异。

2)保护优先性测算结果对比。对比在各数据稀疏方式下将各物种SDM加权叠加获取潜在丰富度的方法与通过智能优化算法获取的优先性结果，分析不同条件下公民科学数据结果与基准数据的相关性，验证智能优化算法是否能够有效改善基于公民科学数据测算保护优先性的结果质量。

3)保护优先区域识别结果对比。识别保护优先区域后，基于一个数据源识别出的优先区域对另一数据源记录点的覆盖程度，判断该优先区域的有效性，了解基于公民科学数据的识别结果的优先区域保护水平，并通过与基于丰富度方法的结果对比，验证智能优化算法能否减小公民科学数据偏差对优先区域识别结果的影响。

3 研究结果

3.1 记录点位特征对比

空间制图显示，系统调查样线较为均匀地分

布在广州市范围内，而公民科学数据在市中心区域高度集中(图2)，二者核密度分布的相关性系数为0.408，相关性较低。

不同类群的记录情况统计显示，公民科学记录的鸟种所属科目分布与样线记录中的分布有较大差异。其中，鹬鹬目、鹭形目的比例远高于系统调查中的比例，而鸽形目、雨燕目和鸡形目的比例相比系统调查结果明显偏低，此外，公民科学记录增加了样线调查中没有的犀鸟目和鹳形目(附图1[①])。对鸟类所属不同科的统计结果显示，系统调查记录的鸟类共65科，同期公民科学观测记录结果包含鸟类属82科，各科间记录量差异较大(附图2[①])。

3.2 物种分布模型对比

在采用1、2、5 km网格进行公民科学数据稀疏化后获得的结果中，分别有118、87、42个有效SDM同时存在于基于样线的模拟结果中。

对比分析2个数据来源得出的SDM结果可知，基于系统调查数据的模拟结果整体AUC值更高。基于系统调查数据和1 km稀疏的公民科学数据的118个物种分布模型的AUC均值相近，分别为0.853 7和0.853 9，但对于具体物种，基于系统调查数据的模拟结果有98个(83.05%)高于公民科学记录的模拟结果。2、5 km稀疏的公民科学数据得出的有效结果AUC均值分别为0.817 4和0.844 5，低于1 km稀疏的结果，这可能由于1 km的稀疏对于本研究的数据集不够充分，导致一些物种存在过拟合情况。对具体物种而言，在2、5 km网格记录中，各物种分布模型的AUC值分别有89.66%(*n*=78)和92.86%(*n*=39)低于系统调查数据的模拟结果，因此可以认为基于公民科学数据获得的SDM结果具有更大的不确定性。

3.3 保护优先性测算及优先区域识别

对比2种来源的数据在采用ABF及CAZ 2种方法下得出的优先性分布结果发现，在相同参数下，基于系统调查数据识别出的高优先性区域有更高的空间异质性，同时有更多较小斑块被识别，而基于公民科学数据的结果中，高优先性区域更具集中倾向(图3、4)。基于公民科学观测数据识别出的优先区域包含较多城乡建设区域等可能不适宜物种栖息的范围，而基于系统调查数据得出的结果较少出现此类问题。基于系统调查数据识别出的优先区域涉及范围更广，布点相对分散，涵盖了很多公民科学记录中缺少点位但系统调查中有物种记录的区域。

在2种来源数据得出的优先区域结果中，使用2 km稀疏的数据结果与系统调查结果的相似程度最高(图5)，2种来源数据识别出的保护优先区域有29.55%重合，分别是1和5 km稀疏的数据结果的1.37和1.61倍(附表2[①])。对于直接加权叠加得出的以潜在丰富度表示的优先性结

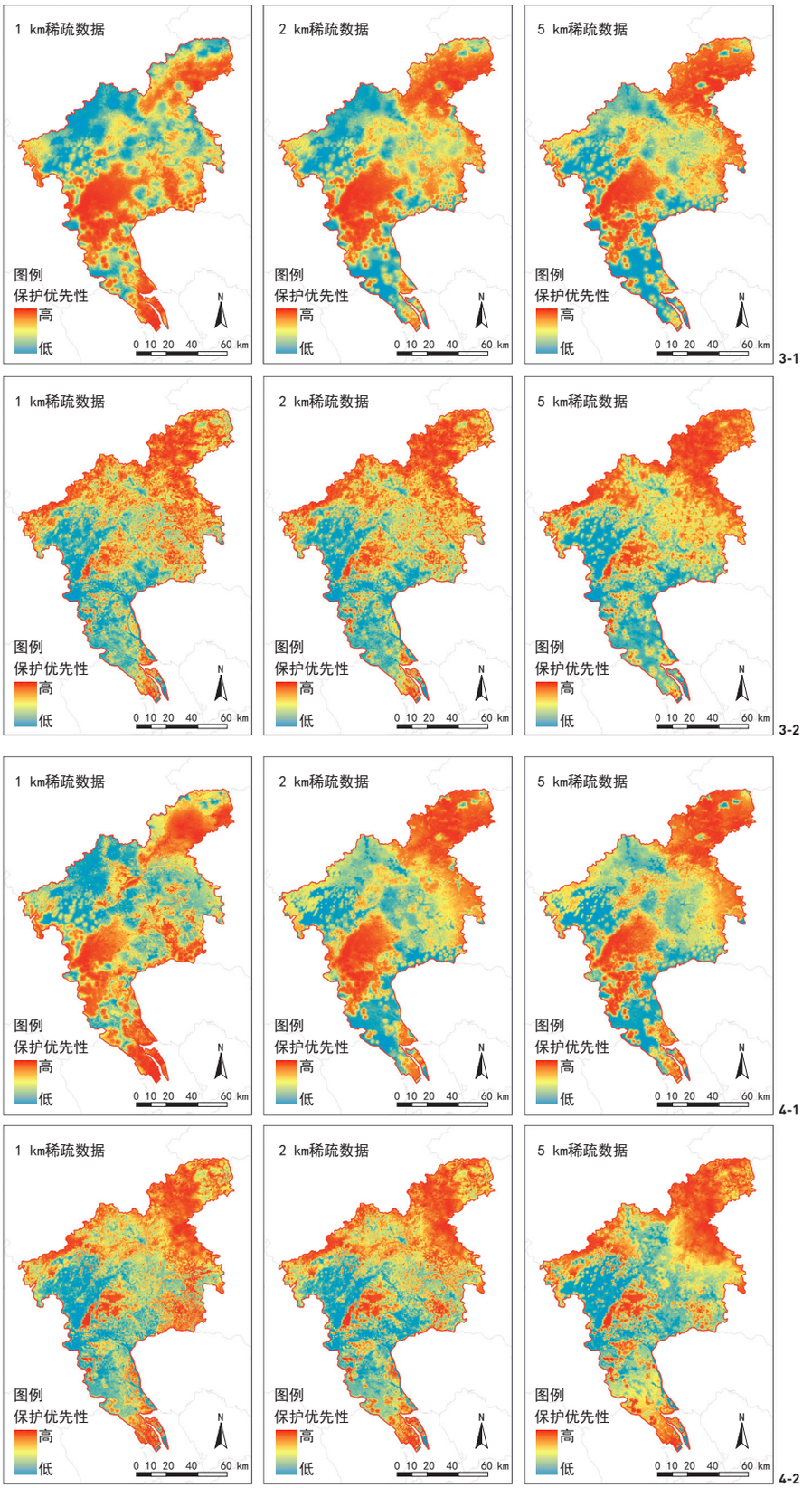

图3 采用加性收益算法在3种数据稀疏条件下计算得出的鸟类保护优先性分布
图3-1 基于公民科学数据的优先性分布
图3-2 基于系统调查数据的优先性分布
图4 采用核心区域算法在3种数据稀疏条件下计算得出的鸟类保护优先性分布
图4-1 基于公民科学数据的优先性分布
图4-2 基于系统调查数据的优先性分布

果，公民科学数据经1、2、5 km稀疏后与系统调查结果的重合比例分别为14.26%、20.40%和29.09%，稀疏物种点位对数据偏差的改善效果随着稀疏网格的增大而提升。

相关性分析结果显示，虽然公民科学数据结果与基准数据结果分布存在较大差异，但采用智能优化算法得出的优先性结果整体优于传统丰富度方法得出的结果(表1)。在100 m精度上，相较于直接叠加SDM获取热点区域，1、2 km稀疏下的公民科学数据基于ABF算法得出的保护优先性与系统调查结果的相关性分别提高10.00%、17.65%，基于CAZ算法得出的优先性分别提高27.30%和30.00%。5 km稀疏下的公民科学数据结果与系统调查数据的结果相关性均超过0.50，其中，ABF算法相较于潜在丰富度的结果有一定提升，但CAZ算法带来的变化不明显。以上结果在200和500 m精度的分析中趋势一致。

分析一种来源数据的识别结果对另一来源点位的覆盖程度结果显示，基于系统调查识别出的保护优先区域可较大程度地涵盖公民科学记录的物种点位(表2)。考虑到鸟类活动范围，分别统计识别出的优先区域及其周边1 km范围进行辅助验证。对于2种来源数据，智能优化算法均可较大程度提升其识别出的优先区域对物种点的覆盖水平。基于系统调查数据，通过优化算法识别出的优先区域及周边1 km范围可涵盖90%以上公民科学记录点，而传统丰富度方法识别结果的覆盖效果较差。对于丰富度叠加方法，公民科学数据稀疏化效果显著，随着稀疏网格变大，其识别出的优先区域对系统调查中的物种点覆盖效果逐渐提高。

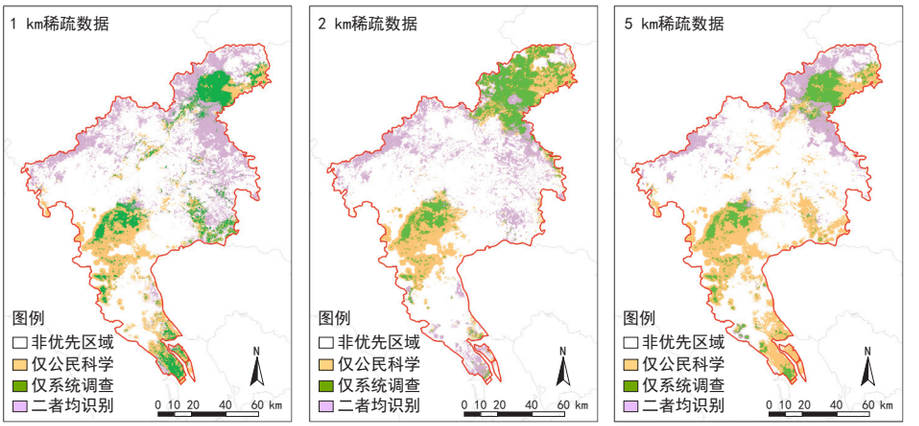

图5 根据2种数据来源在3种数据稀疏条件下保护优先区域分布识别结果

表1 相同物种组合下2种来源数据基于3种测算方法得出保护优先性的相关性

数据筛选方法及分析精度 / 测算方法	1 km稀疏			2 km稀疏			5 km稀疏		
	100 m	200 m	500 m	100 m	200 m	500 m	100 m	200 m	500 m
潜在丰富度结果	0.20	0.27	0.25	0.34	0.38	0.39	0.57	0.51	0.54
加性收益算法结果	0.22	0.34	0.33	0.40	0.44	0.45	0.62	0.57	0.60
核心区域算法结果	0.26	0.47	0.43	0.46	0.50	0.46	0.56	0.51	0.54

表2 一种来源数据识别结果对另一来源点位的覆盖率

统计内容	统计范围	优先区域来源	1 km稀疏	2 km稀疏	5 km稀疏
公民科学数据识别结果对系统调查记录点位覆盖率	优先区域内	丰富度结果	23.81%	36.74%	51.45%
		优化算法结果	62.47%	57.08%	46.56%
	优先区域1 km内	丰富度结果	39.65%	55.54%	75.36%
		优化算法结果	62.47%	69.17%	68.31%
系统调查数据识别结果对公民科学记录点位覆盖率	优先区域内	丰富度结果	48.12%	34.41%	25.57%
		优化算法结果	49.39%	51.23%	49.94%
	优先区域1 km内	丰富度结果	93.11%	83.09%	56.92%
		优化算法结果	94.75%	97.13%	93.22%

4 讨论

4.1 公民科学数据采样偏差明显，数据稀疏效果不稳定

与其他地区的公民科学记录研究结果类似，相比于有限的结构化系统调查数据成果，非结构化鸟类公民科学观测记录可以覆盖更多的物种数量，但存在分布代表性不足、市中心点位过度聚集等问题[8]。与Planillo等的研究发现一致[4]，即使采用稀疏数据的方法，非结构化的公民科学记录数据仍然会增加优先性测度结果中市中心区域的重要性。本研究采用3种不同的网格数据清洗方式对公民科学数据进行筛选，发现对公民科学数据集进行清洗未必能够有效提升物种分布模型的质量，这与相关研究结论一致[26]。对于特定的公民科学数据集，采用何种数据清洗方式能够获得更准确的物种模拟结果仍未有定论。此外，对密集区域点位进行稀疏化不是唯一的公民科学数据筛选方式，未来可结合多种数据筛选方式，如提取单次物种记录较多的报告、来自记录量较多的记录者的报告，以及调查时长较长的报告等，改善非结构化公民科学数据的可用性[3, 27]。

比较各物种的模拟结果发现，基于2种来源数据的SDM结果差异较大，AUC显示公民科学数据获得的结果准确性整体低于系统调查数据结果，佐证了其难以取代系统调查数据[3]，而更适合作为系统调查的补充或在缺少系统调查时使用[26]。由于公民科学记录点较多集中在市中心区域和东北区域，与系统调查结果相比，有近半数物种的分布模拟结果不成比例地集中于这两大片区。虽然通过稀疏化后的结果聚集性整体变低，但相较于系统调查结果，可能由于公民科学数据分布未能充分代表该物种的所有适宜环境特征，导致模拟结果仍有偏差。

基于物种调查数据构建SDM时，数据清理方式和环境因子选择通常由研究人员依经验判断。机器学习方法有助于快速进行不同输入数据的测试训练，使得在一项研究中对比多种不同的数据处理方法及环境变量选择成为可能，在数据数量和质量有限的条件下，能够为研究获得更加可靠的结果；在数据量充足时，通过不断改变训练数据集的样本输入，可以高效地获得更稳健的模拟结果。

4.2 基于公民科学数据测算物种保护优先性的结果特征及应用局限

对于识别出的优先区域，基于系统调查得出的结果具有更高的精细度和更广的覆盖度。这在

市中心区域对比尤其明显，公民科学数据显著集中于该区域，导致识别出的优先区域中缺少市区外的范围，且市中心大片被识别为优先区域，包含了较多人类活动强度较大、自然栖息地质量不佳的区域。而基于系统调查的分析结果将市中心的白云山—天鹿湖—帽峰山、大象岗、大夫山等具有更良好栖息环境的范围清晰地识别了出来，同时涵盖了北部芙蓉嶂—王子山、大封门—南昆山等公民科学记录数据未能识别的重要区域。

本研究在SDM构建中未直接选用土地利用分类对公民科学数据的记录点位进行排除，主要是由于观鸟记录中心的记录中存在较多同一区域如一个公园内或景点内的记录均被上传至同一坐标点的情况，难以据此充分判断物种实际出现的生境。因此，考虑到鸟类能够跨越不同生境类型飞行的特性，采用“到各类用地的距离”这一因子进行物种出现概率学习和模拟。系统调查数据识别结果能够大致显示出自然栖息地边界，佐证了这一方法的可行性，可以认为公民科学数据结果的偏差主要是物种记录点位空间分布不均导致。因此，在使用公民科学观测数据进行生物多样性保护相关模拟和分析结果指导规划时，应结合生境斑块的土地利用情况[27]、物种的生境偏好[28]等因素进一步细化，以获取更符合物种实际生境需求的分析结果。

4.3 智能优化算法辅助降低保护优先性测算的不确定性

相比于使用潜在丰富度识别物种保护优先区域，通过智能优化算法对受保护区域的不同方案进行不断迭代运算，即使这些区域的物种丰富度有限，也可有效覆盖对特定物种具有重要意义的栖息地[18-19]。

在2种算法下，虽然基于公民科学数据的结果对于物种出现点位的代表性明显不足，但通过智能优化算法得出的结果表现相较于传统的丰富度方法更佳。采用2 km网格进行稀疏化的公民科学数据识别出的优先区域与系统调查数据识别结果的重合比例相比于1、5 km稀疏结果有显著提升，可能是由于这一清洗方式与研究区内的物种分布较为契合。对于87种物种组合，基于优化算法识别的结果在加性收益函数和核心区域算法下，公民科学数据与系统调查数据结果的相关性相比丰富度叠加方法分别提高了27.30%和47.10%，可认为在这一数据处理方式下，基于智能优化算法识别保护优先性的方法改善了非结构化调查数据带来的不确定性。

4.4 局限与展望

本研究选取鸟类作为代表类群进行基于公民科学数据的生物多样性保护优先性测算的原因是目前公民科学数据体系尚未成熟，仅有鸟类记录在种类和空间分布上能基本满足保护规划分析需求。研究提供的改善基于公民科学数据测算保护优先性准确性的方法框架，可以应用于其他物种类群和其他地区的研究中。同时，选取了与系统调查同时期即一年内的公民科学数据，在实际规划设计工作中可以适当扩大数据获取时段，以求包含更多可能存在记录点位的区域。

在物种出现点位数据处理中，稀疏化的方法是在每个单元内对每个物种保留一个随机点位，保留点位的不确定性随着网格变大而增加。因此，未来可对同一稀疏网格进行多次运算，对随机产生的多组数据结果构建SDM并使用平均结果进行分析。机器学习为SDM构建中引入多种数据处理方法和输入变量组合提供了高效便捷的方式，但仍未能解决数据缺失区域的问题。本研究采用了基于最大熵模型的SDM构建方法，未来可以进一步应用多种建模方式，如结合随机森林、神经网络等其他机器学习或统计学习的综合方法[15，27，29]，构建不同的物种分布模型，减少单一模型的不确定性，并进一步进行比较，探寻更好地发挥公民科学观测数据优势的方式，提高物种分布模拟及保护优先性测算的质量。

此外，虽然本研究以结构化的样线数据为基准，但同样存在局限。对于一些物种，由于样线数据有限而未能涵盖各类适宜的栖息环境，导致一些公民科学观测到该物种的区域在分布模型中仅呈现较低出现概率。因而，在实际应用中可以将不同来源数据结合，利用多种方法处理数据，以获得更准确的预测结果[27-29]。基于系统调查数据识别的优先区域对物种点位覆盖水平的研究结果提示，识别出的优先区域不宜直接作为明确的空间边界进行保护行动划界，其邻近区域对于物种的栖息很可能同样具有重要价值，需要在实际规划中结合用地条件和管理可行性等因素综合判断。

了解基于公民科学数据进行物种分布模拟的采样可靠性与局限性，有助于在规划和管理决策中采取适宜的方案，避免由于数据偏差引入不适宜的规划措施造成生物多样性保护失效或资源浪费[15]。对于风景园林师，在前期数据有限的研究中，应当在使用公民科学观测记录时谨慎分析数据可能存在的局限性，判断数据调查范围是否涵盖物种的各类栖息环境特征，并尽量综合多种来源数据进行分析[28，30]；对于生物多样性保护研究者，可以在利用公民科学观测记录时，采用多种方式进行数据筛选，并在公民科学数据不足区域补充系统调研，以获得相对准确的物种分布模型；对于公民科学记录平台，为使其记录对生物多样性保护研究具有更好的可用性，可进一步完善不同记录类型，生成更多结构化或半结构化数据，改善整体数据质量[11]；对于保护管理机构、公益组织和相关企事业单位等主体，则可以通过在特定区域增设调查项目、举办物种观测竞赛等形式，增加数据缺乏区域的记录数量，获取质量更好的观测数据，为生物多样性重要区域的保护管理提供充分支撑。

5 结论

保护优先性测度是生物多样性保护规划设计的核心步骤，由于动物分布数据可用性的限制，使根据现有方法进行规划设计能实现的保护效果存在局限。本研究以鸟类为例，以广州市100条样线的鸟类调查数据为基准，将同期公民科学观测记录与其进行对比分析，为优化规划实践中基于公民科学数据测度保护优先性的流程提供了新思路：在SDM构建中，利用机器学习的高效数据处理能力，对公民科学数据进行不同数据筛选方式的比较分析，利用预模拟运算对多种备选环境因子进行筛选；借助智能优化算法，综合物种出现概率及保护等级等因素测度保护优先性，改善物种记录数据带来的偏差。

通过比较基于公民科学数据与系统调查数据获取的保护优先性分析结果，发现相较于传统丰富度测算方法，智能优化算法在大多数场景下均能显著提升基于公民科学数据进行的保护优先性测算质量，具体反映在其结果与系统调查结果的更高相似度，以及优先区域对系统调查物种点的更高覆盖度方面，有效改善了公民科学数据的热点聚集问题。系统调查数据在丰富度方法下得出的优先区域在纳入计算的物种减少时，对记录点位的覆盖能力显著下降，而智能优化算法的结果较为稳定，说明物种较少时采用智能优化算法更佳。此外，研究在不同建模精度下的结果趋势一致，

未来可通过更多研究验证其普适性及适用条件。

研究结果显示，稀疏化公民科学数据可以一定程度上改善采样偏差带来的不确定性，当采用物种丰富度方法测算保护优先性时，稀疏化的改善效果更加明显，但其结果仍与系统调查数据结果有较大差异，且筛选公民科学数据未必总能有效提升物种分布模型的质量。

值得注意的是，尽管研究中采用的数据处理方法和智能优化算法能够改善公民科学数据偏差对保护优先性测算结果的影响，但公民科学数据仍难以取代系统调查数据。因此，在实际应用中可综合多种来源数据和方法进行处理和计算，尽量减小模拟结果的不确定性，以获取相对可信的分析结果。

注：文中图片均由作者绘制。

注释：

① 文中相关附图和附表见链接：http://dx.doi.org/10.13140/RG.2.2.30152.53764。

参考文献：

[1] 阳文锐，李婧，闻丞，等.基于物种分布的北京平原生态网络构建[J].生态学报，2022，42(20)：8213-8222.

[2] 李晖，刘彦，黄伊琳，等.基于Maxent模型的深圳湾鸟类热点生境判别及修复研究[J].中国园林，2022，38(12)：14-19.

[3] Steen V A, Elphick C S, Tingley M W. An evaluation of stringent filtering to improve species distribution models from citizen science data[J]. *Diversity and Distributions*, 2019, 25(12): 1857-1869.

[4] Planillo A, Fiechter L, Sturm U, et al. Citizen science data for urban planning: Comparing different sampling schemes for modelling urban bird distribution[J]. *Landscape and Urban Planning*, 2021, 211: 104098.

[5] Sauer J R, Link W A, Fallon J E, et al. The North American Breeding Bird Survey 1966-2011: Summary Analysis and Species Accounts[J]. *North American Fauna*, 2013, 79: 1-32.

[6] Kekking S, Johnston A, Bonn A, et al. Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity[J]. *BioScience*, 2019, 69(3): 170-179.

[7] Gekdmann J, Heilmann-Clausen J, Holm T E, et al. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements[J]. *Diversity and Distributions*, 2016, 22(11): 1139-1149.

[8] Callaghan C T, Poore A G B, Hofmann M, et al. Large-bodied birds are over-represented in unstructured citizen science data[J/OL]. *Scientific Reports*, 2021, 11: 19073[2024-07-18]. https://doi.org/10.1038/s41598-021-98584-7.

[9] Tye C A, Mccleery R A, Fletcher R J, et al. Evaluating citizen vs. professional data for modelling distributions of a rare squirrel[J]. *Journal of Applied Ecology*, 2016, 54(2): 628-637.

[10] Tiago P, Pereira H M, Capinha C. Using citizen science data to estimate climatic niches and species distributions[J]. *Basic and Applied Ecology*, 2017, 20: 75-85.

[11] Gorleri F C, Hochachka W M, Areta J I. Distribution models using semi-structured community science data outperform unstructured-data models for a data-poor species, the Plain Tyrannulet[J/OL]. *Ornithological Applications*, 2021, 123(4): duab038[2024-07-18]. https://doi.org/10.1093/ornithapp/duab038.

[12] Bradter U, Mair L, Jönsson M, et al. Can opportunistically collected Citizen Science data fill a data gap for habitat suitability models of less common species?[J]. *Methods in Ecology and Evolution*, 2018, 9(7): 1667-1678.

[13] 侯姝彧.中国鸟类保护兼用地优先区域识别与管理研究[D].北京：清华大学，2023.

[14] Guillera-Arroita G, Lahoz-Monfort J J, Elith J, et al. Is my species distribution model fit for purpose? Matching data and models to applications[J]. *Global Ecology and Biogeography*, 2015, 24(3): 276-292.

[15] Woodman S, Forney K, Becker E, et al. eSDM: A tool for creating and exploring ensembles of predictions from species distribution and abundance models[J]. *Methods in Ecoogy and Evolution*, 2019, 10(11): 1923-1933.

[16] Li L, Hu R, Huang J, et al. A farmland biodiversity strategy is needed for China[J]. *Nature Ecology & Evolution*, 2020, 4(6): 772-774.

[17] 任月恒，朱彦鹏，付梦娣，等.黄河流域濒危物种保护热点区与保护空缺识别[J].生态学报，2022，42(3)：982-989.

[18] Kukkala A S, Moilanen A. Core concepts of spatial prioritisation in systematic conservation planning[J]. *Biological Reviews*, 2012, 88(2): 443-464.

[19] Hou S Y, Yang R, Cao Y, et al. A Framework for Identifying Bird Conservation Priority Areas in Croplands at National Level[J/OL]. *Journal of Environmental Management*, 2022, 324: 116330[2024-07-18]. https://doi.org/10.1016/j.jenvman.2022.116330.

[20] Brum F T, Graham C H, Costa G C, et al. Global priorities for conservation across multiple dimensions of mammalian diversity[J]. *Proceedings of the National Academy of Sciences*, 2017, 114(29): 7641-7646.

[21] Silvestro D, Goria S, Sterner T, et al. Improving biodiversity protection through artificial intelligence[J]. *Nature Sustainability*, 2022, 5: 1-10.

[22] Wisz M S, Hijmans R J, Li J, et al. Effects of sample size on the performance of species distribution models[J]. *Diversity and Distributions*, 2008, 14(5): 763-773.

[23] Keating K A, Cherry S. Use and interpretation of Logistic regression in habitat-selection studies[J]. *Journal of Wildlife Management*, 2004, 68(4): 774-789.

[24] Phillips S J, Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation[J]. *Ecography*, 2008, 31(2): 161-175.

[25] Steven J P, Miroslav D, Robert E S. Maxent software for modeling species niches and distributions (Version 3.4.1)[DB/OL]. [2024-04-26]. http://biodiversityinformatics.amnh.org/open_source/maxent/.

[26] Eupen C V, Maes D, Herremans M, et al. The impact of data quality filtering of opportunistic citizen science data on species distribution model performance[J]. *Ecological Modelling*, 2021, 444: 109453.

[27] Henckel L, Bradter U, Jönsson M, et al. Assessing the usefulness of citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a systematic protocol[J]. *Diversity and Distributions*, 2020, 26(10): 1276-1290.

[28] 黄越，顾燚芸，阳文锐，等.如何在北京充分实现受胁鸟类栖息地保护[J].生物多样性，2021，29(3)：340-350.

[29] Tehrani N A, Naimi B, Jaboyedoff M. A data-integration approach to correct sampling bias in species distribution models using multiple datasets of breeding birds in the Swiss Alps[J]. *Ecological Informatics*, 2022, 69: 101501.

[30] 刘海龙，王茜，宋洋，等.北京第二绿化隔离地区以鸟类为主的城市生物多样性保护规划途径[J].中国园林，2022，38(10)：6-13.

（编辑/刘欣雅）

作者简介：

侯姝彧
1991年生/女/黑龙江大兴安岭人/博士/华南农业大学林学与风景园林学院讲师/研究方向为国家公园与自然保护地、生物多样性保护与生态修复(广州 510642)

尚轩仪
1995年生/男/山西太原人/广州市水生态建设中心工程师/研究方向为湿地生态与生物多样性保护管理(广州 510660)

刘 彦
2000年生/女/广东肇庆人/华南农业大学林学与风景园林学院在读硕士研究生/研究方向为国土景观保护和生态修复(广州 510642)

李 晖
1967年生/女/重庆人/博士/华南农业大学林学与风景园林学院教授，博士生导师/研究方向为国土景观保护和生态修复、生态系统服务、生物多样性(广州 510642)

梁健超
1986年生/男/广东广州人/博士/广东省科学院动物研究所广东省动物保护与资源利用重点实验室助理研究员/研究方向为景观生态学、群落生态学(广州 510642)